Interface in Our Ears

Joakim Vedmar

August 25, 2017.

LUND UNIVERSITY

Master's Thesis

Faculty of Engineering LTH at the Department of Design Sciences.

Supervisors: Kirsten Rassmus-Gröhn

Examiner: Mattias Wallergård

Abstract

One of the current trends amongst the major software companies is Virtual Personal Assistants (VPA), such as Apple Siri, Google Assistant and Amazon Alexa. This market is growing rapidly, and the assistants are getting smarter and becoming more capable. However, in this kind of rapid expansion it is easy to lose focus on the user, which may lead to the end user discarding the VPA as another software gimmick.

In this Master's Thesis I focus on the interface in our ears by using various methods in interaction design and user experience in conjunction with VPAs to investigate new ways of developing the digital assistant.

One way to develop VPAs could be to introduce stacked commands so that VPAs can collect and perform several commands at once, giving users more information whilst still being efficient. The feature of stacked commands is where VPAs can perform multiple actions with just one command by performing each command option one after another or by "stacking". This feature of stack commands can give users quick status reports or seemingly perform multiple tasks simultaneously. This thesis illustrates one way to utilise the stack command in the field of home systems.

An early concept prototype was developed and tested how this feature could be displayed, giving visual and audio feedback to the user. The prototype received positive responses during the testing phase and showed that participants preferred this new kind of interaction with the VPA.

The concept of stacked commands can be extended to, but is not limited by, homes, cars and workplaces. A market driven development could further this concept to any eyes-and/or-hands-busy environment where voice activated VPAs would be deemed necessary.

Keywords: Virtual Personal Assistants, User Centred Design.

Acknowledgements

This Master's Thesis has been performed at The Faculty of Engineering LTH in the Department of Design Sciences at Lund University, Lund, Sweden during 2017.

I would like to direct a big thank you to several people without whom I would not have been able to do this project.

Kirsten Rassmus-Gröhn, LTH
Thank you for the invaluable advice and support you've given me throughout this project.

Danica Welch
Thank you for being my biggest support during this project.

Friends and family

Finally, I give thanks to all the participants in surveys, interviews and tests for contributing valuable information and donating their time to help.

Contents

Chapter 1
Introduction

This chapter contains the introduction to the subject of this Master's Thesis. The introduction will serve to give the reader a glimpse into the world of virtual personal assistants (VPAs) and reveal the goals, purpose and problem definitions for this project.

1.1 Background

At the 1962 World's Fair in Seattle, the world's first VPA was presented. It was developed by William C. Dersch at IBM [5] and this device could recognise and respond to 16 different spoken words and calculate simple arithmetic. It was operated by using a microphone that converted the sound into electrical impulses which activated the algorithms in the machine. (Fig 1.1)

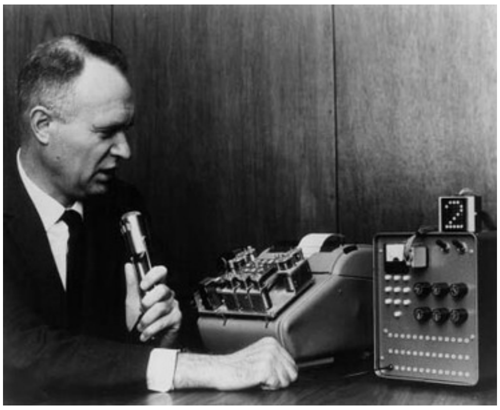

Figure 1.1: William C. Dersch showcasing the IBM Shoebox

In the 1970s the next evolution of speech recognition was introduced which was the Harpy Speech Recognition System [7]. This system could handle a much larger word-base and was created by Carnegie Mellon University, and obtained its funding from the Department of Defense in the USA. With the Harpy System more complex speech recognition software evolved which improved the speed of such systems.

The more modern speech recognition software derived from statistical models, such as the hidden Markov model [2], and with this the ability to recognise continuous speech started to develop. In 1990 the company Dragon managed to demonstrate a 5000-word continuous speech system for the general public. The speech recognition software paved the way for VPAs and both became interconnected.

Since then the market has grown and matured with several attempts made at creating a usable assistant that is also smart. Smartphones started a new revolution in speech recognition systems. When Apple released Siri in 2011 the standard model for modern VPA's was created, which had a more sophisticated design and was widely available to the everyday consumer. With the evolution of the smartphone there has been a rise of more assistants released into the market, such as Google Assistant, Cortana and the Amazon Alexa.

A VPA works primarily by using your voice to talk to the control system but also allows text input. An assistant can respond and talk back to the user by confirming commands and performing various tasks. In standard VPAs a key phrase is used to start the process such as; *Hey Siri; OK Google;* or *Alexa*, but these VPAs can be accessed in other ways, including holding down the home button down on the device until the assistant activates.

In 2016 CNET [1] estimated 100 different commands available on Google Assistant, but as Google does not provide a complete guide to current commands, more commands could be available.

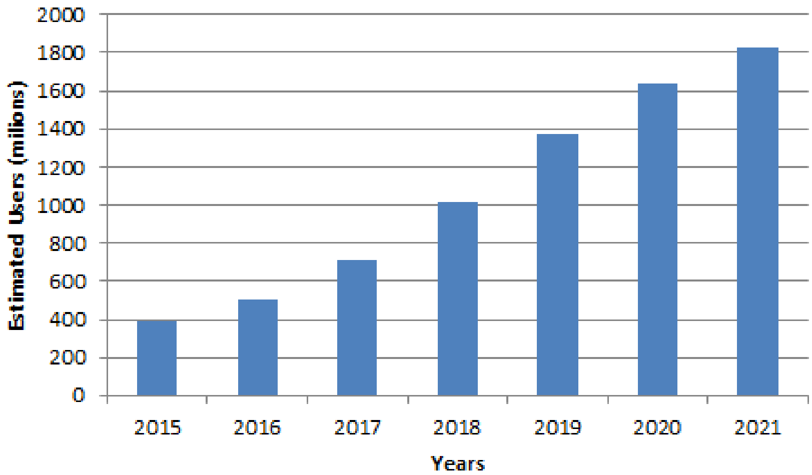

Figure 1.2: Displaying estimated users by year

A survey [3] performed in 2016 estimated that in 2017 the users of VPAs would be an estimated 710 million and growing to a possible 1.831 billion by the year 2021. (Fig 1.2)

With the rise of Internet of Things (IoT), the interconnection of computing devices via the Internet, the assistants will be more connected than ever. Non-graphic interfaces are rapidly growing and with the Voice User Interface (VUI) interaction is possible through a speech system and its VPAs. With IoT the ability of connecting more services to VPAs is possible, such as controlling the robot vacuum cleaner or other household appliances, serving to broaden the options for VPAs creating a well-rounded system.

Still the modern personal assistant is still a young market and has much to learn. The assistants are continually obtaining new and more advanced features, with better speech recognition and the support for new languages is already being implemented. The natural language speech system is the computer's ability to respond to a user in a more natural and less robotic way, which leads to a more fluid interaction with the VPA. Combining this with connectivity, VPAs can handle more advanced and complex voice control actions, both understanding the user and performing the tasks the user wants the VPA to complete.

1.2 Purpose, Goals and Problem Definitions

The purpose of this project is to investigate a new way of using Voice User Interface (VUI) to improve VPAs interaction between user and machine by implementing tools and methods applied in user experience and interaction design. With the rapid expansion of VPAs, it is imperative that they continue to be of a high-quality user experience and strive to improve in the areas that are currently lacking. This thesis is to determine if the VUI can improve usability. The purpose of usability is to achieve specified goals with efficiency and satisfaction for the user.

This project aims for the betterment of VUI and to implement new methods in using VPAs with the focus on User Centred Design (UCD) [21], which is a design process developed to gain understanding of who will be using the product.

The goals for this project are:

- Chart the current habits of users that are using VPAs.
- Suggest improvements for the current VPAs.
- Suggest improvements for the interaction between user and machine with the help of VUI.

To fulfil the criteria, a list of problem definitions needs to be formulated and then examined to create a research basis for the project. Within the thesis the problem definitions will be investigated and then analysed with a conclusive summary.

The problem definitions for this project are:

- How can a VUI help improve the interaction and accessibility between user and machine?

- What does the future look like for VUI?
- Is it possible to further develop the current VPAs?

This Master's Thesis will focus on smart-devices and attempt to address the research questions that will form the structural skeleton of the thesis with interaction design and usability in mind.

Chapter 2
Theoretical Background

This chapter contains the theoretical background of this thesis. In this section the relevant theoretical facts are displayed. This helps explain why certain steps have been performed and serves as a guide throughout the thesis.

If I have seen further it is by standing on the shoulders of Giants.

- Sir Isaac Newton [10], *Letter to Robert Hooke*

The theoretical background will serve to form the basis for the entire project. The theoretical background displayed here has a tried and accepted formula [21] that gives solid direction when designing a user interface.

The background is divided into parts which explains different areas which are relevant to this project. The first part refers to User Interface (UI) which explains the possibilities of how humans and computers can interact with each other. The second part is how User Centred Design (UCD) is important in creating a user interface. By adhering to the usability principles, one can create a user interface that is both efficient and satisfactory to the user. The third part explains how to design an audio-based user interface and visualise the audio system to be able to build a prototype. The final part explains how to survey, test and evaluate such prototypes.

2.1 User Interfaces

A user interface is the space in which humans and computers interact with each other [15]. It is through the user interface that a user can interact with a machine system. This type of interface began as a command line based system, which did not adhere to the principles of usability, and has now developed into the Graphical User Interface (GUI). A GUI is a useful way to show the user what is happening on the machine, but GUI also has disadvantages as you must physically see the screen to be able to interact. This is proved to be a drawback for people with visual impairments and is not the best way to cater to all users in all situations.

2.1.1 Voice User Interface

The Voice User Interface (VUI) is a user interface which is based on voice, meaning people interact with this system by speaking [11]. A VUI has the greatest potential to enable people that are unable to move their hands or physically see the screen. These situations include certain disabilities and occupations in eyes-busy/hands-busy situations, such as drivers. GUI is perceived as a primary interaction mode, placing the voice interface as its secondary interaction mode, which unfortunately has resulted in voice interface properties not being fully developed.

The problem with VUI not being a two-dimensional medium, such as a screen, which enables the combination of persistence (ability to view for an extended period) and selective updating (ability to type a value into a field anywhere on the screen without changing the rest) is a drawback in this one-dimensional user interface.

Donald Norman brings up a known problem in interaction, that natural user interfaces are not natural [12]. He explains that each interface must have a set of basic rules with

> *"well-defined modes of expression, a clear conceptual model of the way they interact with the system, their consequences, and means of navigating unintended consequences."*

- Donald Norman [13], *Natural Interfaces Are Not Natural*

One of the problems with a VUI is that it is an invisible user interface, unless paired with a GUI. A GUI is easy to explore compared with a VUI where the user needs prior knowledge of the voice commands. These voice commands rely on the user's ability to memorise all commands which can sometimes be hard for the user, it can also become frustrating when the system misinterprets the user's commands.

