Exploring a voice user interface to convey information in an e-commerce website

Christopher Liljestam

Interaktionsdesign
Bachelor
22.5HP
VT/2019
Supervisor: Alicia Smedberg

Author: Christopher Liljestam
University username: AF9105
E-mail: christopheryljestam@gmail.com
Abstract

Screen readers for visually impaired users are poorly optimized for e-commerce websites hence the exclusion of the content. It creates a societal need for accessibility of the content in e-commerce websites for the visually impaired users. This study explores how six blindfolded participants could co-design a Voice User Interface (VUI) in an e-commerce website to convey its information that creates a good user experience for visually impaired. The result of a co-design workshop with interaction design practices showed that a VUI should be humanlike and convey relevant information. Failed speech recognition and overwhelming information had a negative impact on the user experience. To cope with the problems, the VUI should provide more control to the users by conveying explicit confirmations and retrospective information from past shopping trips. Due to the difficulties in finding visually impaired participants, the design process was not completed hence the ideation needs an additional design process.

Keywords: social innovation, co-design, voice user interface, visually impaired, e-commerce, accessibility, interaction design, humanlike, blind navigation
Acknowledgements

I am grateful to all my teachers who have shared their knowledge and guided me for the past three years. I would like to thank my supervisor Alicia Smedberg for the constructive feedback and advices which helped me to confront the problems that I faced. I want to thank the participants in the co-design workshop for their interesting insights and input. I also want to thank my girlfriend for her love, support, and motivation to do my best. Lastly, I want to thank Junkyard for giving me permission to use their website for my thesis project.
# Table of contents

1. Introduction ........................................................................................................... 6  
   1.1 Aim ................................................................................................................. 6  
   1.2 Problem area ................................................................................................. 6  
   1.3 Research question .......................................................................................... 6  
   1.4 Junkyard ......................................................................................................... 7  
   1.5 Contribution .................................................................................................... 7  
   1.6 Target group ................................................................................................... 7  
   1.7 Delimitations .................................................................................................. 7  
   1.8 Ethics ............................................................................................................... 8  

2. Theory and related work ......................................................................................... 8  
   2.1 Social Design and Social Innovation ............................................................. 8  
   2.2 Interaction Design .......................................................................................... 9  
   2.3 Accessibility .................................................................................................... 9  
   2.4 Co-design ........................................................................................................ 10  
   2.5 Who are the visually impaired? ....................................................................... 10  
      2.5.1 Exploring navigation tools for visually impaired ...................................... 11  
   2.6 What defines a voice user interface? ............................................................... 13  
      2.6.1 Human-like theory .................................................................................... 14  
      2.6.2 Conversational theory ............................................................................. 15  
   2.7 Related work ................................................................................................... 15  

3. Methods .................................................................................................................. 17  
   3.1 The lab ............................................................................................................ 18  
   3.2 Co-design workshop ....................................................................................... 18  
   3.3 Cooperative Principle ..................................................................................... 18  
   3.4 Prototyping ....................................................................................................... 19  
      3.4.1 Low-fidelity prototyping ......................................................................... 19  
      3.4.2 Wizard of Oz ............................................................................................ 19  
      3.4.3 Experience prototyping .......................................................................... 20  
   3.5 Brainstorming .................................................................................................. 20  

4. Design process ....................................................................................................... 21  
   4.1 Pre-research stage ........................................................................................... 22
1. Introduction

1.1 Aim

The intended aim of this thesis is to explore through interaction design practices how a voice user interface (VUI) could be designed to convey the content of an e-commerce website through the perspective of visual impairment. The purpose is to improve the navigation and accessibility for visually impaired users in e-commerce websites. The practices are utilized in a controlled co-design workshop with sighted participants who were exposed to the perspective of vision impairment. The expected outcome of the research is to get a better comprehension of how a VUI could convey the information in an e-commerce website in order to deliver a good user experience.

1.2 Problem area

The existing technology such as screen magnifiers and screen readers provide accessibility for visually impaired users in their digital surroundings on the Internet. For instance, the screen readers read the digital content loudly thus the visually impaired can comprehend the content and navigate in the graphical user interfaces (GUIs). However, GUIs of e-commerce websites consist of several problems for the visually impaired. The screen readers cannot detect images, animations, navigational buttons, as well as the layout of tables and charts (Sohaib & Kang, 2016). This problem leads to the societal issue of excluding the visually impaired users from e-commerce websites as they do not have access to the user experience as regular users. The exclusion of visually impaired users creates a societal need for accessibility.

1.3 Research question

The thesis consists of the two questions. The main research question is as following:

How could sighted individuals co-design a voice user interface in an e-commerce website to convey its information through the perspective of visual impairment?

Followed by the sub-question:

How may sighted designers deliver a good user experience for visually impaired users?
1.4 Junkyard

Junkyard is the biggest street fashion online store in Scandinavia. It is an e-commerce website where the target audience is the younger generations. I got permission to use Junkyard for research purposes. The website was utilized as the e-commerce environment by the participants in the co-design workshop.

1.5 Contribution

The work is a contribution to society as it investigates a need for a societal problem. The outcome of the work contributes to designers who are willing to design a VUI for visually impaired users. It will aid designers who seek a deeper understanding how to design a VUI in order to create a good user experience for visually impaired users. The ideation is for designers who want to continue the work of improving the accessibility for visually impaired users in e-commerce websites with a VUI.

1.6 Target group

The intended target group for the thesis includes people with low-vision and blindness between the age of 18-35, who want another approach to navigate and access e-commerce websites. I chose this target group as it corresponded well with Junkyard’s target audience. However, as my project progressed, I realized that it was a difficult target group to reach. Therefore, I had to change the target group to the age 18 and above, to increase my chances to find visually impaired users.

1.7 Delimitations

The research of this paper does not involve visually impaired participants as I did not find any interested individuals to participate in the activities. Therefore, I was forced to recruit sighted participants to the workshop. The process of trying to recruit visually impaired is presented in the section 4.2.

It is understandable that the sighted participants are not actually visually impaired hence they do not have the same experience of being permanently visually impaired compared to individuals who are. Therefore, if I had visually impaired participants, I would have more in-depth and sophisticated results regarding their needs in the workshop activities.

As the thesis explores how VUIs convey the content of e-commerce for visually impaired people, it will only focus on the aspect of hearing to navigate. Otherwise, visually impaired can rely on other senses as well such as smelling or touching but these senses won’t be considered in the co-design workshop nor the ideation.
1.8 Ethics

The research of my thesis project will follow the ethical standards that are defined by the Swedish Research Council (Swedish Research Council, 2017). All the participants in the co-design workshop had to sign a consent form for the approval of being photographed as well as audio and video recorded. They could refrain to participate in the activities and the documentation at any time. In the consent form, they were also acknowledged that their data won’t be used outside the context of this thesis as well as the publication of this thesis in the database of Malmö University. I informed the participants the information above orally as well.

2 Theory and related work

The second chapter acknowledges the relevant theories and work surrounding the thesis, in order to give the reader some needed context. The chapter will cover topics such as the social design and innovation, interaction design, accessibility, co-design, the visually impaired, voice user interfaces and humanlike and conversational theory.

2.1 Social Design and Social Innovation

Designers need to consider the social and moral responsibilities before designing a product or a service. Victor Papanek (1971) states that the designers’ responsibilities go far beyond the responsibility of how well the designs are received at the market. A designer needs to examine the potential of the design based on moral and social values. Whether a design of the product or service should serve as a beneficial contribution for the society (Papanek, 1971).

Ezio Manzini defines social innovation as new ideas (products, services and models) that simultaneously meet social needs and create new social relationships or collaboration (Manzini, 2015, p.11); Innovations that are good for society as well as enhancing the society’s engagement to act. Therefore, a VUI for visually impaired users in e-commerce websites could be a design for social innovation as it meets the need of accessibility.

Social design is a design activity which involves people to solve social sensitive issues which are not dealt with by the market or by the state. The participants are usually people who do not have the finances or political position hence they utilize social design to make their voices heard (Manzini, 2015).

Manzini states that social design and social innovation shares the same notion of design. However, they refer to different activities and have very
different indications in the adjective term of “social”. Therefore, social design is not a design for social innovation. Manzini argues that social innovation utilized the term “social” in the sense of social forms such as a society has been built. Moreover, social design utilizes the term in the sense of indicating extreme problems with social sensitive issues such as poverty which the market or government have not fixed (Manzini, 2015).

The actual design for social innovation is defined by Manzini as following (Manzini, 2015, p. 62):

“Design for social innovation is everything that expert design can do to activate, sustain, and orient processes of social change toward sustainability.”