2.2 User Centred Design

The User Centred Design process [21] outlines each phase throughout the design and development life-cycle, whilst maintaining focus on the end user.

The UCD process is generally divided into these four phases (Fig 2.1):

- **Specify the context of use**
  Identify the users and what they intend to use it for.
- **Specify requirements**
  Identify the goals that must be met for the product to be successful.

- **Create design solution**
  Produce solutions to address the requirements and create a usable design.

- **Evaluate designs**
  Perform usability tests with real people to see if the requirements and needs have been addressed.

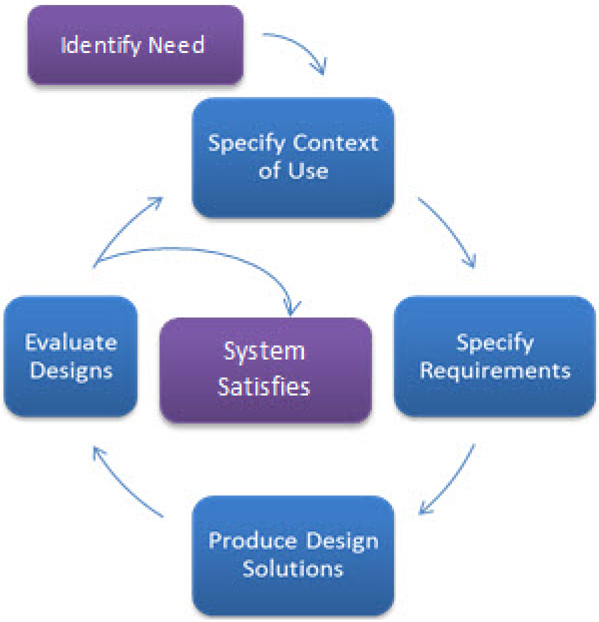

Figure 2.1: The User Centred Design Process

To create a well-rounded prototype a theoretical base is needed to expand on the ideas of design. Two pioneers in human-computer interaction are Ben Schneiderman and Donald Norman. Both individually published books in 1987 [17] and 1988 [14] which covers the basics of interaction design. To develop a new prototype, it is recommended to use certain rules and principles that define the needs of a usable system. It is essential to designing and analysing a prototype that the VPA follows these rules and principles.

2.2.1 Eight Golden Rules by Schneiderman

When designing a user interface there are several recognised standards that have been tried and tested over several decades. One of these is Schneiderman's Eight Golden Rules [17] of interface design:

1. **Strive for consistency**
   Strive to have some consistency over the entire user interface. Actions that resemble each other should have similar reactions, as well as different sounds and feedback should remain consistent to ensure that the user does not feel lost or confused in the system. Exceptions should be comprehensible and limited in number to avoid frustration for the user.

2. **Cater to universal usability**
   Recognise that the system will have diverse users with different technical backgrounds and the user interface should cater to the needs of the many. This would include features such as additional feedback for novice users, as well as additional selections to expedite the process for expert users.

3. **Offer informative feedback**
   For every user action there should be some form of informative feedback to confirm that an action has been recorded. For different levels of actions, the response should vary, such as frequent actions should have a modest response and infrequent major actions would warrant a more substantial response.

4. **Design dialogs to yield closure**
   The user interface should be organised in the way of a story. There should be a beginning, middle and an end. Progressive steps lead the user to feel a sense of accomplishment in succeeding a task and move on from the task at hand.

5. **Prevent errors**
   Design the user interface in a way that prevents the user from making mistakes. The system should be designed so that if a user makes any mistakes, the system will detect and guide the user to repair the mistake instead of returning to the beginning.

6. **Permit easy reversal of actions**
   Allow the user to undo actions that they have made to relieve their anxiety. This is pivotal in making sure that the user explores unfamiliar options in the system.

7. **Support internal locus of control**
   The user should have support from the user interface so that they believe they are in control of the system. The system should respond to the actions of the user and not the other way around.

8. **Reduce short-term memory load**
   The user interface should be designed in a way to alleviate short-term memory issues due to the limitations of human information processing. This makes it imperative that the user interface is designed in a way that the displays are kept simple and the presented information is kept to a minimum.

These rules need to be applied to simplify the user's interaction with the system by making sure that the user can traverse through the system with ease.

2.2.2 Design Principles by Norman

Another of the recognised design standards is the principles by Donald Norman in his "Design Principles to Make Everyday Things Usable" [14]. Norman reiterates the need for a product to be usable and shows several examples that are completely dissatisfactory to the user because there is no recognised way to interact with the product. It shows that even with the best of intentions with a product if it is not satisfactory to the user it will not be used. To combat this problem Norman created a set of design principles to make everyday things usable with a positive experience.

- **Affordance**
  Affordance depicts how an object should be designed giving a visual clue to its use and function. An example would be the use of a button which is raised from a flat surface to indicate that it can be pressed down.

- **Mapping**
  It suggests a relationship between two things, such as an action triggering a resulting reaction. Example would be the manipulation of the steering wheel which changes the direction of the car.

- **Constraints**
  By restricting interactions, the system is easier to use by reducing the amount of errors that may occur. There are a number of different methods for restricting actions; these include logical, physical and cultural constraints.

- **Visibility**
  Users need to know what can be done in a system. This is best signified by visual feedback. By making functions more visible the user will be able to interpret what is accomplishable.

- **Feedback**
  The acknowledgement that an action has been performed is done by feedback; this demonstrates the accomplishment of said action. This allows the user to recognise that a change has happened and can proceed with the activity.

- **Consistency**
  This refers to having operations look and perform similarly every time an action is performed. Designing a system that has rules which must be obeyed and therefore an action has the same reaction every time.

By incorporating these principles there is a better chance that a product proves to be useful. Upholding these principles, the VPA can become the everyday thing that is both useful and usable. By combining both the rules and the principles it will together complement each strategy in making the VPA more usable.

2.3 Designing an Audio Interface

Sound provides information in different ways [14] and it can tell us when things are working and when they are not. The use of sound as an indicator is valuable and serves a beneficial function. A successful sound feedback creates an understanding between the sound and the information that is being conveyed. A sound feedback can convey information for example, when a phone call is beginning to be connected, if it does not give you the signals that you would think then this may be an indication that something has gone wrong with the connection.

Sound is a good feedback source, but it can also be annoying and sometimes become distracting when not implemented properly in the interface. Sounds are also difficult to be kept private, as well as annoying to the people around you, unless the volume is set low or headphones are used.

2.3.1 Auditory Display

The definition of auditory display [4] is the use of sound to communicate between a computer and a user. The goal of this is to enable understanding of the changes that occur in a computer system. Sonification is a core component of auditory display and is the act of conveying information and data to the user via non-speech audio medium.

There are a number of functions for which auditory display should be able to perform and this information display can be showcased in these three broad categories:

1. **Alarms, alerts and warnings**
   This refers to the sound that indicates that an action has or is about to occur. Example of this could be a message signal which indicates that the user received a message but does not present the details of the message.

2. **Status, process and monitoring messages**
   The current status of the system should be presented to the user by changing the sound structure when an action occurs.

3. **Data exploration**
   The system should be able to convey information about an entire system or at least part thereof.

2.3.2 Voice Control and Speech Recognition

Obstacles to speech recognition [17]

- Increased cognitive load compared to pointing.
- Interference from noisy environments.
- Unstable recognition across changing users, environments, and time.

There are a number of different variations of technologies pertaining to speech recognition, which is described as follows:

- **Discrete-word recognition**
  The technology of discrete-word recognition recognises individual words spoken by a specific person and has in general 90 to 98% reliability for 100 to 10 000 or more word vocabularies. A method of the system learning the variations of speech of the user is to let the specific user say the whole vocabulary to yield a higher accuracy. The discrete-words method works well during special circumstances but unfortunately is not a good general interaction medium.

- **Continuous-speech recognition**
  This is the method which many sci-fi movies have adopted as the standard for a speech recognition platform. Instead of interpreting all interaction word-for-word it is able to understand speech as used when people converse with each other. The difficulty for this method is recognising the boundaries between the spoken words and is something that still needs improving.

- **Non-speech auditory interfaces**
  Other options of technology are the auditory output which can be combinations of sound and music that combine to make the user aware of system changes. This is often referred to as auditory display and sonification and is one of the underlying information presentations.

There are two terms that are often included in sound interface which are auditory icons and earcons. These are two sides of the same coin. An auditory icon is generalised as familiar sounds, examples being door opening, ball bouncing, and paper being balled up. By giving a familiar sound effect, it helps reinforce the visual metaphors in a system. An earcon is an abstract sound effect which must be learned for the user to understand. This can be likened to an audio cue that gives the user an indication that a message has been received.

2.4 Design Process

To show that the project will be able to develop the current VPAs, a prototype will be created. To create a prototype there are recognised methods that helps the designer to achieve their intended goals. In this project a set of tools such as surveys, interviews and researching current VPAs created a platform for the gathering of data needed to complete the prototype. There are different levels of prototypes that can be built, with each having its positives and negatives.

2.4.1 Questionnaires

By gathering data from the users, it helps to create an understanding of people's perspectives. By performing surveys early in the design process, it can help understand the habits and the usage of VPAs.

Questionnaires are a tool in conducting surveys [9] and can be narrowed to a series of questions with the intent of finding information about the user from their perspective. It is useful for trying to understand the user for whom the product is designed. The possibility of conducting the questionnaire online proves to be easy for the test subjects as they can choose to perform the study within their own time frame and does not require a meeting with the stakeholders. The questionnaire format can include:

- "Yes" and "No" checkboxes.
- Multiple choice checkboxes.

- The Likert rating scale.
- Semantic scale (that can range from example: very good to very bad).
- Open ended questions.

The questionnaires focus on the research's objectives and hypothesis and can provide the quantitative data for analysis and presentation. With the quantitative data it is possible to present statistical data in a relevant field.

2.4.2 Interviews

A face- to-face interview is a survey method [20] designed to obtain larger quantities and more in-depth information with the intention of probing the respondents on their answers. With this tool it is possible to include more open-ended questions and ask the respondent to explain their reasoning behind their answers.

There is a difference in comfort level for the respondents in comparison with a questionnaire, that makes it possible for the interviewer to clarify the questions and rephrase for the respondent to understand the question clearly. There are several disadvantages for interviews as it is a time-consuming activity and requires resources, such as appropriate locations and possible travel time, to be able to conduct the interviews

2.4.3 Prototyping

By gathering data to identify the needs of the end user [19] it is easier to obtain peoples perspectives. Methods including questionnaires, interviews and test studies can help to obtain different perspectives, with a varied range from a general input, such as questionnaires, to a more specific input generated from a test study. With the needs identified a requirement specification can be created that lists the needs to be met for the final product. By identifying requirements early, it helps to recognise avoid errors being made in later iterations. This requirement specification does not need to be in depth but only to help to visualise what abilities the final product needs.

Emerging from the requirement specification a conceptual model, a representation of the system, can be outlined for how people can interact with the product. The conceptual model highlights what features can be used and how the user interacts with the system.

To build a product it is wise to start with the building of prototypes so that way you can explore your ideas and show your intentions for features early in the process. This can be done with many different levels of prototyping; anything from a paper drawing (Low Fidelity) to searching through a series of processes on a fully developed site (High Fidelity). There are benefits to prototyping but ultimately these culminate to a major consideration - cost. It is considerably less expensive to make changes in the design early in the development process, but this also allows for feedback from selected users and the ability to discover design errors before production.

As previously stated there are many different prototypes ranging from Lo-Fi to Hi-Fi which are based on Graphical User Interface (GUI), which allows for visual interaction between user and machine.

A Lo-Fi prototype is a sketch based prototype allowing for none or minimal user interaction. This kind of prototype is done on paper it's easy to create and is helpful in showing the design in a way that invites users to be more comfortable in suggesting changes. Using a Lo-Fi prototype model tends to be a simple and inexpensive and is often used in the early stages when exploring the ideas for new features.

In contrast a Hi-Fi prototype is a high-tech representation of the design that allows the users to see what the design will look like as a finished product. These kinds of prototypes would be presented more in final stages of interface development and is more suitable for collecting performance data and displaying or demonstrating actual products to stakeholders.

Each prototype comes with a set of compromises that leads the designer to focus on the key issues of what the prototype should produce. By doing this at different stages a designer can explore systems in for which a Lo-Fi could be used, which has a more exploratory setting, or Hi-Fi which has a narrower approach. Two of the most common compromises in development which are produced from these limitations are either horizontal prototyping (entire interface with no underlying functionality) or vertical prototyping (highly detailed system but for only a few features).

2.4.4 Usability Testing

According to Rubin and Chisnell [16] the ability to ensure that the creation of products is useful is by performing usability tests. Usability is an integral quality for making a product easy to learn, effective and efficient. To make sure a product is satisfying to the user there are reliable methods for assessing the product.

In this thesis the main type of testing is called an assessment test, which is one of the popular ways to test a prototype. The assessment test is usually conducted midway through a product development cycle after early design concepts have been established. The purpose of this kind of test is to examine and evaluate the main features of the design by creating tasks that test relevant parts of the prototype; the goal for the test moderator is to find any errors or deficiencies in the product. The test is created to evaluate how effectively the concept has been implemented.

If the users can interact with the system and can perform tasks it is possible to find relevant answers or discrepancies in the system. By interacting with the system, the test moderator can find if the interface is confusing. Is the graphics and terminology clear and comprehensible? Do the users get stuck at any point? If so then try and determine the cause for such an event.

The findings will be based on qualitative and quantitative results from the tests. To ensure the proper steps are taken, a test plan will be created to avoid mistakes and to standardise each task to provide the same circumstances for each participant.

Chapter 3

Methodology

This chapter contains methodology section of this thesis. In the methodology section, relevant stages in the thesis are explained and it allows the reader a full view of the project.

The methodology section serves the purpose of combining the relevant theoretical background with the designing and testing of the prototype in this project.

The design process of this thesis was conducted in an iterative and incremental manner. This process can be categorised in three stages which are background research, prototyping and usability testing.

Firstly, background research explains the different parts of how information was gathered to research the relevant parts of VUI and a VPA. The next stage was prototyping which explains the process of how the prototype was designed and developed. The use of prototypes was a way of creating tools to test and analyse ideas and by visualising concepts to see if it were usable. The final stage, the usability testing where the methods on how the prototype will be tested and what kind of testing will be used is explained.

3.1 Background Research

In this Master's Thesis a large portion of time was dedicated to the gathering of information to better understand the fundamentals needed to complete the project. The research occurred in stages where firstly different sources of information were collected in the studies, surveys, interviews and was used to give broad but precise data into VUI and VPAs.

3.1.1 Literature Study

A study in literature was conducted and relevant information in interaction design was gathered from library database searches which resulted in many sources relevant to interaction design. This provided a good base to work within and, with the help of various internet resources, resulted in a stable and factual background to which this thesis was founded.

A good start to the project was the book on “Interaction Design: Beyond Human-Computer Interaction by Helen Sharp” [19] as it had good scientific coverage over the whole field of interaction design and many references that lead to other sources of information. Books such as Design Principles to Make Everyday Things Usable by Donald Norman [14] and Designing the User Interface by Ben Schneiderman [17] provided a solid background into usability as they had relevant information on how to design a prototype of good usability. By researching various aspects of interaction design and other relevant material lead to such pioneers such as Gould and Lewis’ and their Three Principles in user-centred approach which helped to influence the project but was ultimately not used in the thesis.

3.1.2 Survey

A survey was performed, and the template for which can be found in Appendix A, where the participants were asked about their current usage, likes and dislikes of current VPAs. This survey was created in Google Forms which was useful to map out the current usage in today’s virtual assistants. Google Forms had the ability of collaborating information and create a visual display of that information in a chart form to receive a prompt overview of all the sections within the survey.

A broad array of users was selected for the survey to identify all possible user issues within the prototype. An effort was also made in acquiring both Swedish and English speakers to eliminate language barriers and to make the prototype as global as possible, thus giving the best results possible within the survey. It was also important to get as wide range of ages as possible to see if there would be any difference in usage, therefore changing any possible results in the survey.

During the survey the participants were asked several questions regarding; their habits of using a smartphone; their usage of a VPA; how often they used VPAs and why; and how happy they were about current VPAs, this information was used to gather relevant data about VPAs. A certain question was asked of the participants to determine what environment the participants would deem a VPA would be appropriate or inappropriate to use. This created a good indicator for where the user would use their virtual assistant and provided ideas for the next step in the thesis. The participants were also questioned about what features they would like to see in future VPAs and if those features would persuade them to use a virtual assistant more often.

3.1.3 Interviews

A series of interviews were performed to secure a more detailed view of how frequently VPAs are used in current systems. The questions were based on the survey and can be found in Appendix A. The purpose of extending the initial answers and obtaining a more in-depth response was to allow a deeper understanding into the daily uses of digital assistants and for what purpose the end user would use these services. This allowed the participants to further elaborate on the questions, as these tended to open-ended, and required them to explain their thought process.

Five participants were interviewed with various degrees of expertise in virtual assistants. These interview sessions were performed in their respective homes to make them feel comfortable and at ease. This type of informal interview was not in any way recorded and lasted for approximately 30 minutes and began with a short introduction of the thesis and a few easily answered questions, such as name and age. Following on from this each interview was conducted in an informal way to allow a more free-flowing discussion. The limited number of people that were interviewed did not provide a statistically accurate basis on which the thesis could be founded; instead this created a variety of viewpoints for the uses of digital assistants.

3.1.4 Test Study

The test study of digital assistants was conducted by the author to discover the range of capabilities. By testing each VPA within the same environment, with the same commands and by the same person, the digital assistant could be individually analysed to discover any variations in the results that were produced.

The assistants chosen to be in the study was Apple's Siri, Google Assistant and Amazon Alexa. A list of commands was created to test the VPAs and discover their abilities and how they performed. This testing was performed to compare the survey answers, with specific focus on the current abilities and performance of the already marketed digital assistants. This helped to refine the selections for the prototype.

These tests included intelligence testing to monitor the artificial intelligence functions in current VPAs, to see whether the VPAs could perform complex tasks such as contextual understanding. The VPAs were also tested to see if they could understand commands and possible intentions of the user. Speech recognition was frequently tested to discover whether the digital assistants recognised and understood the speech patterns and if they understood the intention from the user. Complicated command was thoroughly tested to see the kinds of commands the digital assistant could perform and to discover any abilities or performance that the digital assistant lacked.

The commands tested ranged from simple commands such as, "Google Batman" to more complex commands with contextual understanding such as, "Who is Batman?" and "What is his name?", to test if the VPA understood relevant follow-up questions.

3.1.5 Requirement Specification

Before starting work on the prototype, a requirement specification was created, so that the prototype could be evaluated and tested. The specification was established using the data from previous phases to recognise gaps in current VPAs and by using functional requirements [19] to capture what the product should do and how it reacts to certain commands.

A scenario was created that demonstrated what tasks the prototype would be able to perform. Test cases were established to show how the prototype would react when it was used, including; what the pre-and post-conditions would be; what phrases it would react to; and what type of visual or audio feedback it was going to produce. In conjunction with this requirement specification a flow chart was created that mapped out some of the tasks and their respective responses.

3.1.6 Prototyping

Identifying the needs and habits in previous phases helped to recognise gaps in current VPAs. To achieve the goals set out in the project a need was identified and the resulting conceptual model was created to explain how users could interact with the system. By using the literature studies, surveys and test study to grasp concepts of what users want from a digital assistant the development of an idea for the prototype could be realised.

A discovery during the previous phases was that people wanted to use more voice control actions via virtual assistants, but that current home systems were not developed well enough or were not affordable to be able to connect the average home. However, a single point of interest was discovered during the test study that not one digital assistant was able to show a complete status report for more than one item or task at any given time. Therefore, to develop a prototype that would be able to check and change the status of multiple objects at once with a single command would be an ideal concept.

The first mention of stacked commands came from an article that mentioned the fact that Google Assistant lets you add custom shortcuts [18] and that the feature of stacking these commands can let the user perform more than one command at a time. With this idea of stack commands, it could become possible to get good overviews over a complete system just by performing various commands after each other.

3.2 Prototype

This prototype was created to be able to test and visualise status reports by using stacked commands. By stacking individual commands one after another it enables the product to perform several commands at once.

To completely utilise the stack commands, a home system was ideal to comprehensively visualise such a system. A home system was selected for the prototype as it was easily relatable to the average user and could already be seen as a future step for VUI, as established by those participants that were surveyed.

Creating both a GUI and VUI these can help to illustrate how a system responds to commands. By showing a GUI, this can illustrate added features such as, toggle buttons, whereas the VUI uses voice responses to illustrate the system. By combining both interfaces this can create a more complete system that does not discriminate against a possible future user. A VUI can be created in a similar fashion to a GUI, but a VUI does not have a long-lasting feedback such as text on screen, as a GUI. To combat this, it is important to have voice feedback that is clear and concise, thus by having short phrases spoken back to the user it helps the cognitive load for the user.

In this project both a VUI and a GUI were created to form one joint prototype. By providing better feedback, such as incorporating the same words spoken as shown on the screen, this helped the user to be guided throughout the interface.

There were obstacles in creating the prototype, due to a lack of hardware to implement this kind of system. Instead to visualise the system Google Assistant was used with pre-programmed shortcuts to create a visual display with a pre-designed picture and a corresponding command phrase. The lack of hardware connection meant that the actions that were to be performed was not physically possible and so the user would have to assume that the prototype had performed the voice control actions.

Another obstacle with this Google Assistant system was that the user was required to select the picture to obtain a full screen version, so unfortunately a full immersion would be missing in the final prototype. The voice response would be generated from a speaker connected to a nearby computer that had pre-recorded phrases using a Text-to-Speech (TTS) application that was then played back to the user after each command to fully illustrate the complete system.