Manzini further explains that design for social innovation need to include the stakeholders who bring their own experience-based knowledge and design capacity through co-design activities, in order to detect issues and solve them. Therefore, a design for social innovation is the designer's contribution towards a social change (Manzini, 2015).

2.2 Interaction Design

Rogers, Sharp, and Preece defines interaction design as designing interactive products to support people in their everyday and working lives (Rogers, Sharp & Preece, 2002, p.6). The field of interaction design is very user-centred which implies that interaction designers strive to create user experiences that enhance and extend users’ way of work, communicate, navigate, and interact. A VUI in an e-commerce website would be an approach to extend how the visually impaired people navigate and interact with the website.

To design a product to support people, the interaction designer must be knowledgeable about the target audience and how the design could provide usable support. Therefore, in an interaction design activity there are two aims; the first aim is to understand the users within the target audience in order to identify their needs. The second aim is to produce a foundation based on requirements in order to move forward into developing the design (Rogers, Sharp & Preece, 2002).

2.3 Accessibility

The visually impaired users’ access to e-commerce websites regards accessibility which has several definitions. The International Standards Organization (ISO) defines accessibility as (Connor, 2007, p.5):

“The usability of a product, service, environment or facility by people with the widest range of capabilities (ISO TC 16071).”
The definition of accessibility articulates the crucial factor to users’ diversity when it comes to user-centred design. The accessibility of a design should aid diverse users to achieve their similar specified goals.

The consequences of not designing for accessibility are the risks of creating barriers for disabled users. The barriers will exclude the disabled users as they will not have the access to the content of website compared to other users. However, a good accessible design can improve the experience for regular users as well. Designers that are responsible for excluding users from the content are not always aware of it (Connor, 2007).

2.4 Co-design

The workshop of this thesis is a co-design activity hence further context about the co-design is needed. In the area of participatory design, there is two terms called co-design and co-creation which are often confused by each other. According to Stappers and Sanders (2008), co-creation serves as an act of collective creativity that is shared by two people or more. However, Stappers and Sanders (2008) states that co-design is collective creativity which is applied in a design process. Stappers and Sanders further explain co-design as following (Sanders & Stappers, 2008, p.14):

“(…) co-design is a specific instance of co-creation. Co-design refers, for some people, to the collective creativity of collaborating designers. We use co-design in a broader sense to refer to the creativity of designers and people not trained in design working together in the design development process.”

The explanation implies that a design process with the influence of co-design changes the role of the user. In a traditional design process, the user is viewed as a subject which the designer with researched knowledge interviews and observes the user in order to build upon the known knowledge. However, in co-design, the user becomes a designer based on their experience and plays an important part in the extension of knowledge, ideation, and design development (Sanders & Stappers, 2008).

2.5 Who are the visually impaired?

According to WHO, 1.3 billion people live with some form of vision impairment. Circa 80% of these people have vision impairment that is avoidable, the 20% have either severely impaired vision or blindness whose vision cannot be fixed. That is an estimation of 253 million people who live with inevitable vision impairment (WHO, 2018).

The vision impairment decreases the individual’s ability to see which causes low vision or blindness. Visual acuity is a measurement for visual impairment which measures the sharpness of the human eye regarding shapes and detailed objects at a given distance (Marsden, Stevens, & Ebri, 2014).
People with moderate to severe visual impairment have a visual acuity worse than 6/18 to 3/60. Blind people have a visual acuity worse than 3/60 which implies that a blind individual can only see objects from 2 or less metres compared to 60 metres as an average sighted individual (Bourne et al., 2017). Therefore, individuals with severe visual impairment or blindness need mobility aids in order to be independent while navigating their surroundings. Two challenges arise to achieve independency which are accessing written content and navigating with safety and efficiency (Giudice & Legge, 2008).

The visually impaired people must take advantage of their other senses than vision while navigating, such as hearing, smelling, and touching. It becomes a vital demand to interpret sensory feedback to learn the surroundings. For instance, they use long canes to learn locations of steps and doorways to avoid missteps or collisions. Moreover, they must interpret the auditory feedback of a traffic light to determine when the traffic is green, to ensure the crossing is safe. Without any reliable vision, there is also a risk for them to lose their track in the environment which makes it difficult for them to know their current position in relation to their destination (Giudice & Legge, 2008).

According to Giudice and Legge (2008), a guide dog and a long cane are the most common mobility aids for the visually impaired. They both have the same purpose which are to detect and identify obstacles in the path of the visually impaired.

The recent years of development of technologies has significantly improved the visually impaired access and navigation in the digital space and its content. For instance, technologies such as screen readers, screen magnifiers, and electronic Braille displays. The digital tools will be further explained in the subsequent section.

### 2.5.1 Exploring navigation tools for visually impaired

The navigation tools for visually impaired that utilize technology is called electronic travel aids (ETAs). For instance, navigation tools based on technology such as GPS, sonars, and cameras (Giudice & Legge, 2008). Conveying information from an e-commerce website makes the VUI a digital navigation tool to an extent as the visually impaired could navigate to access its content. Therefore, it is necessary to explore how visually impaired users interact with navigation tools to interpret the feedback. According to Giudice and Legge (2008), there is a tendency for technology-driven tools to fail with establishing themselves. The tools might work in theory but fail to satisfy the visually impaired users’ needs in practice.

Based on real-world navigation tools, Giudice and Legge (2008) state that a tool is effective if it conveys selective information that is specific according to the current environment. The complexity of a tool depends on the amount of information it conveys. However, Giudice and Legge do not mention the consequences if the tool does not select relevant information. Kristjánsson et
al. (2016) argues that if a tool does not select relevant information to convey, then it risks to sensory overload the visually impaired with information which must be avoided.

Giudice and Legge (2008) further mentions that empirical research is required to ensure that the tool is user-friendly and provides enough information. Giudice and Legge (2008) also highlights the aesthetic aspect of a tool where the design of a tool should not give an intrusive feeling. Some visually impaired users value the aesthetic greater than the tool’s capability to improve the navigation.

The tools showcased below are some of the technology-driven tools the visually impaired utilize in order to navigate or access information in the digital environment.

2.5.1.1 Screen readers

A screen reader is a software navigation tool which utilizes text-to-speech on GUIs which conveys the content one line at the time, from top to bottom in a fast manner. The visually impaired users navigate by the keystrokes on a keyboard which makes navigating in digital space possible and the content of web pages accessible. The visually impaired can also interrupt the speech of the screen reader by navigating between the sections of the interface (Bose, 2014).

However, there is limitations regarding accessing web pages. The screen readers rely heavily on the *HyperText Makeup Language* (HTML) of the web pages thus the visually impaired will receive another kind of experience than the developer visually intended for the sighted users (Bose, 2014).

A costly and popular choice of screen reader is JAWS which also supports electronic Braille displays. Electronic Braille displays provides Braille text by generate tactile dots that is coherent with the visual text on the digital screen (Bose, 2014).

Lastly, as mentioned in the section 1.2, due to the limitations of screen readers and how they operate, it becomes a societal problem in e-commerce websites for the visually impaired. The screen readers cannot read images, animations, navigational buttons, as well as the layout of tables and charts (Sohaib & Kang, 2016). Screen readers were a part of the co-design workshop in the section 4.4.5.

2.5.1.2 Screen magnifiers

Visually impaired with some functional vision navigates with screen magnifiers software that zoom into the screen thus the visual content becomes enlarged and perceptual. The magnified area centres around the current position of the cursor which covers the full screen or just a part of it. Some screen magnifiers can invert colors so the light text of the content has a black background or yellow text on a blue background. In the screen magnifier ZoomText, the user can customisable color mapping where specific
colors can be mapped. It can showcase text in negative contrast hence it will prevent screen glare which can otherwise be harmful towards the users’ eyes (Lenkhorn, Vans, Ing, Urniawan, & Utcliffe, 2003).

According to Vigo and Harper (2013), the partial magnified area is also a drawback for screen magnifiers as it does not convey the context of the whole web page. Screen magnifiers were also a part of the co-design workshop in section 4.4.5.

2.5.1.3 Voice user interface

In recent years, VUI has exploded in popularity in a shape of a voice assistant in smart devices, such as smart speakers and smartphones (Good Rebels, 2018). For instance, Google Assistant in Android devices and Siri in Apple products. The established VUIs have a vital role in the co-design workshop in the section 4.4.4.

Visually impaired users benefit from a voice assistant due to its auditory feedback and voice-based interaction. They can access information online e.g. asking for the weather or conduct other inbuilt functions e.g. math calculations (Pradhan, Mehta & Findlater, 2018). In the subsequent section, VUI will be further explained.

2.6 What defines a voice user interface?

A voice user interface is defined by Schnelle-Walka as “user interfaces using speech input through a speech recognizer and speech output through speech synthesis or pre-recorded audio.” (Schnelle-Walka, 2012, p. 1)

Schnelle-Walka mentions the three different basic elements that are required in a VUI. Automated Speech Recognition (ASR) defines the grammar of the user’s speech input which the system apprehends as either a prompt or a command. Dialog logic is the second element which based on the speech input, determines what actions the system will do next. In this case, a dialog refers to spoken words which are exchanged between the system and the user. The last element is Text To Speech (TTS) which serves as system audio feedback that is either an informative response or a prompt to the user (Schnelle-Walka, 2012).

In the dialog segment, there is a term called Natural Language (NL) which is explained by Susan J. Boyce (Gardner-Bonneau & Blanchard, 2008, p. 43):

“(…) used to describe applications in which the caller can speak more than a restricted set of command words or menu choices and the system will respond appropriately.”

She further describes that the goal of NL, is to predict what the user might say at any given time during the dialog. Therefore, a NL system consists of many different variations of prompts and commands, thus the NL can recognize the users’ variated speech input. However, the NL system is not always a solution for every application. A NL system is ideal to replace
applications with several functions, where the user must go through a redundant number of steps to achieve the goal (Gardner-Bonneau & Blanchard, 2008). Therefore, an e-commerce website could be an ideal setting to implement VUI as the customers need to go through several steps in order to find and purchase clothes.

2.6.1 Human-like theory

A VUI utilizing NL is an example of the notion anthropomorphism. Anthropomorphism is defined by Lemoine and Chérif as “the tendency to attribute human characteristics in inanimate objects” (Lemoine & Chérif, 2019, p. 29). In this paper, the definition of anthropomorphism regards humanlike characteristics in a VUI. Therefore, I will refer anthropomorphism as humanlike in this paper.

According to Qvarfordt et al. (2003), users’ perception of a system as humanlike or non-humanlike influences their descriptions of their interaction with a system. For instance, interactions with a humanlike system are described as conversations while interactions with a non-humanlike system are described as commands. Ovarfordt et al. (2003) also mention that users prefer a humanlike system once they have experience it. Burgoon et al. (2000) agrees by implying that VUIs designed with humanlike characteristics are rated highly by users in terms of utility. Burgoon et al. further states the following (Burgoon et al., 2000, p. 571):

“Interfaces fostering an increased sense of involvement and mutuality would be an optimal design for human-computer collaboration, such as in assisting in the decision-making process”

Burgoon et al. (2002) also explain that these collaborations make users more active in the interactions where relationship building is possible due to the humanlike characteristics of the VUI.

A VUI which strives to achieve human characteristics such as human intelligence, belongs to the field of artificial intelligence (AI). The field of AI is defined by Poole and Mackworth as “the synthesis and analysis of computational agents that act intelligently” (Poole & Mackworth, 2010, p.3). According to Poole and Mackworth (2010), a system is behaving intelligent when it makes appropriate choices within its programmed capabilities to fulfill its purposes. It is also considered intelligent if it learns from experiences. In order to test the intelligence of a system in contrast of human intelligence, Alan Turing created the Turing test in 1950. The test is an imitation game where a human interrogator asks an AI-driven witness any questions via a text interface. If the interrogator cannot distinguish the AI as a system, then it will be considered intelligent and passes the test (Poole & Mackworth, 2010).
2.6.2 Conversational theory

The humanlike characteristics make the dialog between the system and the user into a conversation which Oxford dictionaries defines as a talk, especially an informal one, between two or more people, in which news and ideas are exchanged (Oxford dictionaries, 2019). Radlinski and Craswell (2017) argue that both sides of a conversation are initiated based on the word exchange in the definition, hence the term mix-initiated. According to Radlinski and Craswell, the necessary implementation of a conversational system search depends on the complexity of a task. A simple task that merely requires one search to find expected relevant information does not need a back-and-forth conversation with a voice assistant (Radlinski & Craswell, 2017).

Radlinski and Craswell provides a list of five different properties for conversational search systems that should be achieved to create a good user experience. They call this set of properties for RRIMs properties (Radlinski & Craswell, 2017, p. 120):

- **User Revealment**: The system helps the user to express or potentially discover their true information need, and possibly also long-term preferences.
- **System Revealment**: The system reveals to the user its capabilities and main purpose, building the user's expectations of what it can and cannot do.
- **Mixed Initiative**: Both the system and user can take the initiative as appropriate.
- **Memory**: The user can reference past statements, which implicitly remain true unless contradicted.
- **Set Retrieval**: The system can reason about the utility of sets of complementary items.

2.7 Related work

Exploring previous related work in the field is necessary in order to comprehend the outcomes of these work and how one's own work can be further build upon or compared. It can encourage exploration regarding the design of how a VUI could convey in an e-commerce website in order to deliver a good user experience.

Myers et al. (2018) conducted a research where their aim was to comprehend the obstacles users encounter by interacting with an unfamiliar VUI and observe which strategies are utilized in order to overcome them. In the study, the researchers used a VUI calendar manager named *DiscoveryCal* on 12 sighted participants in controlled sections with pre-written tasks. For instance, creating events, modifying events, and deleting events. The research does not involve visually impaired people, it is still essential to include it in this paper as the obstacles and strategies revolving VUI can occur in this
paper's workshop as well as it regards the user experience. In their study, Myers et al. (2018) discovered by behavioral patterns, four obstacle categories and ten strategies that were used by the participants.

Myers et al. (2018) states that the most common obstacle category was errors regarding the system's flawed capability to recognize the utterances by the participants which resulted in misunderstandings. To overcome the obstacle, the participants commonly articulated their utterances as a strategy by speaking louder or slower. Further strategies such as creating a new utterance by adding more information or simplifying the input by utter keywords were utilized as well, in hope to facilitate the comprehension of the system. Their research also showcased that the participants recalled the prior correct utterances while facing the obstacles which implies that the participants learned the VUI and reflected on the knowledge. However, the participants relied more on guessing and exploration than recalling from their knowledge.

According to Myers et al. (2018), the flawed speech recognition of the VUI was not a big threat against the user experience as the participants knew that improved articulation was the solution. The other three obstacles were a greater threat as they caused confusion and frustration which lead to other strategies driven by frustration. One of these obstacles was failed feedback from the system where the feedback was unclear or not provided to the participant. The other obstacle was the software errors of the system which included software bugs that hindered the system to work as intended. The last obstacle discovered in their study regarded users' unfamiliarity with the VUI where the users did not know the supported functions nor the correct way to structure their utterances for the system to understand (Myers et al., 2018). According to Myers et al. (2018), the obstacles lead to flawed strategies such as quitting the task completely, accepting the incorrect edit after several attempts, and rapidly conduct attempts in frustration without any thought.

As my project does not include visually impaired participants, it is important to learn their opinions about VUI and how it should convey information, from another project.

An example for such a project is a study by Olsson (2017), who investigated how to design interfaces for visually impaired to convey contextual information in stores and public transportation. In the study, Olsson (2017) conducted interviews with 10 visually impaired participants regarding interfaces including VUI and the problems they encounter while navigating outside. Olsson (2017) also conducted a workshop with six people with design backgrounds in order to ideate a concept. The final concept was an application for shopping which was based on diverse technologies including VUI. In Olsson's concept (see figure 1), to avoid overwhelming the visually impaired user with information, the VUI conveys only the three most popular items on the lists before asking the user to continue hearing the whole list. The concept was not made into a prototype thus it was not tested in practice with the target audience.
Figure 1. The VUI conveys the three most popular items before asking the user for more information to prevent overwhelming information (Olsson, 2017).

Based on the interview findings, (Olsson, 2017) mentions that the visually impaired participants highlighted the problems of a VUI where the VUI did not understand the utterances or did not provide coherent information. For example, the voice assistant Siri did not read the result from a Google search (Olsson, 2017). This finding is equivalent to the findings of Myers et al. (2018) regarding failed feedback. Olsson (2017) also mentions another finding where the visually impaired evaluate highly the humanlike voice of the VUI. They expressed that the voice needs to be pleasant and humanlike. A visually impaired participant did not want to interact with a VUI if it sounds too robotic. According to Olsson (2017), the visually impaired participants were generally positive towards VUI where some of them are using a VUI in a daily basis.

This paper will build upon the previous work by strengthen the insights of the co-design workshop in the discussion chapter and aid the ideation of the design opportunities in the section 4.6.