The GUI that was created utilised icons that displayed the corresponding item, such as an icon of a door for the door setting (Fig 3.1) and an icon that represented its current state of being. The prototype did not allow any interaction with the picture, but would in later iterations be able to change the state by clicking the toggle button.

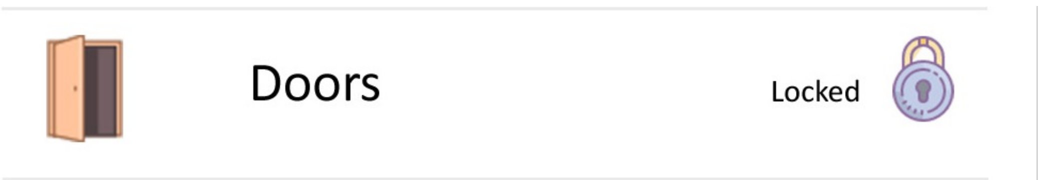

Figure 3.1: Displaying one door object in the GUI

The user interface was based on Google Assistant as this was the most competent virtual assistant found in the test study and was the most viable assistant with the device that was used in prototype testing. This prototype design was then created with the help of Adobe Photoshop and the icons came from Icons8 [6] which creates open software icons.

To illustrate the prototypes capabilities a flow chart was created in Lucidchart [8] which is a tool to be used to create flowcharts. The flowchart was created (Fig 3.2) to visualise each command with the resulting feedback and how the user would navigate through the system.

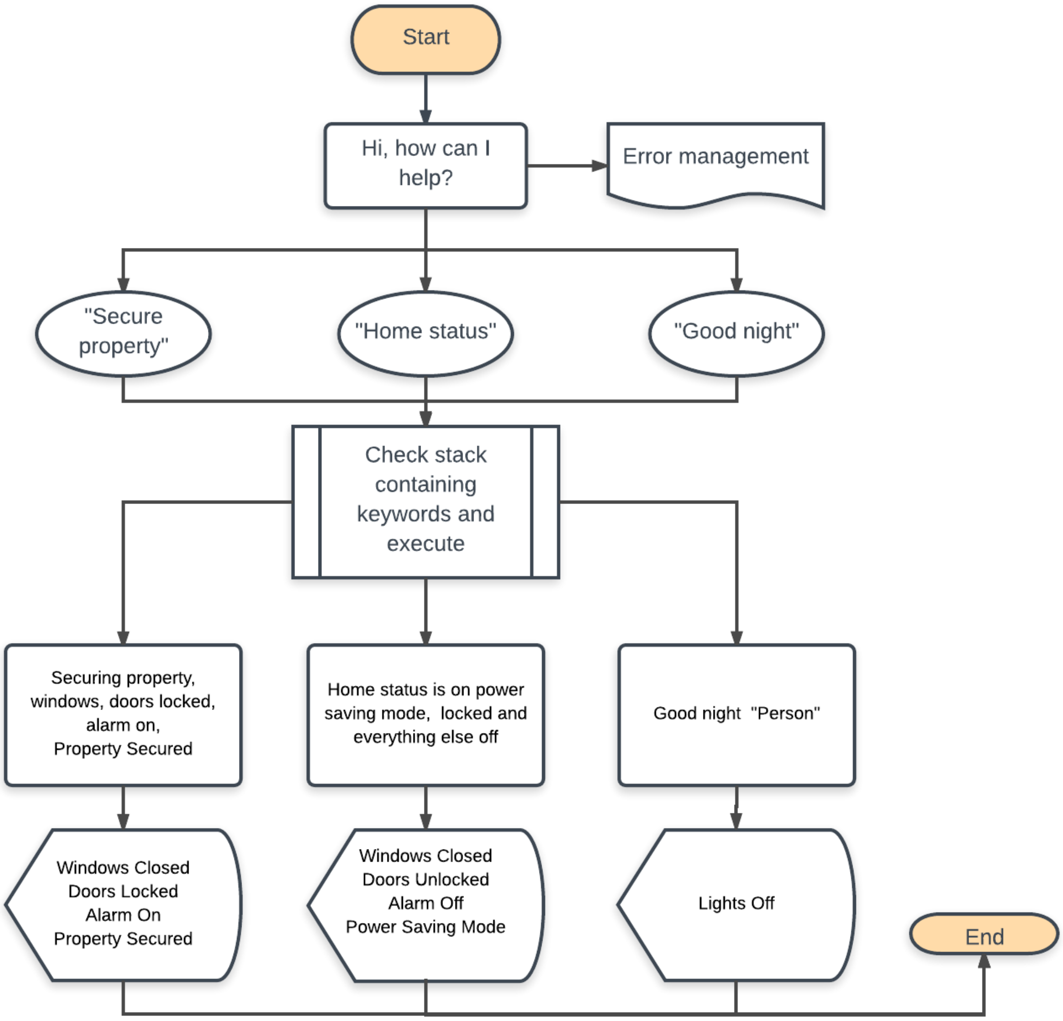

Figure 3.2: Displays the flowchart created to symbolise the features in the VPA prototype.

3.3 Usability Testing

To conduct this usability test, a test plan was created which can be found in Appendix B. This tried to eliminate any problems and prepared the test inclusions. To start the test plan, a series of problems were created to set out to ensure the questions were satisfactorily answered. These problems are as follows:

- Would this prototype help the development of VUI?

- Does it give the user enough feedback?
- Is the result satisfying for the users?
- Would the user ever use the product again if it was implemented to completion?

The test began with a brief introduction to the participant of the project and what the purpose and goals were for this thesis. Afterwards a description of the prototype and the available commands that were to be used were given to the participant. The set tasks for the test included a scenario in the participant then chose from one of the available commands and spoke into the phone; the phone then generated a specific picture and voice feedback that described the result.

The participant would during the test be asked to think aloud and to explain why they made that decision. By asking the participants to "think aloud" this helped the test moderator to verify cues from the user verbalising their thought patterns and explain why they are performing the tasks in a particular way. By choosing this method no time performance tests were made as the task of verbalising thoughts can be time consuming.

There was a total of six different response sections within the scenario that was given to the participant.

An example of one of these response section was: *"On your way home from your skiing holiday, you are thinking about what you need to do when you get home, but then you realise you can have a nice warm and cosy house waiting for you when you get there. What command works best for getting things ready at home?"* This scenario was created so that the user could imagine themselves within the given situation and performing the controls available. More response sections of the scenario can be seen in Appendix B.

The problem definitions would form the basis for the testing and for the questions that were asked to the participants. The problems resulted in a few questions that were formed so that the participants could explain their thoughts about the prototype; this is also known as qualitative data. The qualitative questions that were asked are as follows:

- Did the prototype give sufficient feedback?
- Did the prototype give sufficient feedback to visualise the system?
- Do you feel like you were in control of the system and could reverse any actions?
- Would you use this system again?
- Would you prefer giving one command for each item instead of one complete command?
- How do you feel it went?

During the test there was also quantitative data that needed to be collected. Since it is a voice based system there are several different variables that could influence the system, such as a fail in the speech recognition. This was not included among the errors that occurred which would only have tested the phones AI ability to recognise the right words. The following is the questions for the quantitative data:

- How many errors were recorded?

The test sessions were approximately 30 minutes which included both the test and the debriefing interview.

Chapter 4
Results

This chapter contains the results section of this thesis. In the result chapter all the relevant results gathered from previous chapters are recorded and compiled giving the reader all the results that are relevant to the project.

The results section serves to visualise all relevant results gathered from the methodology section, expanding on the methods used and compiling the results.

This chapter is categorised in the same three parts as the previous chapter. The parts were background research, prototyping and usability testing. The first part consists of the background research, which delves into the results from the survey, interviews and the test study. The next part is the prototyping where the results from the prototype are reported and where the design choices of the VPA interface are explained and a view of the interface is displayed. The final part is the usability testing where the results from testing are visualised and explained.

4.1 Background Research

The gathering of information was an important step in analysing the user habits of the VPA. To provide such results different methods were used to record the habits. The methods used were a survey, interviews and a test study. This is where the results of those methods are recorded.

4.1.1 Survey

The survey conducted rendered responses from people of all ages with most of the responses being in the 18-35 range. (Fig 4.1)

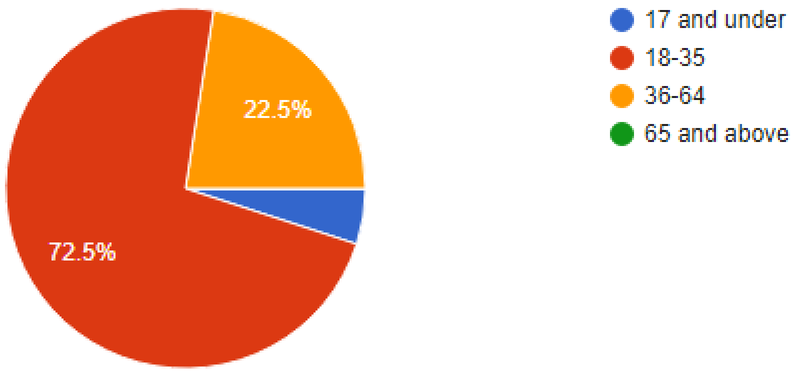

Figure 4.1: Pie chart displaying a percentage of the participants age in the survey

All the respondents used a smartphone and operated one every day. Of the respondents over 60 percent used voice actions on either a smartphone or a tablet. The survey rendered 40 responses from people of all ages and different backgrounds.

The responses for not using voice control were:

- No need or reason for it.
- Embarrassing to use.
- Doesn't produce the right words that I say.
- Have no use for it.
- Quicker to type than explain what you want to do.
- Prefer visual input
- Old habit. So used to navigate using fingers

Of the respondents that did use voice control actions there was a diverse range over how often the respondents used a VPA. (Fig 4.2)

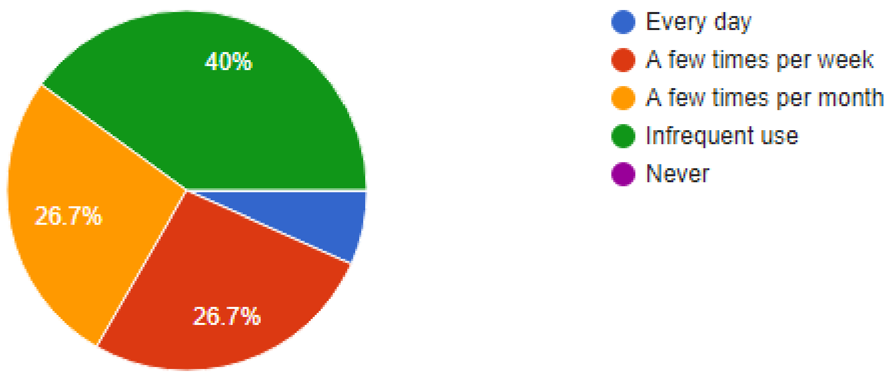

Figure 4.2: Bar chart displaying how often the respondents used a VPA

The reason why people used voice control actions was the following:

- Use it in the car
- For fun
- Spelling and reminders
- Because it is easy
- Hands-free experience
- Quick and easy way to Google something
- Easier because of dyslexia.