3 Methods

The third chapter includes all the different interaction design methods that are utilized in this study. It will provide a comprehension through informative definitions of these methods and their relevancy to the study. The methods are the lab, co-design workshop, the cooperative principle, prototyping as low-fi prototyping with Wizard of Oz, and experience prototyping, and brainstorming.
3.1 The lab

The co-design workshop followed the circumstances of a laboratory. According to Koskinen, Zimmerman et al. (2011), in a laboratory the subjects are removed from their natural environment and brought into a controlled area where they are involved in experiments. A lab environment helps researchers to focus on one thing at the time which make it easier for researchers to gain answers for their hypothesis. The most important matter in a laboratory are the experiments. A successful research is when the experiments manage to build a basic relationship between the participants and the items of interest, which the researcher can modify to obtain the wanted results (Koskinen, Zimmerman, Binder, Redström, & Wensveen, 2011).

3.2 Co-design workshop

A co-design workshop is an interaction design practice with a co-design approach. As mentioned in the topic of co-design from the 2.4 section, participants in the workshop co-design with the designer, in order to ideate and develop knowledge and concepts. The experienced designer must provide tools and techniques to facilitate the participants ideation and expressions. Therefore, the designer possesses the role as a facilitator in a co-design workshop. The designer aids the participants by guiding and leading in order to encourage the participants to express their creativity (Sanders & Stappers, 2008). The workshop structure and goals are further explained in section 4.3.

3.3 Cooperative Principle

The Cooperative Principle was created by the linguistics philosopher Paul Grice (1975) which is divided in four different categories, with the purpose to sustain a cooperative and efficient conversation. As Burgoon et al. (2002) mentioned, VUI are suited when users want to collaborate, therefore a humanlike VUI should follow the cooperative guidelines. The categories are quantity, quality, relation, and manner. Each category has a maxim which consists of rules that each person of the conversation needs to understand and follow. The maxims with their respectively categories are as following (Grice, 1975):

- **Quantity:** The speaker should not say more information than necessary
- **Quality:** The speaker should say information that he believes is true.
- **Manner:** The speaker should try to be clear and explain so it makes sense for others.
- **Relation:** The information should be relevant in the conversation.

The Cooperative Principle was utilized as a method in the co-design workshop to see how the sighted participants as well as the established voice
assistants followed Grice’s maxims while conveying information, see section 4.4.3 and 4.4.4.

3.4 Prototyping

According to Rogers, Sharp and Preece (2002), a prototype can be in any shape and form, from a paper-built wireframe of an application to a technical and sophisticated version of a software. Rogers et al. state that a prototype allows the stakeholders to interact with the intended ideation of the product. With usability testing, the stakeholders can evaluate its feasibility based on their experiences and exploration of the use cases. Furthermore, prototypes can be tested by designers as well to detect or solve problems at any stage of the design process. Testing prototypes in the intended scenarios answers questions as well as supports designers regarding their design choices, e.g. if functions of the system are appropriate and enough for the intended use case or if the audio feedback of the system has a sufficiently long duration (Rogers, Sharp & Preece, 2002).

3.4.1 Low-fidelity prototyping

Prototypes are commonly divided into two different groups which determines the fidelities of the prototypes: Low-fidelity prototypes and high-fidelity prototypes. Rogers et al. (2002) mention that the aesthetics of low-fidelity prototypes are not the same as the final product due to their different materials, for instance paper and cardboard. Low-fidelity prototypes are commonly produced in the early stages of the design process due to their low expenses, simplicity to develop and modify, and short creation time. The purpose of low-fidelity prototypes is to encourage exploration of ideas which are never intended to be in the final product (Rogers, Sharp, & Preece, 2002).

3.4.2 Wizard of Oz

One of the methods in low-fidelity prototyping is the Wizard of Oz method. The origin of the method is from one of the scenes in the film with the same name. In the scene, Toto the Dog pulls back a curtain which reveals the secret of the wizard who is a regular man that pulls levers and turns on switches (Rouse, 2014). The practice is identical to the scene whereas the user interacts with a software which is operated by another human (Rogers, Sharp & Preece, 2002). In other words, the responses of the system are human controlled hence illusion of the intended experience.

The Wizard of Oz method was applied in one of the activities in the workshop as it felt as the most natural method, to illustrate the potential of a VUI-based e-commerce. It allowed the participants to roleplay as voice assistants and blind customers. The method was used in section 4.4.3.
3.4.3 Experience prototyping

Experience prototyping is a form of prototyping which can be explained as (Boulton, Devriendt, Brunn, Derudder, & Witlox, 2011, p. 424):

“(…) any kind of representation, in any medium, that is designed to understand, explore or communicate what it might be like to engage with the product, space or system we are designing.”

Experience prototyping allows designers, users and stakeholders to understand the concept by experiencing it themselves rather than observing a demonstration or the experience of someone else. Through experience prototyping designers can identify issues and design opportunities (Boulton et al., 2011). Boulton et al. further identify three different kinds of activities where experience prototyping can be valuable in a design process (Boulton et al., 2011):

- To understand already existing user experiences and context
- To explore and evaluate design ideas
- To communicate ideas to an audience

The two first mentioned activities in the bullet list, are highly relevant for my project as they were applied in my workshop, see section 4.3.2. The sighted participants had to experience the sense of being visually impaired in order to provide relevant insights for the research. Furthermore, the concept of a VUI in an e-commerce website were tested and explored by the blindfolded participants.

3.5 Brainstorming

Brainstorming is an effective individual or group method for encouraging quantitative ideation in terms of creativity and problem solving. Brainstorming can be applied any time during the design process whereas new ideas or solutions are needed. However, the method is mostly used between the early to middle stages.

During a brainstorming session, it is essential that the quality of ideas and solutions are not being criticised, as criticism can cripple and discourage new wild ideas (Wilson, 2013). I utilized brainstorming to create the activities in the workshop as well as brainstorming ideas for the design opportunities, see sections 4.4.1 and 4.6. Brainstorming encouraged me to be more creative which helped me to create activities that were creative and useful.
4 Design process

The design process of this thesis never got completed, nor did it involve visually impaired people as I failed to find any interested. The intentions were to conduct an iterative design process with a final prototype that includes the solution to the design opportunities. As I did not find any visually impaired users, I did not have the opportunity to conduct any ethnographic fieldwork. The ethnographic fieldwork was replaced by a controlled co-design workshop with six sighted participants who experience prototyped as visually impaired users. This is not ideal as Norman (2013) highlights that ethnographic fieldwork is needed in order to understand the true needs of the target audience in their natural environment. However, the co-design workshop was necessary in order to take the design process forward. The co-design workshop took place after the failed attempts to recruit visually impaired people. It provided me insights from a visually impaired perspective which I could use for my ideation.

In figure 2, one can see a visual presentation of the design process and its lacking stages. As mentioned earlier, the design process is lacking visually impaired users. Furthermore, it is lacking prototypes based on the ideation, user testing with prototypes of the target audience is missing as well as iterations. Lastly, the design process is lacking the redefined solution to the design opportunities. The design process does provide design opportunities and ideation based on the workshop which need further empirical research and ideation in an additional design process.

Before conducting the process of finding visually impaired participants, I conducted a literature study to select coherent academic research towards the research questions. The study was essential to explore and learn the societal problem and the visually impaired people including their digital tools. Academic research was also needed to gain a further understanding about VUIs. Furthermore, the literature study was essential to understand which relevant methods to select for the co-design workshop. Course literature have been used alongside with the academic texts through the main sources which were ACM Digital Library, Libsearch, eBook Collection (EBSCOhost), and Google Scholar.
Figure 2. The figure illustrates the design process and its missing stages to reach the final solution. (Author’s documentation, 2019).

4.1 Pre-research stage

Before conducting the literature study, I had an interest to investigate conversational voice interfaces in an e-commerce website like Junkyard. The interest originated from an article written by Bentahar from Forbes. In the article, Bentahar (2018) states that voice searching is changing the e-commerce industry due to the rapidly increased usage of the established voice assistant compatible devices for shopping purposes (Bentahar, 2018). I began contemplating how an inbuilt voice assistant could convey information in a regular e-commerce website and how it could be designed to provide a good user experience.

I had to select a main target group for the paper. I created a mind map thus I could lay out the different routes of target groups that I could choose from (See figure 3).

My first intentions were to involve regular e-commerce shoppers that represent the target audience of Junkyard. The target audience of Junkyard consists of the younger generations, approximately the age group from 18 to 35 years old. However, the potential of a conversational VUI in an e-commerce website for regular visitors felt minor as they could already have a great user experience. I wanted to increase the value of the thesis thus the research would become beneficial for a target group. Therefore, I wrote down three different kinds of disorders as the reader can see in figure 9 below. These disorders were dyslexia, vision impairment, and dexterity. I selected
visually impaired users as the target group because I identified a societal need for a voice assistant in an e-commerce website.

Figure 3. I created a digital mind map of the different target groups I considered for this paper. The yellow highlighted lines showcase the route I took (Author’s documentation, 2019).

### 4.2 The attempts to find visually impaired people

#### 4.2.1 Contacting organizations

In order to find visually impaired people, I began to contact the *Swedish Association of Visually Impaired* (Synskadades RiksFörbund) which is the main organization in Sweden for the blind and partially sighted people (See figure 4). The organization is divided into several sections for each local region across Sweden. I contacted four organizations through email which were located within their respective municipalities nearby Malmö, south of Sweden. As my intended target audience was directed towards the youth, I contacted another branch of the association with the name *Unga med Synnedsättning* (US) which focuses on the youth within the visually impaired people. I sent an email to the office of US as well as the office which represents the south of Sweden (US Syd).

Two of the four organizations responded on my email the following day and wrote that they merely consist of elderly members. Two days since I wrote the email, the clerk of US Syd responded and told me that he was going try to find people who might be interested. Meanwhile, US forwarded my email to US
Syd as they were unable to help. Six days later, the clerk contacted me to say that all his contacts have either declined or not answered yet. The outcome of these emails forced me to change strategy, to increase my chances to find participants.

![Figure 4. A compilation of the emails I wrote to the different organizations with visually impaired members, in order to recruit participants to my research (Author’s documentation, 2019).]

The difficulty to recruit participants made me reach a crossroad where I decided to change the age group of my target group, by removing the age limit of 35 years old thus the elderly visually impaired people could participate in the research as well.

Lastly, three weeks after the contact with the first organization, I called the Swedish Association of Visually Impaired of the province Scania itself where the municipalities I previously contacted are located. The clerk said that the elderly does not use digital tools to access the Internet as much as the younger people do. He also tried to gather a group of interested people, but this attempt failed as well.

### 4.2.2 The findings of the recruitment

As I did not manage to gather any visually impaired participants, I could only proceed from the insights I gained from the failed recruitment in order to investigate the factors behind the failed attempts. This paper has a topic directed towards a very niche group of people. The term niche is defined by Oxford dictionaries as *denoting or relating to products, services, or interests that appeal to a small, specialized section of the population* (Oxford dictionaries, 2019).

According to Nestor, approximately 100.000 visually impaired people of the Swedish population are registered in the Swedish vision centers (Nestor, 2015). Visually impaired have a small population in Sweden hence I had to go through several organizations to reach them which is time-consuming. Furthermore, the contact persons in the organization must find interested people which takes time as well. Without any visually impaired participants, I had to include sighted participants in a co-design workshop in order to continue the design process. As stated by Giudice and Legge (2008), the visually impaired rely on their hearing to interpret auditory feedback in their
surroundings. Therefore, the main approach for visual impairment in the workshop was blindfolders, in order to force the sighted participants to rely on their hearing.

4.3 Co-design Workshop

4.3.1 The goals of the workshop

I made a mind map to have an overview of the different goals that I wanted to achieve with the co-design workshop (See figure 5). All the activities were based on the goals which were divided into two main areas: awareness and conversation. The purpose of the “awareness” area was to bring awareness to the user experience for visually impaired regarding how their current digital tools convey information. In the “conversation” area, there was a need for a deeper understanding of the conversation between the system and user with a visual impairment. In order to understand how the information could be conveyed through the perspective of vision impairment so I could identify design opportunities and ideation. These activities also investigated how the sighted participants and the established voice assistants applied the guidelines of cooperative principle by Grice from section 3.4.

Figure 5. The mind map showcases the two areas: awareness and conversation (Author’s documentation, 2019).
4.3.2 Designing the co-design workshop

4.3.2.1 Brainstorming activities

With the goals in mind, I began the ideation by brainstorming activities in three different categories which were as following (See figure 6):

- Icebreakers (orange)
- VUI (green)
- e-commerce/Junkyard (yellow)

Figure 6. The workshop activities I ideated by brainstorming on post its within the categories VUI (green), website/e-commerce (yellow), and icebreakers (orange) (Author's documentation, 2019).

I wrote the activities on post-its and divided them with specific colors, representing each category respectively. The labels of the categories served as the topic in the activities. The purpose of the icebreakers was to expose the sighted participants with playful activities with blindfolders so they could acquaint themselves with the sense of being visually impaired.

Drawing and instruct activity

The icebreaker activity was called the “drawing and instruct activity” which was inspired by the book “Let’s co-design our dying issues” from 2015, where the participants in a co-design workshop designed their own coffins. It interested me how each coffin looked unique and artistic. Therefore, I wondered how each drawing would look like if a person with vision impairment drew it. In the drawing activity, the participants were divided in groups of two and tried to draw the silhouette of a horse. One of the group members were drawing the horse blindfolded while the other group member aided the person with instructions. Each participant tried to draw and
instruct. This activity was in the “conversation” area as the purpose was to comprehend the importance of auditory detailed information as visually impaired. Furthermore, the activity prepared the participants for the other activities by letting them be acquainted to blindness.

**Evaluate and critique VUI**

I created three activities for the VUI. All the activities regarding VUI were included in the workshop. The first created activity in the “VUI” category was inspired by my former teacher Pelle Ehn from a workshop activity in my previous co-design course, in a lecture about future workshop. In Ehn’s activity, we used post-its to write down our critical thoughts that we placed on a whiteboard in front of the class, hence we had to say our written thoughts loudly after each placement. Thereafter, we discussed all the different criticism and paired the recognisable patterns between them. However, my workshop activity was not merely based in criticism but also use cases, positives, and room of improvements. The main aim of the activity was to acknowledge the participants with voice assistants by sharing their own use cases, thoughts, and experiences as sighted users of the systems. Therefore, this activity is in the “awareness” area from the workshop goals. When the participants shared their experiences and thoughts, it prepared the participants for the rest of the VUI activities as they acknowledged some capabilities and possibilities of a VUI. The workshop included a quadrant where each region had a question each based on the previously mentioned factors which were: use cases, criticism, positives, and room of improvement.

To avoid exclusion in the activity, the participants who had not used a voice assistant before wrote down potential use cases along with potential positives and negatives.

**Interact with established voice assistants in turns**

The next activity was called “interact with established voice assistants in turns”. In the activity, participants sat around a table blindfolded with both, an iPhone with Siri and an Android phone with Google Assistant on the middle of the table. To make the activity fair between the systems, both Google Assistant and Siri talked Swedish. The participants could either ask or command anything to the established voice assistants in turns. After every task done, the participants would discuss the feedback and their insights that they have gained. After a round with Google Assistant was complete, another round with Siri would begin. The activity was within the areas of “conversation”. The purpose of this activity was to encourage the participants to analyse how helpful and useful these voice assistants currently are for visually impaired users.

**Roleplay as voice assistant and blind customer**

This is the main activity of the workshop which is called “roleplay as voice assistant and blind customer” where the participants were divided in groups of two. Each group utilized a laptop with the Swedish version of Junkyard.
One group member roleplayed as a “blind” customer at Junkyard. The other group member conducted Wizard of Oz prototyping as the “voice assistant/system”, thus the stimulation of a voice assistant in Junkyard. In the paper, the participants roleplaying as blind customers will be referred as “blindfolded participants” while the participants roleplaying as voice assistants will be referred as “wizards”. The activity is under the “conversation” area as it heavily focuses on the conversation between system and user.

The activity had two purposes. The first was to detect the expectations and needs that the visually impaired customers would require from a voice assistant in Junkyard, in terms of conveying information. The second purpose was to detect potential problems and which strategies were utilized in order to solve the problems.

The voice assistant duty was to aid the visually impaired customer to navigate and find clothing of liking. In order to encouraged exploration and prevent too short trips between clothing, the blindfolded participant had to learn five specific details of the clothing item before purchase. They had to filter the searching by asking for details such as brands, clothing, color, and price, till they eventually find an item of interest. After the blindfolded participant has purchased the items, the participants exchanged roles so they could experience both roles.

**Screen reader and screen magnifier**

In the category e-commerce, four different activities were created but only one activity made it to the workshop.

The activity was called “screen reader and screen magnifier”, the participants familiarized with a screen reader and a screen magnifier, to bring awareness how the digital tools convey information to the visually impaired users in order to access digital content. Therefore, the activity was in the “awareness” area. They used the software-built screen readers as well as screen magnifiers in Microsoft Windows and Apple macOS.