On the question of which voice actions they used. (Fig 4.3)

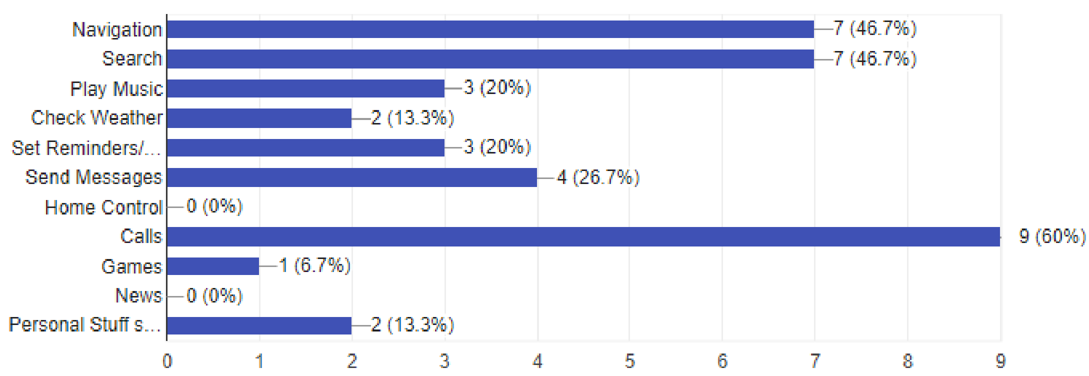

Figure 4.3: Survey results displaying frequently used voice control actions

A question that was raised in the survey was how satisfied people were with the current digital assistants which resulted in: (Fig 4.4)

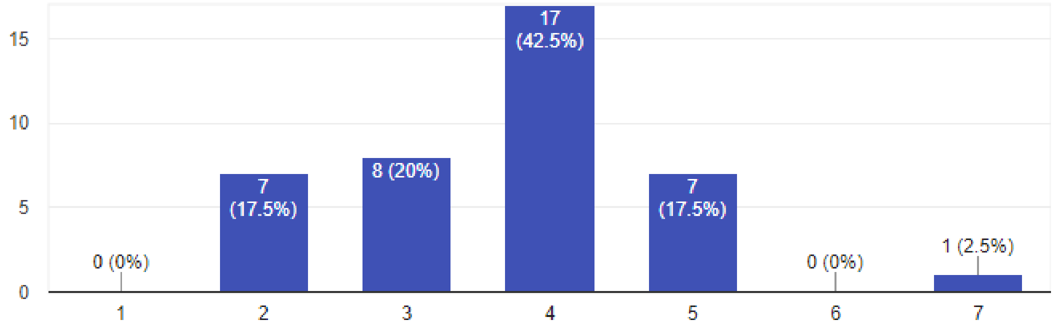

Figure 4.4: Survey results depicting how satisfied the participant are with current VPAs with 1 being very unhappy and 7 complete satisfaction.

This showed that the respondents are generally not satisfied with current VPAs.

Another survey question that was asked was to show how the respondent thought voice control actions could be made better and make them use it more often. The result of these questions was:

- Better AI
- Nothing, don't see the use for it.
- If they can handle more complicated commands
- Better speech recognition
- More intelligent and connected to more services
- Needs more languages such as Swedish

On the question of whether the respondents would like to see any new features on their digital assistants they answered:

- None that I can think of.
- Home automation control but I guess that is in the future.
- More adapted to in-home use.
- More connectivity to IoT home appliances with the ability to control them.
- Translation ability that could read or hear voices and play it back in native tongue.

- Not needed right now, more focus on the usability.

To get a feel of where the respondents were most likely to use a digital assistant the question was asked of where they felt most comfortable using this feature and where they felt they could not. (Fig 4.5) (Fig 4.6)

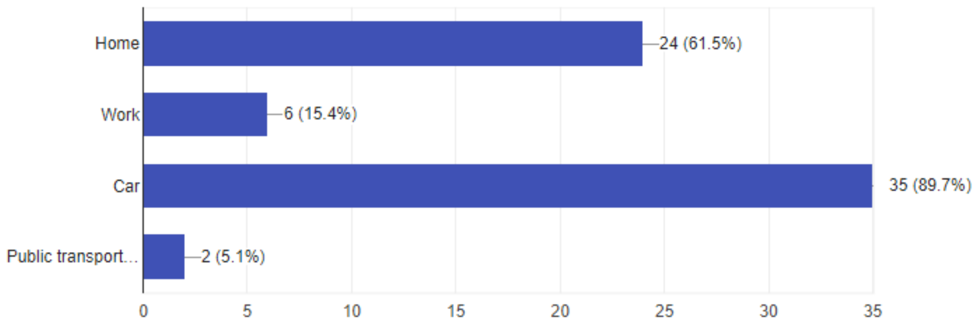

Figure 4.5: Survey results depicting the environment where the user is most comfortable using a VPA

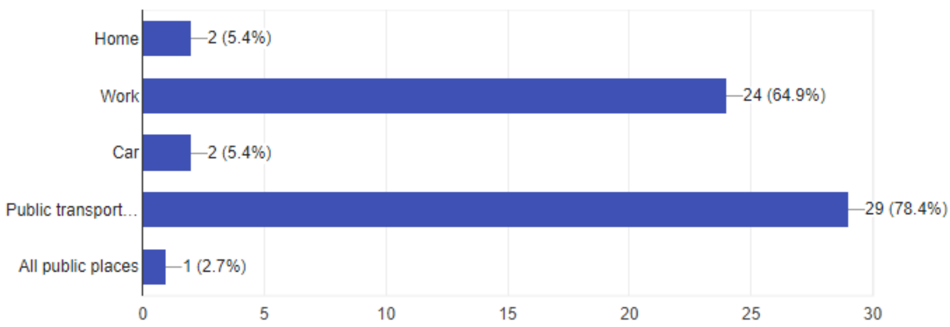

Figure 4.6: Survey results depicting the environment where the user is not comfortable using a VPA

With this survey multiple valuable notes were gathered on the habits of the respondents and for what they currently use VPAs.

4.1.2 Interviews

The interviews were conducted in an informal environment where a discussion and further elaboration on their answers was made.

Interviews were conducted with five people of varying ages and experience of using a digital assistant. This yielded a return of useful qualitative data to fully grasp what each interviewee felt and their habits of using voice to control actions.

The participants were of different ages with the youngest being 25 and the oldest 60. Their education also varied from high school graduate level to higher education within university as well as differing occupations from hospitality managerial to engineering positions.

The participants all rated themselves with an above average experience in using VPAs and all had used a VPA at least once per month, with one of the participants having used it daily. The most frequent voice control actions were to Google, for entertainment purposes, phone calls and navigation. When the question arose as to whether they liked the ability to use current digital assistants the consensus was yes, but that it had several improvement options. On the question of what could make voice control actions better the participants were also in agreement that it needed to have better contextual understanding and be able to understand complicated commands. One issue that occurred for all the participants that were interviewed, except one, was that they had to speak English, even though they were native Swedish speakers, as it did not provide as broad a spectrum of commands using Swedish as it did English.

Most of the participants responded that they would like to see digital assistants used more in their own homes, and that with the rise IoT applications that it could be better integrated in the future. They all discussed the ability to control more in their homes with their smartphones, not only to with voice control actions, but that currently this was not a complete enough system or affordable to be able to completely connect your entire home.

Of the five participants that were interviewed all had mentioned that they were not comfortable using their voice to control actions on public transport or at work and with one adding that using voice control actions in public would be considered rude.

4.1.3 Test Study

A comparison study between Apple's Siri, Google Assistant and the Amazon Alexa was conducted by checking set voice commands to see if the virtual assistants managed any differing results. It was found that each system had its own strengths and weaknesses.

During the test study it was found that the different virtual assistants handled all the simple commands brilliantly but failed when any commands arose where the assistants had to handle two or more things at once. Google Assistant provided the best solution by its ability to understand contextual conversation, such as the command "Who is Batman" and afterwards "What is his name" (Fig 4.7), something the other assistants didn't completely understand.

This generated positive feedback in the design process going into the prototype phase. From the test study of digital assistants an important breakthrough was found that users could only perform one command which generated one action at a time, this made it hard for the user to obtain complete knowledge of the entire system.

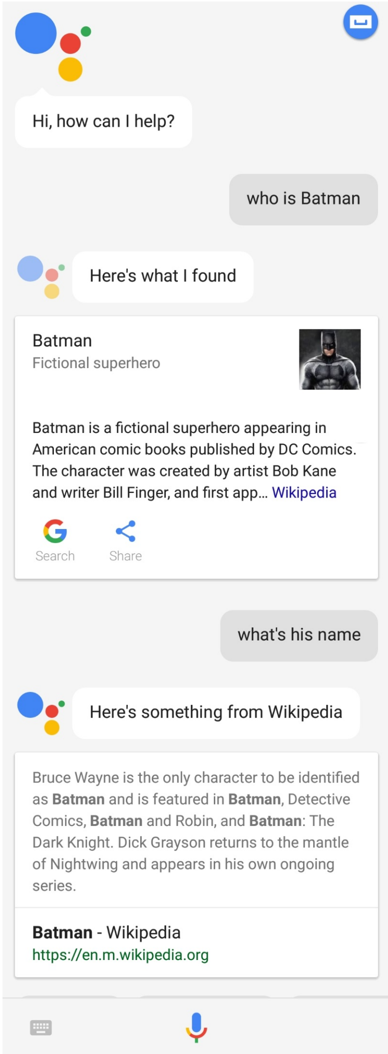

Figure 4.7: Displaying Google Assistants ability to understand contextual conversation.

4.2 Prototyping

By using stacked commands, the option of setting a single command to perform several other commands at once can be implemented. This is the main feature of the prototype, which illustrated how this command system could be performed. An example of stacked commands would allow a movie to start whilst simultaneously dimming the lights with a single command of "Watch a romantic movie". This example can be created by stacking two commands such as "Watch a romantic movie" coupled with a command "Dim the lights". This feature of stacking commands is not currently available in current VPAs.

4.2.1 GUI

This prototype was developed to be able to test in a home system environment, as well as to test if people preferred the ability to use a single command for a multiple command response. To visualise this interface multiple objects were listed with corresponding icon and status of the object.

The GUI that was developed displayed different characteristics with icon, name, the status and a toggle button. Each object contained the name of the object such as “Alarm”, an icon, such as an icon resembling an alarm signal, and finally the status of that object written with a corresponding toggle switch that could control the status of the object individually.

The icon was chosen to be a visual representation of the object, such as a picture of a window in place for the “Windows” object. (Fig 4.8) This was designed to help the user quickly browse over the system and see which objects were represented and their status. To help each object have a suitable icon for the visual representation a design choice was made, and universal icons were chosen to avoid any misunderstandings within set cultures. The icons were resized to fit each object within the same number of pixels in the list and to look consistent. The icons were chosen to illustrate what the object was, and to avoid misunderstandings, easing the use of the system for people with differing language skills.

The name of the object was displayed after the icon in a clear font to make it easy and quick for any user to read.