They were once again be in groups and discussed the digital tools. However, this time, no participants were blindfolded due to the difficulty to navigate with a screen reader without any experience. Afterwards, they have interacted with the both tools, there was a discussion about them and the experience.

4.3.2.2 The structure of the workshop

I started to plan the structure of the workshop when all the activities were ready. However, I recruited six sighted participants for the workshop, two participants had a design background. The workshop took place in Malmö University.

An overview of the workshop planning was essential, in order to organize the activities chronologically. Therefore, I made a timeline from start to finish
with pieces of paper and began to allocate the five activities in the timeline (See figure 8).

I dedicated a piece of paper for the materials needed for the workshop (See figure 7). These materials were for instance, blindfolds, the voice assistant compatible devices, pencils, post-its and so on. I used the pink post-its so I could simply add to the workshop plan. It allowed me to determine and connect the materials for each activity.

![Figure 7. The essential materials for the workshop (Author’s documentation, 2019).](image)

As one can see in figure 8, the activities are placed next to their estimated duration times. The materials are connected to the respective activities. The total time of the workshop was estimated to be 2 hours and 30 minutes. I allocated the activities to create a sense of coherency throughout the workshop. The workshop intended to have seven main activities (excluding the introduction). However, two unmentioned activities were removed as they would not contribute to my research as much, in comparison with the other activities. The workshop structure was as following (from top to bottom):

- **Introduction** - Introducing the project and the workshop
- **Evaluate and critique**
- **Drawing and instruct**
- **Roleplay as voice assistant and blind customer**
- **Interact with established voice assistants in turns**
- **Screen reader and screen magnifier**
4.4 The findings of the Workshop

4.4.1 Evaluate and critique

The participants utilize a VUI for several different kind of tasks. The most common tasks were mainly for navigation, translation, and to turn on different products. The tasks were as following:

- Turn on music
- Google translate
- Navigation while driving or bicycling
- Listen to a book
- Set alarm
- Interacting with wireless earphones where a visual interface is missing.
- Navigation through the bank’s telephone exchange
- Start music
- Google translate, to hear how one pronounce English words

The reasons behind these tasks are that it benefits them to multitask, as they can use their hands for their primary tasks. The interaction with a VUI can be distant as well as it reduces the number of steps to conduct a task. According to the participants, a VUI simplifies the weekday.
I noticed that there was a broad of the annoyances regarding failed speech recognition (See figure 9). Several participants said that they expressed frustrations and confusion when it occurred. Failed speech recognition can severely impact the responses of a voice assistant which results in incorrect answers. One of the participants wrote down that he felt uncertainty while interacting with a voice assistant as he did not know if his input was articulated enough due to his dialect or pronunciation. There were some participants that highlighted the problem with noisy voices in the background, which make it difficult for the user and the VUI to understand each other. The voices from the background would impact the results as well because the voice assistant would receive it as input.

Figure 9: The participants placing post-its on the “criticism” region (Author’s documentation, 2019).

The participants did not place so many post-its at the improvement region of the quadrant. However, one participant mentioned the desire to sometimes turn off the information from a voice assistant as it can be long and overwhelming. Information that the participant has heard several times before. One participant wished to add a function, to raise and lower the input volume thus the user could control the background voices.

I further noticed that a few participants had a tough time to brainstorm their thoughts because they have never used a voice assistant before. See figure 10 below for an overview of the quadrant.
4.4.2 Drawing and instruct

Every participant that tried to draw the horse silhouette with no-vision valued specific information tremendously as they relied merely on their hearing for information (See figure 11). The instructors conveyed descriptive information based on letters, geometrical shapes, and horse body parts as it was easier to for their minds to imagine. It is information that the blindfolded participants have already seen hence they had a sense of how to draw it. However, such descriptive information as horse body parts is not ideal for real users whose have been visually impaired since birth, as they have never had any vision. The sizes of shapes and the speed of the drawing were frequently used as well by the instructors.

Figure 10: A closer view of all the post its in the quadrant (Author’s documentation, 2019).

Figure 11. The blind participants relied heavily on audial information while drawing (Author’s documentation, 2019).
Below one can see the different horses which were drawn in the activity (See figure 12 and figure 13).

Figure 12 and 13. Two of the horse drawings from the activity (Author’s documentation, 2019).
4.4.3 Roleplay as voice assistant and blind customer

After the previous activity, they were now acquainted with the sense of being visually impaired. Onto the next activity, the participants quickly immersed themselves in their roles while I was observing and guiding them if they had any questions (See figure 14). Each wizard began the conversation by presenting themselves and welcoming the blindfolded participant to Junkyard. They presented themselves as voice assistants and revealed their main purpose, so the blindfolded participants knew what to expect from the wizards which was appreciated. Moreover, a wizard in a group proceeded the conversation by asking if the blindfolded participant wanted to hear about sales. Once agreed, the wizard proceeded to convey information about the sales. The blindfolded participants combined different categories and subcategories together while searching in hope to get more accurate results. For instance, color, brand, and category of clothing, for example: can you find me white Adidas shoes? The combination of different categories worked as the wizards now conveyed information about the

During the entire activity, every wizard utilized a humanlike voice while having a conversation with the blindfolded participant. The wizards conveyed relevant details slowly to the blindfolded participants such as color, brand, prices and so on. They further described the clothing by providing patterns, placement of the logo, and material in order to give the blindfolded participants relevant information. These descriptions were essential for the participants to find their desired clothing items. The blindfolded participants also mix-initiated the conversations by for instance, asking the system to convey details about another clothing. The provided information by the wizards does not violate the Grice’s maxims “relation” and “quality” as it is relevant and accurate information for the blindfolded participants. Neither does the information violate the maxim “manner” as the wizards conveyed it in a manner which made sense for the blindfolded participants.

However, in the discussion session after the activity, the wizards felt that it was difficult as a voice assistant to determine what information is relevant for the users which is a matter of one’s personal preferences. A wizard said the following (Author’s field notes, 2019):

“It is difficult to know what to say as the website has so much visual information”.

The topic of proactive was mentioned where the system should learn the user’s preferences from prior shopping trips in order to convey manageable information. The risk of conveying too much information, is that it overwhelms the user which violates the maxim of “quantity” by Grice (1975).
The blindfolded participants asked the wizards to repeat in the occasions when they did not understand the output nor remember. After some rounds, the participant who wished to interrupt an established voice assistant, utilized the function with the group member. The participant said as following in the discussion (Author’s field notes, 2019):

“A big difference is when I started to interrupt the speech of the voice assistant as I knew it was going to repeat the same things again which makes it overwhelming. I already know the process, so I chose to interrupt because I already decided the item that I wanted.”

The topic of overwhelming information was brought up several times in the discussion. It has a connection with the huge amount of data that was available on the website. The wizards were not always certain how to represent the data of the clothing items. A wizard mentioned that it was easier to convey the information when the offering was less. Another wizard realized that it was easier to convey each row of clothing to the customer before giving information about the next row (See figure 15). A prompt was then given in the end of the wizard’s speech to confirm if the blindfolded participant wants more information. These prompts worked as confirmations which were frequently used by the wizards as a strategy to give more control to the participants to prevent them being overwhelm with information.
Figure 15. A screenshot of rows with clothing items presented by Junkyard (Author’s documentation, 2019).

The participants said that the conversation became more fluid and personal over time. It was easier to convey the information as the wizards began to understand what relevant information the blindfolded participants required. By every round trip, the blindfolded participants learned and utilized the different functions of the “voice assistants” to experiment. For instance, a blindfolded participant heard about Junkyard’s free freight, so the participant wanted to go back from the shopping cart and add shoes to the cart.

4.4.4 Interact with established voice assistant in turns

The participants started to interact with Siri, and Google Assistant afterwards. They asked and commanded different tasks, for instance, math problems, navigation, facts, translation, and music (See figure 16).
Figure 16. The participants interacting with Siri in turns (Author’s documentation, 2019).

The participants felt frustrated with Siri at times. As she did not read the information, instead she encouraged the user to see the result which is not possible (See figure 17). From the perspective of an individual who rely on auditory feedback, Siri violated the maxims of Grice (1975) severely.

Figure 17. The participant asked how many letters exist in the alphabet. Siri encouraged the blindfolded participant to read her provided web search (Author’s documentation, 2019).