The status was displayed in both text form and visual form to help the user in differentiating the objects status. The visual form was displayed with two different toggle icons in two variations. One of the toggle icons that visually represented a lock, appeared in two different shapes, one of which was locked and the other unlocked. The other icon was a classic toggle switch which represents an ON/OFF switch. To help colour blind people this switch was also represented with both the text feedback and the toggle being moved to the right and to the left.

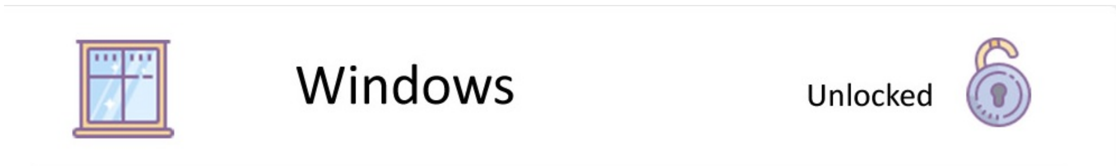

Figure 4.8: Displaying the "Windows" Object in the GUI

After the list of objects there was a confirmation box which gave the user extra feedback to visualise that something had occurred. This box had a green background colour to symbolise that the action had been completed and to give closure for the user.

4.2.2 VUI

The interface also included voice feedback which was given to the user through a speaker system separate to the phone. Unfortunately, the prototype was unable to play back the necessary information on the phone as Google Assistant does not allow the use of third-party speech applications being used in conjunction with displaying pictures. The voice feedback that was generated included both Google Assistants own feedback and the voice feedback through the speakers.

The VUI did not need additional feedback as the GUI but instead focused on simple commands such as “Windows closed” to help ease the cognitive load in the user’s mind. By allowing just two phrases more stacked commands able to be performed and remain comprehensible. To help the user in confirm their action the same confirmation box in the GUI was also reiterated in voice. A command such as “Home status” would have a GUI and the voice response (Fig 4.9) as follows:

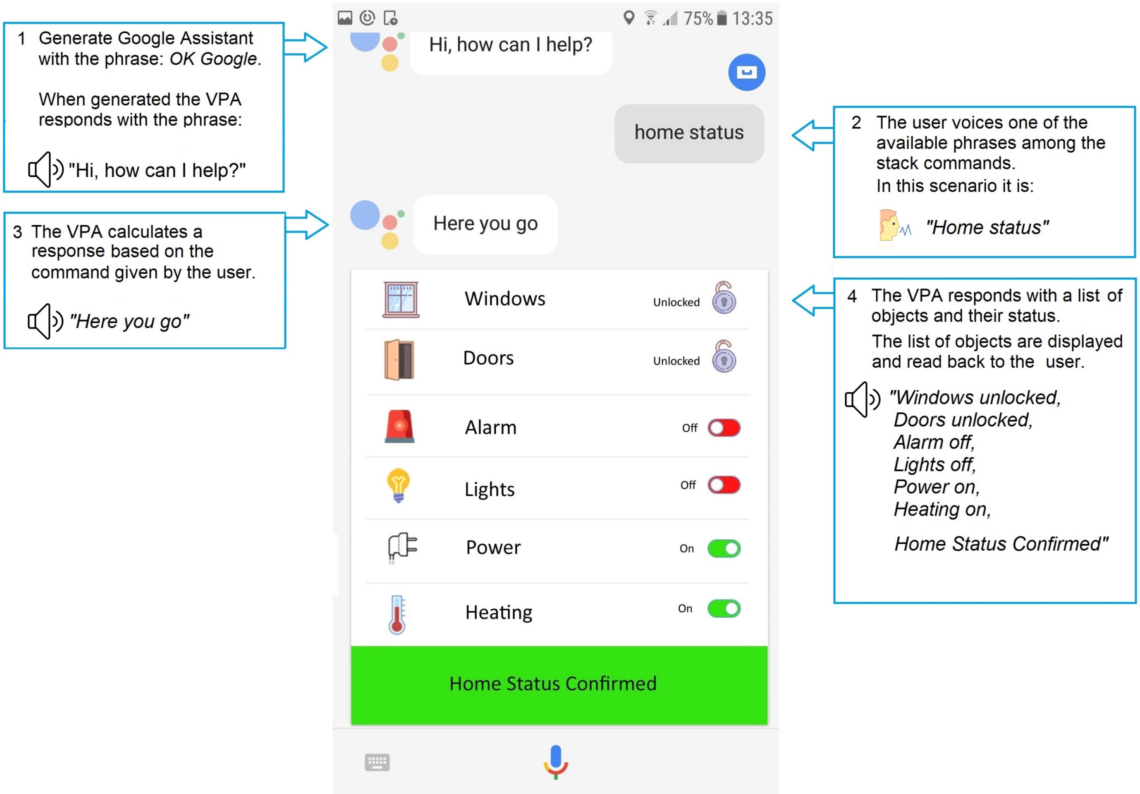

Figure 4.9: Prototype displaying the GUI and VUI.

The voice commands were chosen to promote natural speech to the virtual assistant and help promote talking to the assistant as a person instead of forcibly thinking about it as a computer. This would be adjustable and customisable for each user in future iterations. It was also chosen to be simple and intuitive commands for the user to say in those specific situations.

To illustrate the scenarios and different response actions a list of commands was available in the prototype. These were:

- Home Status
- Secure Property
- Power Saving Mode

- I'm Coming Home
- I'm Home
- Good Night

These commands controlled six objects in diverse ways as each applied a different change to the user interface. A "Home status" command gave the user an overall status rapport for each of the objects in the list whereas a "Secure property" command (Fig 4.10) secured the home by changing the status for each object, such as enabling the alarm and locking the doors, to facilitate a secure home.

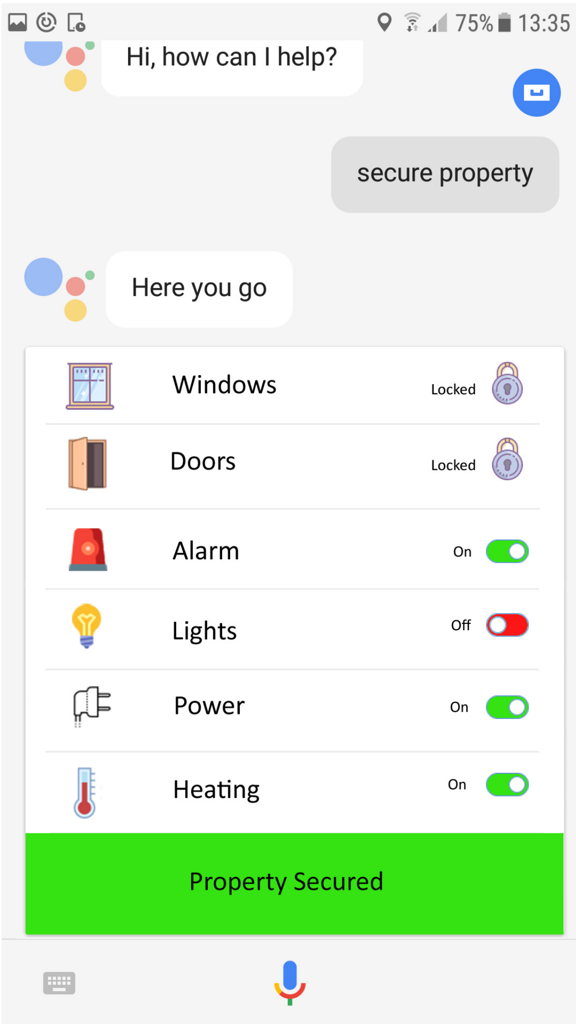

Figure 4.10: Prototype displaying the "Secure property" command

4.2.3 Usability Testing

The testing was performed on seven people of various backgrounds. The participants were within different age brackets from 16 to 66 with half of them being from Sweden and the rest from English speaking countries. This was to ensure that any problems with language skills did not need to be addressed since the entire test was performed in English. All the participants had different degrees of technical backgrounds but all of them owned a smartphone that they used daily.

The debriefing interview indicated very positive feedback. The first question asked about the testers thoughts on the prototype and the responses are as follows.

- Good
- It was very easy to use but at the same time efficiently done.
- Would be a wonderful system to have
- Don't use my phone very much but if I did this would be very useful.
- Smart, easy and helps to alleviate the stress load.
- The concept is very relevant and up to date with current technology advances,
- I think this would reduce stress being able to command the house to do certain tasks.
- I also liked that I could check multiple things at once such "home status" gave several points of information rather than having to check each one individually, this would also eliminate the possibility of missing something.
- Would be interesting to know if I can personalise the commands to suit my preferences like temperature being adjustable or I want the windows to stay open but the doors to lock when I say secure property.
- I believe the idea basis is well thought out. It would be advantageous to be able to do multiple commands at once with voice control from my phone or even a google home box (smart device)

The next two questions asked the participants about how they thought the test went and if they would use a similar system based on the prototype again. This resulted in overwhelming positive remarks on the prototype, with the participants commenting that it was good and that they would like to use it again.

During the debriefing the participants were asked to elaborate on why they would like to use the prototype again and the results are as follows:

- A good way to simplify your life
- Ease of use and practical
- It's easy.
- Very effective and will become useful for everyone in the future
- User friendly and I think will become very popular.
- It's easy.

- I believe it makes life easier for any person. In several circumstances it would save me time and is less hassle then trying to type several different things into my phone, one after the other.

To further elaborate on the prototype all of the test subjects thought the prototype gave sufficient feedback. They didn't feel that the prototype overwhelmed them with too much information but that it gave short and essential responses. The participants also believed that the prototype provided them with enough feedback to visualise the whole system.

The participants gave an overall positive feedback when asked about if they had a feeling of control over the system and that the system allowed them to reverse any actions. They had not tested any individual commands but also felt that since the prototype controlled multiple objects it would be intuitive to be able to control individual objects too.

Two of the participants wanted to add the option of making the commands completely customisable to suit their individual preferences, for example by changing the windows to be open instead of closed in "Secure property". The participant also wanted to add more customisation in more specific ways, such as setting an exact temperature for the house instead of just adjusting the heating on or off. From one the participants they thought that as the prototype did not give a genuine response from an enabled application on a smart device this made it hard to visualise the final product. Since the prototype was not connected this kind of response was expected. The participants all said that they preferred giving one command to multiple objects instead of doing multiple commands to control all objects individually.

During the usability testing the amount of errors was recorded, as shown in Table 4.1. The types of errors recorded were of two different varieties, there were speech recognition failure and an incorrect phrase spoken in the scenario. Speech recognition failure resulted in a command incorrectly interpreting "Good night" as "Good nights", which prompted a completely different command than intended.

Table 4.1: Displaying the amount of errors that occurred during the test

Error Type	**Total Number of Errors (7 Participants)**
Speech Recognition	3
Wrong Command	2
Total	5

Chapter 5
Discussion

This chapter contains the discussion section of the thesis. In this chapter the relevant results from the previous chapter are discussed, analysed and compiled.