Siri interrupted while the participants were going to speak which caused a lot of confusions. For instance, if the participants did not say anything instantly after the triggering command “Hey Siri”, Siri would abruptly interrupt by saying phrases like “I am listening” or “What do you want?”. It violated the “manner” maxim of Grice (1975) as it did not make any sense to interrupt in the middle of an utterance. A participant asked for navigation to the neighborhood Värnhem in Malmö where Siri responded that she could not provide timetables, but she could provide a route which she never did. She revealed an interaction she could do but failed to conduct it. The participant wished to proceed the conversation by saying “yes” to Siri instead of starting a whole new conversation. The constant need of triggering the assistant breaks the flow for the participant which gives an unnatural feeling. Afterwards the participant said as following (Author's field notes, 2019):

“It is difficult to determine when to say and not to say, the user should determine when the conversation with the voice assistant is over by saying stop.”
The participants got surprised when Google Assistant provided relevant information loudly when a participant asked the assistant about a ticket to Hongkong (See figure 18). Google provided information about the lowest available priced round-trip ticket between the nearest airport to Hongkong. It also asked the participant the dates of the trip. The feedback did not violate any maxims of Grice (1975).

When a participant asked Google Assistant what she is capable of, she provided a visual list of her features. She mentioned only a few features to build the expectations of the participant as well to avoid overwhelm the participant with information. However, there was a wish for a prompt by Google, so the participant could hear the rest of the list. Thus, the Google Assistant failed to initiate a prompt to the participant. The participant liked that it conveyed some functions it was capable of. I noticed how the participants asked Google Assistant more than Siri as Google Assistant provided a better and consistent user experience. It made the participants more committed to try the Google Assistant.

![Figure 18](image)

**Figure 18.** Google Assistant provided a lot of informative and relevant information regarding a ticket to Hongkong (Author's documentation, 2019).

Both voice assistants had severe problems sometimes with the speech recognition which caused frustration and confusion. The voice assistants misunderstood the participants’ utterances which provided incorrect answers. The participants commonly retried the utterances and improved their articulation by speaking louder, these were the same strategies used in the study of Myers et al. (2018). For example, when a participant asked Siri what is 68 multiplied with 68, Siri misunderstood the question as 68 multiplied with 8. The participant tried the utterance again but this time, he
articulated his utterance by speaking louder. In this try, he got the correct answer. His interaction with a voice assistant was the first time so the strategies were based on guessing which also occurred in the study of Myers et al. (2018). However, the same participant wanted Google Assistant to translate the Swedish greeting word “Hej” to Spanish which is “Hola”. However, Google Assistant misunderstood the input as “Haj” which means “shark” (See figure 19). The failed feedback frustrated him, so he decided to settle with the answer. The more experienced participants relied on prior experience to articulate the utterances, however sometimes they forgot to articulate and needed to be reminded.

![Google Assistant interpreted input incorrectly](image)

**Figure 19. Google Assistant interpreted input incorrectly (Author’s documentation, 2019).**

In the discussion afterwards, the participants said that they felt a better connection with Google Assistant because she sounded more like a human compared to Siri. Google Assistant provided more relevant and accurate answers which she read loudly to the participants. However, both systems had mutual problems with speech recognition and the unnatural feeling of constantly trigger conversations.

**4.4.5 Screen reader and screen magnifier**

To navigate with a screen magnifier in full screen mode was not a big problem for the participants (See figure 20). They already knew the layout of the website hence the navigation was easier. However, one participant tried another website and mentioned that if a website is unknown, it takes a while to get the whole context of the website.

The participants thought that the screen reader conveys information too fast which give them no time to follow the information. Therefore, it was difficult
to sort out the relevant information which overwhelmed the participants. They would consider utilizing voice assistants instead of screen readers in an e-commerce website if they would become visually impaired due to voice assistants’ capability to convey relevant information easier.

Figure 20. A screen magnifier with the full screen mode on Junkyard (Author’s documentation, 2019)

4.5 Design opportunities

After the workshop, I gathered all the insights from each activity and created an affinity diagram to map my insights and analyze them by dividing them into groups (See figure 21). According to Wilson (2013), an affinity diagram is a hierarchy of information in groups and sub-groups which designers utilize to organize their qualitative data. The organization of all the insights in related groups facilitated my process to identify design opportunities.

The affinity diagram highlights what the participants desired with the established voice assistants and the wizards. They desired the VUI to convey information with human characteristics including humanlike voice and intelligence. They also desired the system to convey relevant information such as details about clothing or sales information. The affinity diagram also highlights the overflow of information and failed speech recognition which were the participants’ main problems in the workshop. Several times during the workshop, I noticed how these problems caused frustration and confusion which made the user experience worse.

Regarding the failed speech recognition, the participants had to figure out by themselves that they had to rely on strategies such as retry and articulate the utterances louder to ensure that the established voice assistants understood their input. This creates a design opportunity towards the system where it can aid the visually impaired user with the speech recognition by conveying auditory feedback. The design opportunity creates the first question:
1. How can the system communicate feedback to the visually impaired user to prevent failed speech recognition?

The participants mentioned that it was difficult to convey the relevant information due to the huge amount of visual data available in Junkyard and individual preferences. The participants and the wizards collaborated with each other to avoid overwhelming the user with information. Therefore, the blindfolded participants desired control to cope with the overwhelming information. The participants and wizards ideated themselves by giving prompts and commands to each other to increase the control for the participants. For instance, the wizards conveyed prompts to the blindfolded participants to confirm proceeding into further details about the clothes while the wizards utilized a humanlike intelligence to remember details from previous trips. The collaboration between system and user creates another design opportunity hence the question.

2. How can the system and the visually impaired user collaborate to select relevant information without overwhelming the user?
4.6 Brainstorming

With the insights from the workshop organized and the questions from design opportunities identified, I conducted a brainstorming session to generate ideas to prototype. As one can see in the figure 22, I added the ideas below each question. I was inspired by Wilson (2013) to focus on quantity rather than quality in order to encourage creative ideas. Therefore, I generated several ideas on each question instead of only one idea. The brainstorming session also allowed me to be creative and explore upon the ideation from the workshop. regarding how the system can convey information to the aid users to prevent failed speech recognition and overwhelming information. As this paper’s topic regards the technology of VUI and auditory interpretation, all the ideas were based on how the VUI could convey information by auditory feedback.

![Image](image.jpg)

Figure 22. The brainstorming session (Author’s documentation, 2019).

I picked out one idea from each question that could be developed into prototypes in an additional design process. To further understand how a VUI could be designed to convey information and deliver a good user experience, I created visual flows with simple dialogues so called callflows. Cathy Pearl (2016) explains callflows as diagrams that illustrate the paths that users can take in a VUI. The callflows will not showcase every interaction the user can do. They only display examples on interactions that are appropriate regarding the design opportunities. I ensured that the ideas did not violate the maxims of Grice (1975) nor the RRIMS properties by Radlinski and Craswell (2017).

The first callflow represents the idea regarding failed speech recognition which can be seen in the figure 23. The system repeats the inquiries made by
the visually impaired users and initiates a prompt for confirmation, the idea was based on a participant in the section 4.4.1 who felt uncertainty whether the system recognized his utterances. As the participants utilized confirmations a lot in the workshop, I felt it was necessary to include them in the ideation. According to Pearl (2016), these confirmations are examples of explicit confirmation where the system forces the user to confirm the information. Pearl (2016) further explains that explicit confirmations are important if the risk of failed speech recognition is high hence it is even more appropriate to include them in the idea. Through the repeated utterances, the user would hear if the system comprehended the entire inquiry correctly and then explicitly confirm it as correct. If the interpretation was flawed, the user will hear where in the sentence it was flawed and improve the articulation upon those words. The system could encourage the user to retry and inform the user solutions to the problem such as articulate the utterance. It would be helpful for users who are unfamiliar with the VUI to learn the solutions but also to remind regular users as well.

Figure 23. The callflow of the idea regarding the failed speech recognition (Author’s documentation, 2019).

The second callflow represents the idea that regards overwhelming information which was inspired from the concept of Olsson (2017) and the ideation from the workshop (See figure 24). The callflow begins when the system reveals itself.
Like the concept of Olsson (2017), the system conveys the four clothing items in the first row, as one can see in figure 15. When it is done, it will require an explicit confirmation from the user, in order to proceed by conveying the clothing items in the second row. I got inspired by the discussion and ideation regarding retrospect in the workshop so I contemplated how the system can further offer aid to the user with its AI. To let users set their own preferences regarding the amount of information they want to hear. These preferences could be saved for future shopping trips till the users change their preferences. Explicit confirmation and to set individual preferences would potentially help the users to cope with the information in their own pace. This idea can be further explored where the system conveys large amounts of information e.g. current sales.

Figure 24. The callflow of the idea regarding overwhelming information (Author’s documentation, 2019).