The purpose of this chapter is to compile the results through analysis, evaluation and discussion of the results. The prototype will be evaluated through relevant theoretical background to reveal if the standards that were set out in the design phase have been upheld.

This chapter is divided into three parts. The first part is where the results from the previous chapter are analysed to see how they performed. By analysing the information gathered from previous chapters it will be possible to draw conclusions in the following chapter. The second part is prototype evaluation, where the prototype will be evaluated through the relevant theoretical background, including the eight golden rules. Prototype evaluation is performed to ascertain if the prototype could be described as a usable system. Also in this section the problem definitions are evaluated to assess if the project has answered the research questions that were set out in the introduction. The final part is where the remaining limitations in the project are discussed and analysed to establish if there is any future work in this field.

5.1 Findings

The findings from the results chapter will be analysed and discussed here.

5.1.1 Survey and Interviews

The survey gave a good indicator of how the current VPAs are working. The participants had mixed data about the use of VPAs, the majority of which were not using a VPA at all or if they were it was used just for fun. This signifies that the possible potential for assistants has not yet been unlocked. The participants did not see the purpose of using an assistant, especially in public spaces, because they felt embarrassed or believed that it was rude or annoying to the people around them.

It was also indicated that VPAs simplified life for some participants, for example people with reading disorders such as dyslexia, where the VPA provided relief for the user as the assistant could provide results faster and easier than manual input. One of the indicated advantages for using a VPA was the fact that it became much faster to find answers in certain areas, such as looking something up on Google.

The participants were currently not getting the results that they wanted because of the lack of a smarter AI, better speech recognition and language choice shortage. This lead to frustration due to the willingness to use a VPA but finding that it lacked because of an immature system. The survey also asked the participants which locations they would feel most comfortable using a VPA and found that the comfort of their home or car was where they saw the most potential. The participants concluded that there was no lack of features but instead that they required better interaction.

5.1.2 Test Study

The test study was conducted to see exactly what kind of commands the assistants could manage. In the study one saw how far the VPAs had been developed since they first began. As VPAs now had solutions for setting up appointments or reminders and were even better at understanding contextual situations using natural speech. The ability to understand follow up questions associated with initial commands, such as “Who is Batman?”, and followed with “What is his name?”. Understanding these kinds of commands leads the user into believing that they have interacted with an actual personal assistant instead of just giving the VPA a command.

It was found that none of the assistants examined could handle a command that required multiple tasks to be completed. The assistants did not provide a decent overall viewing of the system, such as multiple status reports, which lead the user into believing that they could only modify objects but could not control the system.

5.1.3 Prototype and Usability Testing

The prototype was designed to ease the cognitive load of the user by being able to process multiple things at once. It was designed for high usability and it was important that the prototype was accessible and catered to the needs of people that had the most interaction with a VUI. It was therefore designed to provide the same features for everyone.

The user interface was designed so that the same information was both displayed and read aloud to accommodate for different disabilities. The prototype was also designed to understand simple commands but still being able to perform multiple tasks. By performing simple actions, the system is more intuitive for the user and therefore could help with users interacting more with a machine on a daily basis.

To make the graphic interface less sensitive to errors the objects had a toggle button along the side of the screen to be able to adjust individual objects in case the voice command did not perform the task the user had intended. This would have also been possible with just a command directed towards individual objects. To open the system this way ensures that there were multiple options of adjusting existing commands as people have different preferences when performing tasks, whether it is voice or visual based.

To make the prototype a comprehensive audio system, steps had to be made to turn it into an auditory display system. To make sure that the system was completed correctly there were necessary steps made so that the prototype adhered to these three broad categories: alerts, status and data exploration. The prototype was designed to showcase these options in different ways.

With the testing several results were received and gave the prototype an overwhelmingly positive feedback. All the participants liked this way of interacting with the system and gave good reviews. The feedback from the participants was that the option of doing multiple things at once with the prototype was easy, good and efficient. The overall view was that they would like to use this kind of feature in the future to simplify life and emphasised that it was an easy system to operate.

With a new system it is common to become frustrated and feel lost by the limited amount of feedback received. In this prototype there was a clear beginning, middle and end which allowed for various types of feedback throughout the system. The platform that was used was Google Assistant which provides tactile, visual and oral feedback within the system.

To create a simple interface that everyone could understand there was a need to give sufficient feedback. The post test questions included asking whether the prototype gave enough feedback and the participants all answered that they were happy with the amount of feedback they received. By giving both oral and visual feedback there was usually redundant information given but since the interface only included the name and status reports of the object it formed just enough information to complete the understanding. However, this could have changed if the number of objects had increased.

The participants thought that they were in control of the objects even though the system was not developed to be able to test this. A participant also commented that if you could control multiple objects with a phrase it would be natural to be able to control individual objects as well.

After the prototype test the participants were asked if they would like to add any further features to the prototype. A stand out opinion was that the option of being able to change their preferences, which would personalise each command to individual tastes and needs, would be a good feature to add. In future iterations it would allow the user to add in exactly what multiple objects they would like to control.

The testing yielded some errors in speech recognition and a wrong phrase used in the scenario. The error in speech recognition was not part of the project, but voice based systems currently have these kinds of problems, this only happened with a very low frequency and could probably be disregarded in this instance. The wrong command error occurred when the participant responded with an incorrect phrase in the prototype and yielded a different conclusion than intended. One that this happened was when the participant was eager to try the prototype and responded before the scenario had been concluded.

Another error arose when the final command of "Good Night" was meant to be said, which in the scenario would lock the doors when the user went to bed but the phrase "Secure property" was spoken instead. This error occurred because of the lack of explanation beforehand of each the commands. To the participant it was natural to assume that they wanted to secure the property when going to bed not knowing that the "Good night" command performed a similar action with a slight change. This proves that the customisation of the commands is indeed a good idea to implement in future iterations of the system where the user can assign each list of objects to its own desired phrase.

5.2 Prototype Evaluations

When designing a prototype, a good start is to use industry standards in designing a usable system. In this project there have been two different viewpoints that have been considered when designing the interface and subsequent prototype. These have been the *Eight Golden Rules* by Schneiderman [17] and *Design Principles to Make Everyday Things Usable* by Donald Norman [14].

5.2.1 Eight Golden Rules

Starting with the Eight Golden Rules and the first rule to *Strive for consistency* the prototype should have actions that resemble each other. The prototype has done this by making each interface like the others and making the same actions available for the user in all variations.

The second rule to *Cater to universal usability* is evident in that the prototype provides different types of feedback. Although there was not an advanced model it was such a simple but efficient tool that it eliminated the need for unnecessary complexity. The platform in which the prototype was based also allows input not just by voice but also by text further catering to diverse users with different abilities.

The third rule to *Offer informative feedback* is achieved where the prototypes speech recognition recorded the dialogue spoken, then let the user know that it had recorded the action and then processed the information until it reached a conclusion. After each command a confirmation box was displayed at the end that signified that the command was accepted and completed.

The fourth rule to *Design dialogs to yield closure* is shown when the prototype allowed for a beginning, middle and end feedback. It allowed the user to receive information for each action was processed and helped the user to know that an action had begun, was being processed and showed the result when it understood the command. The prototype would after each command operate each object individual object and adjust the system so that the result was portrayed in the real world such as a door is locked.

The fifth rule to *Prevent errors* where the prototype could have handled a command either by interpreting the command incorrectly or understood the command differently than intended. Unfortunately, these kinds of errors occurred but happened less frequently as the system matured.

The sixth rule to *Permit easy reversal of actions.* The prototype did not have this implemented. The toggle button for each object was not adjustable but was visible to see that there was a toggle button where you could have set the status. This was discussed during the test phase and each participant understood that the prototype would allow easy reversal of actions.

The seventh rule to *Support internal locus of control* was supported by the prototype conveying to the user knowledge on the status of each object after each command and a confirmation box that showed that the command had succeeded. The prototype testing asked the participants if they felt in control of the system and all of them responded yes in that regard.

The eighth and final rule *Reduce short-term memory load* was supported in two different ways. The first was by only visualising six objects in the list of objects this is slightly less than what the memory load can handle. This is something to be cautious of in future iterations to not overload the user. By providing simple and intuitive commands people did not have to think about which command they would like to have used but instead perform the command which was natural to them.

5.2.2 Design Principles to Make Everyday Things Usable

The first principle Donald Norman speaks of is *Affordance* and to uphold this principle in the interface the object had to be clear with where you can interact within the system. For each object there were two icons in the interface, one of which was an icon that depicted what kind of object it is and the other toggle button. These two icons served two different functions but only the toggle button was interactive. There were efforts made to make the toggle look more selective as the toggle button has two different modes; On/Off; or Unlocked/Locked, which is illustrated in the different response sections of the scenario.

The second principle is *Mapping* where the prototype has an action and a reaction to a task. In the prototype there were no real-world actions as it was not a connected prototype but for the future this would be one of the next iterations and would show the actual status of objects.

The third principle is *Constraints* whereby restricting the system makes the system more usable. In this case the prototype only allowed two different modes On/Off and Unlocked/Locked. The interface only allowed input either by voice or by selecting the toggle button which restricted the system.

The fourth principle is *Visibility*. This principle raised the need for a GUI, which was hard in a voice based system, to display the actions that were available. The interface did not visualise any actions before they were spoken and only displayed the results after they were spoken. This system relied heavily on the user knowing which commands were available, but this could have been corrected by asking the VPA what actions it could have performed.

The fifth principle is *Feedback*. The interface provided a visual and an intended audio feedback when interacting with the system which guided the user through a simple interface. During the testing the participants were asked about the feedback and all of them were happy with the amount of feedback they received and thought it provided sufficient information to the user.

The last principle is *Consistency* and the interface looked and performed similarly every time an action was performed. In the prototype all the objects were treated the same and performed in a similar fashion.

5.2.3 Evaluating the Problem Definitions

To start this project there were a few problems identified and about which the project was then based.

The first problem was to assess if a VUI can improve the interaction and accessibility between user and machine. To analyse this the test phase proved that the features of stack commands could help the user analyse a system faster and easier. It also gave the user the ability to scan through a custom set of objects which the user was interested in. By having a simple and intuitive system that did this and with sufficient feedback it helped to alleviate the usability issues and ease the mental stress. The participants all gave good reviews about the prototype even though it was an incomplete system which gives hope for the future iterations of this kind of software.

The second problem was to examine what the future had in store for VUI. The future looks bright for VUI because since the beginning of this project there has been at least two major VPAs that were released, and they also received good reviews. With this the major VPAs that are currently available are continually being upgraded to include more features and receiving smarter artificial intelligences. For the future of the prototype and stack commands it is up to software companies to include more complex actions that can solve multiple things at once in their assistants.

The third problem was to check the possibility of developing the current VPAs. The prototype proved that further development of the assistant is still possible using a voice based system. By improving software to better the speech recognition and improves the systems

AI it will help users and make it more usable. So, with a more mature system it will hopefully be even more popular than before. The feature of stacked commands will hopefully take the VPAs to an even higher level of interaction for the user.