As I mentioned in the beginning of this chapter, due to time restrictions and no visually impaired participants, the design process of this paper remains unfinished with no prototypes based on the ideation. The next step forward for this work is explained in chapter 6.
5 Discussion

The purpose of this thesis was to explore how a VUI could be designed in an e-commerce website to convey its information from a visual impaired perspective. In this chapter, I will discuss the insights from the co-design workshop with the relevant theory and work. I will also mention what other sighted designers need to think about in order to deliver a good user experience for visually impaired users.

The empirical research showed that a VUI with humanlike characteristics is vital for designing a VUI in e-commerce settings and overall. The wizards in the “Roleplay as voice assistant and blind customer” activity utilized a humanlike voice as it was the best approach to convey information to the blindfolded participants in a collaborative manner. Burgoon et al. (2002) mentioned that a humanlike VUI is best suited for collaborations which make users more active while the system aids with users’ decision making. The collaboration between the blindfolded participants and wizards was vividly apparent in the e-commerce website. The participants were indeed very active while the wizards helped them to decide which clothes to buy. Overtime, the conversations became more personal and the wizards had a better understanding which information the participants wanted as well as clothes. This indicates that they were building a relationship with each other as Burgoon et al. (2002) stated a humanlike VUI would do. To achieve the sense of collaboration, the workshop proved that the maxims of Grice (1975) must be followed which is equivalent to the property of “user revealment” by Radlinski and Craswell (2017). The established voice assistants were a good example of this. The participants preferred Google Assistant than Siri as Google Assistant were more consistent in conveying relevant information with a more humanlike voice where Siri performed worse. The better human characteristics of Google Assistant made the participants feel more connected and provided a better user experience. This is similar to Qvarfordt et al. (2003) argument, where they argued that users prefer a humanlike system once they have experience it. The need of humanlike characteristics of a system is not only limited to sighted users but visually impaired users as well. Visually impaired people in the research of Olsson (2017) expressed that they prefer a humanlike voice that is pleasant and provides relevant information. Therefore, sighted designers should use human characteristics to deliver a good experience for visually impaired.

The problem of failed speech recognition occurred several times during the participants interactions with the established voice assistants, where the systems misunderstood the participants hence the failed feedback. It caused a lot of frustration and confusion for the participants because it did not convey information that they needed to achieve their goals. Failed feedback also occurred in the study of Myers et al. (2018). However, according to Myers
et al. (2018), failed speech recognition was not a big threat towards the user experience as their participants already knew that improving their articulation was the solution to the problem. Some participants in the co-design workshop also knew the solution due to prior experience with voice assistants but they still forgot to articulate. Myers et al. (2018) defined failed feedback as an obstacle that was a bigger threat against user experience but did not explicitly mention that failed speech recognition could be a reason to failed feedback. Therefore, failed speech recognition should not be taken lightly as it impacts user experience negatively by resulting in failed feedback. Failed speech recognition should be considered while designing a VUI for visually impaired users. To cope with the high risk of failed speech recognition, the sighted designers need to take advantage of explicit confirmation which Pearl (2016) also argued for, in section 4.6. Explicit confirmations could help the users to understand where in the utterance it was flawed.

The workshop also proved that relevant information conveyed by a VUI in an e-commerce website is very important for users who rely merely on hearing. Giudice and Legge (2008) makes the same argument where they say that an effective tool for navigation needs to convey selective information that is specific according to the current environment. The environment in this paper is the digital environment of Junkyard where the blindfolded participants asked e.g. information about the different clothing in the current row or freight information. Through the wizards, they wanted to achieve relevant information without it they would feel lost in the digital space. The blindfolded participants took advantage of the “set retrieval” property by Radlinski and Craswell (2017) where they searched for several categories at once. For instance, their utterances included preferences such as the category of clothing, price, and brand in the same search so the wizard could convey accurate information. The wizards in the workshop began the shopping trip by conveying what they could do which built up the blindfolded participants’ expectations. This is the property of “system revealment” by Radlinski and Craswell (2017). The revealment helped the blindfolded participants to start and proceed their shopping trip.

The risks of conveying too much information is overwhelming the visually impaired user with information. Kristjánsson et al. (2016) calls the problem in another synonym as sensory overload, as it overloads a sense with information. According to Kristjánsson et al. (2016), overwhelming information must be avoided in a VUI. This was proven during several activities in the workshop where the blindfolded participants expressed that they got frustrated by being overwhelmed by information, as it was time consuming and repetitive. In the “Roleplay as voice assistant and blind customer” the wizard felt that it was difficult to convey information to the blindfolded participants as each participant has their own personal preferences in information and clothing. To convey the information without
overwhelming the user cope, the participants wanted more control mostly in the shape of the explicit confirmations. Explicit confirmations can also be described as a part of the “mixed initiative” property by Radlinski and Craswell (2017) as the system takes the initiative to prompt the user for confirmation. The explicit confirmations also keep the conversations fluently without the need of constant triggering which was a desire by the participants to keep the VUI humanlike. The users would have control of the conversation, from the beginning to the end of the conversation. The other approach which helped to user to control the overwhelming information in the workshop was the benefits of a system’s AI in terms of memory. The participants discussed how the system can retrospectively help the users by utilizing its AI, to remember and convey the shopping clothing or address information to avoid repetitive and time-consuming information. The memory aspect of the system is equivalent to the “memory” property by Radlinski and Craswell (2017) as it utilized information or statements from users in prior shopping trips unless they are contradicted.

6 Self-Critique and Future Work

While recruiting visually impaired people, I should have planned a back-up plan in the meantime as it was uncertain if I would be able to recruit any visually impaired. I lost valuable time when I could have begun the planning with the workshop in an earlier stage hence, I would have time to prepared it better. For instance, I could have encouraged the participants to utilize a voice assistant before the workshop. This might have improved or increased the results in the “Evaluate and critique VUI” activity as everyone would have fresh experiences with a voice assistant.

The next step of this work would be an additional design process where sighted designers can create low-fi prototypes based on the callflows and test them with visually impaired. The callflows are for designers who are willing to design VUI for visually impaired to solve the societal problem. Giudice and Legge (2008) mentioned, empirical research is required to ensure that the VUI is user-friendly and conveys enough information for the visually impaired. The VUI might not satisfy the visually impaired users’ needs in practice. Therefore, at this state of this project, it is not possible to determine if a VUI would solve the societal problem. However, as it relies heavily on collaborative conversations and conveying information based on auditory feedback, it has shown to have potential to improve the accessibility towards a social change for visually impaired people in e-commerce websites. Furthermore, the implementation of VUI could also be investigated in different settings where VUIs have the potential to provide accessibility for visually impaired users.
7 Conclusion

The visually impaired people do not have the same access to e-commerce websites as regular users due to the limitations of their digital tools such as screen readers. This creates a societal issue where visually impaired are alienated from the content of e-commerce websites. I identified a societal need for a VUI in e-commerce websites to improve the accessibility to the information hence I felt a responsibility as an interaction designer to explore how a VUI could convey the content.

I conducted a design process which was never completed as I had difficulties to find visually impaired people due to their small population. Therefore, iterative prototypes as well as fieldwork with the intended target audience is lacking in the paper along with the final solution. My main research question was: “How could sighted individuals co-design a voice user interface in an e-commerce website to convey its information through the perspective of visual impairment?”. Followed by the sub-question: “How may sighted designers deliver a good user experience for visually impaired users?”. To answer these questions and to proceed the design process, I designed and conducted a controlled co-design workshop with several activities based on interaction design methods such as Wizard of Oz and experience prototyping. The six sighted participants in the workshop were forced to have blindfolders so they could pretend being visually impaired customers in an e-commerce website.

Based on the insights from the workshop, the participants expressed several times that failed speech recognition and conveying too much information caused frustration and confusion which affected the user experience negatively. The participants desired a VUI to convey relevant information with humanlike characteristics including voice and intelligence. To cope with the problems, the VUI conveyed explicit confirmations and statements from prior shopping trips in order to provide control to the users, which the participants also desired. A humanlike system is the best approach for collaborations which provides a better user experience. To achieve a humanlike system, sighted designers need to follow the Grice’s maxims and achieve the RRIMS properties.

In this paper, two design opportunities regarding failed speech recognition and overwhelming information were identified. I created two callflows based on the ideation to showcase for designers how these ideas can be prototyped. These prototypes should be tested with visually impaired users in an additional design process. The empirical research would be essential to determine how a VUI in an e-commerce website could fulfill the needs of visually impaired as a design for social innovation regarding the societal problem.
8 References