5.2.4 Remaining Limitations and Future Work

The prototype is still at a concept level and needs to be fully implemented which could be achieved with the help of the Google Assistant API (which is used to implement software for a VPA) or other VPA APIs. Even with the ability to implement the software that is needed there are still some limitations that could hinder the development.

The first thing is that of Internet of Things applications that need to support the ability to perform tasks from a third-party app so that actions on the connected objects can interact with the VUI. To go with this, the applications need to run smoothly. Speed is a key factor in order to have a natural speech pattern with the virtual assistant - it is not ideal for a command to take a long time.

To completely implement the prototype there are several features that need to be implemented beyond just the connectivity with the objects. The user interface implemented along with all its features such as the icons for the objects and the toggle button for the status of the object. By clicking the icons, it should lead the user into advanced settings into that specific object so that changes can be made.

Feedback is an important stage in the implementation to give the user guidance throughout the system. The prototype couldn't handle voice feedback through Google Assistant so external feedback had to be played to the user to simulate the system. This is something that needs to be addressed in the implementation to give the user both visual and audio feedback and to be accessible for the users. There was no error handling supported in the prototype except for Googles own handling of errors. So, this is something that needs to be addressed and researched in a future iteration.

Even though the test results for the prototype were overwhelmingly positive the prototype was still just a concept when tested which means that further testing needs to be done with a finished application to fully evaluate the application.

An interesting request from one of the participants was the option of customising each command and phrases, which would require a settings option that also needs to be implemented, where each can be customised to whatever command the user finds suitable and which objects the user would like to have set. The feature of stacked commands is not limited to just a home system but can also be used to visualise other systems as well. By making each command completely customisable for each individual person it can be used to illustrate car interfaces, which is a whole other market that could be used. By finding new markets, stacked commands can be a catalyst into expanding the VUI system as it has endless possibilities for the visualisation of an endless number of products.

Chapter 6
Conclusion

This chapter contains the conclusion section of the Master's Thesis. In this chapter the evaluations from the previous section are answered and conclusions are drawn.

The qualities that a VUI has is its ability to help bridge and extend the user's ability to traverse a system, not excluding any disabilities. VUI is also growing exponentially and is catching up to the GUI that has dominated the field for many years.

Improving the current VPAs is a challenge because of the rapid implementation of new features being added to the system. So far, voice based interfaces have not been able to support multiple objects that can be visualised in a system, this is where stack commands could be integrated. By stacking commands and executing them almost simultaneously it is possible to visualise a whole system of objects that currently only shows objects individually.

By improving the interaction between user and machine stack commands can help with efficiency and relieve the mental stress load that users have in everyday life. The user interface was designed to be usable and accessible to everyone, including those with impairments. The concept of stacked commands has the advantage that it can be applied to any system with multiple objects needing visualisation, either using an audio or visual system.

Bibliography

[1] Jason Cipriani. *The Complete List of 'OK, Google' Commands*. 2016. URL: https://www.cnet.com/how-to/complete-list-of-ok-google-commands/ (visited on 08/20/2017).

[2] National Research Council. *Funding a Revolution: Government Support for Computing Research*. National Academies Press, 1999.

[3] Felix Fichter. *Digital Assistants – Always at Your Service*. 2016. URL: https://www.statista.com/chart/5621/users-of-virtual-digital-assistants/ (visited on 08/20/2017).

[4] Thomas Hermann. *The Sonification Handbook*. Logos Verlag, Berlin, Germany, 2011.

[5] IBM. *IBM Shoebox*. 1976. URL: http://www-03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html (visited on 08/20/2017).

[6] Icons8. URL: https://icons8.com/ (visited on 09/02/2017).

[7] Bruce T Lowerre. *The HARPY Speech Recognition System*. Pittsburgh Pennsylvania, 1976.

[8] Lucidchart. URL: https://www.lucidchart.com/ (visited on 09/02/2017).

[9] Charlotte Magnusson and Kirsten Rassmus-Gröhn et al. *User study guidelines*. URL: http://www.certec.lth.se/fileadmin/certec/publikationer/HaptiMap_d12.pdf (visited on 08/20/2017).

[10] Isaac Newton. 1676. URL: https://en.wikiquote.org/wiki/Isaac_Newton (visited on 08/20/2017).

[11] Jakob Nielsen. *Voice Interfaces: Accessing the potential*. 2003. URL: https://www.nngroup.com/articles/voice-interfaces-assessing-the-potential/ (visited on 08/20/2017).

[12] Donald Norman. "Natural User Interfaces Are Not Natural". In: *Interactions* 17 (2010), pp. 6–10.

[13] Donald Norman. "Natural User Interfaces Are Not Natural". In: *Interactions* 17 (2010), pp. 8–9.

[14] Donald Norman. *The Design of Everyday Things*. New York : Basic Books, 1988.

[15] Ravihansa Rajapakse and Mashika Dias et al. "Audio User Interface for Visually Impaired Computer Users: in a Two Dimensional Audio Environment". In: *International Journal of Computer, Control and Information Engineering* 6 (2012), p. 5.

[16] Jeffrey Rubin and Dana Chisnell. *Handbook of Usability Testing: how to plan, design and conduct effective tests.* John Wiley Sons, Indiana, United States of America, 2008.

[17] Ben Schneiderman. *Designing the User Interface*. Pearson Education Inc, 1987.

[18] Ben Schoon. *Google Assistant now supports custom shortcuts for lengthy commands*. 2017. URL: https://9to5google.com/2017/05/17/google-assistant-now-supports-custom-shortcuts-for-length%20y-commands/ (visited on 10/15/2017).

[19] Helen Sharp. *Interaction Design: Beyond human-computer interaction*. 2nd Edition John Wiley Sons, Chichester, England., 2006.

[20] Sarah M Sincero. *Personal Interview Study*. URL: https://explorable.com/personal-interview-survey (visited on 08/20/2017).

[21] Usability.gov. *User Centered Design Basics*. URL: https://www.usability.gov/what-and-why/user-centered-design.html (visited on 08/20/2017).

Appendix A
Survey and Interviews

This is a survey that collects information on the use of Digital Assistants (eg. Siri and Google Assistant). The information to be collected will be anonymous and used as part of a Master's Thesis in Interaction Design.

How old are you?
(Checkbox) -18 , 19-30, 30-45, 45-65, 65+

What is your current occupation?
(Checkbox) Working, Student, Unemployed, Retired, Other

Do you have a smartphone or tablet? (iPhone, Android, Windows Phone), Other)
(Checkbox) Yes/No

What kind of operating system do you have on your phone?
(Checkbox) (iPhone, Android, Windows Phone), Other)

How often do you use your smartphone/tablet?
(Checkbox)
Every day, A few times per week, A few times per month, Infrequent use, Never

---

Did you know that there are voice actions on your phone such as Siri, Google Assistant, Google Now, Cortana?
(Checkbox) Yes/No

Do you use these actions?
(Checkbox) Yes/No

If not? Please explain
(Open ended answer)

How often do you use voice control actions?
(Checkbox)
Every day, A few times per week, A few times per month, Infrequent use, Never

Why do you use voice actions?
(Open Ended).

Which voice actions do you mainly use?
(Checkbox) + other (open ended)

Would you be using more voice actions if it was better?
(Checkbox) Yes/No

What would make voice actions better?
(Open Ended)

What would make you use voice actions more on your smartphone?
(Open ended)

Are there any additional voice actions that you would like to add? In that case what actions?
(Open ended)

Are you generally happy about the current voice controls that are on the market right now?
(Likert scale)

Why would you want to use voice control actions?
(Open ended)

When do you use voice control actions?
(Open ended)

In what environment can you see yourself using voice control actions?
(Checkbox) Home, Work, Car, Public transport i.e bus/train/etc

In what environment do you think it is not appropriate to use voice control actions?
(Checkbox) Home, Work, Car, Public transport i.e bus/train/etc

Appendix B
Test Plan

Goal
The development of a test plan is to get some structure of the test to minimise errors occurring during the test. By standardising a test, it can help to eliminate the human factor involved when testing a prototype. To be able to evaluate and draw conclusions from the prototype a test is required.

Problem Definitions

- Would this prototype help the development of VUI?
- Does it give the user enough feedback?
- Is the end result satisfying for the users?
- Would the user ever the product again if it was implemented fully?

Data to be collected
Qualitative data will be collected from the following questions in a debriefing:

- Do you believe that this prototype gives you a better understanding of the status of objects?
- Did the prototype give sufficient feedback?
- Did the prototype provide you with enough feedback to visualise the system?
- Do you feel like you were in control of the system and could reverse any actions?
- Would you use this system again?
- Would you prefer giving one command for everything instead of one complete command?

- How do you feel it went?

Quantitative data will be collected through

- How many errors recorded.

Selection
The test will be conducted on carefully selected people that give a wide range of people from different backgrounds and age brackets.

Setup
An introduction where the project is described and what the purpose and goals of the thesis are. A description of the prototype will also be given and what voice control actions are available for this prototype and explain that the phrases would be fully customisable to the individual person but not in this prototype.

During the test the tester will be asked to ”think aloud” in order to explain their actions and why they choose a specific command.

Task overview
A series of scenarios will be read out and for each scenario the tester will pick and choose from the available commands. When the tester has decided on which command they choose a visual feedback will be displayed on the screen and an audio feedback will be played.

After the test a questionnaire will be asked to be answered by the tester where they will be asked to give feedback on the prototype.

Test environment
The test will be performed at the home of the testers to ease the stress that might occur during the test.

How will the results be reported?
The results will be compiled, analysed and evaluated in the final report of the Master Thesis.

Task List - Scenarios

1. Home Status
   You are going on vacation, you've just hit the road after a hectic morning of packing up the kids and making sure you have got all the things for your holiday. As you're stuck in traffic on the highway you realised you forgot to check everything before you left. You are sitting in your car which is connected to your smartphone where you've been following the instructions on the navigation on the way to your holiday house. Since you have a very connected home you check the status of your home.

   "What would you do to check on your home?”

2. Secure property

   Now that you've discovered the status of your home, you realise that you have forgotten to turn on your alarm.

   "What is your next step?"

3. Power Saving Mode

   Now that you've secured your property whilst you are away, you've been sent an email notifying you of your upcoming power bill. To save on costs while you're away from home for an extended period.

   "What can you do to save on power at your home?

   You are now happy to spend time on your holiday in the knowledge that your home is secured and you are saving money whilst you're away.

4. I'm Coming Home

   On your way home from your holiday, you are thinking about what you need to do when you get home, but then you realise you can have a nice warm and cosy house waiting for you when you get there.

   "What command works best for getting things ready at home?"

   You are now less stressed about arriving home late with the family, knowing that things will be ready for you when you get there.

5. I'm Home

   As you pull up in the drive way of your home you realise you've packed the house keys in a bag somewhere in the car and the kids are asleep. To make things easier for you to get access to your home.

   "Which command would you use on your smartphone?"

6. Good Night

   Once the kids are placed in bed, and the car is unpacked from your holiday, you collapse in bed and realise you've left the door unlocked and the lights on downstairs. You are too tired to get out of bed.

   "What is the best command to use from the comfort of your bed to secure your home for the night?"