Adding Voice Interaction To A Wearable

Matthias Cordes

Supervisors:
Khiet Truong and Randy Klaassen

24. April 2018

University of Twente
Faculty of Electrical Engineering, Mathematics and Computer Science
Program: Creative Technology

Abstract

This report is about the bachelor graduation project "Adding voice interaction to a wearable". It documents the steps that have been conducted in order to answer the research question: "How should voice interactions be designed for wearables in the context of habitual change?". A high potential is being seen in the field of voice interaction technology and many opportunities can be spotted on how this technology can improve our lives by making conducting certain tasks more efficient. After firstly investigating, what the the state of the art of different context relevant subjects like voice interaction technology, voice interaction design, wearables and habitual change is, an initial project idea has been developed within an ideation phase. The generated idea has then been used as the foundation for a first user test using a lo-fi prototype. After that the user and system requirements have been specified in the specification phase. Subsequently, a second prototype has been developed in the realization phase. Finally, in the evaluation phase, the hi-fi prototype has been used in the context of a user study. The results of the study have been discussed and recommendations for future work been made.

Acknowledgements

I want to thank my project supervisors Khiet Truong and Randy Klaassen for their guidance, constructive criticism and friendly advice during the time of working on this project. Besides, I want to thank all the participants of the conducted user studies within this project for their time and their trust.

Matthias Cordes

Table of Contents

List of Figures

List of Tables

1. Introduction

Voice technology has become increasingly popular not only as a way of retrieving information from the web, but also as a tool to effectively communicate with machines. When thinking of voice technology, smart speakers such as Amazon Echo or Google Home are often being associated with it. These devices are equipped with virtual assistants that have been available for smartphones for quite some time already, but are rarely being used by the consumer (Milanesi, 2016). Voice interactions often felt awkward and are rather an entertaining gadget than a serious solution for the communication between humans and their smartphones. Nowadays, companies like Amazon, Google, Microsoft and Apple compete on the technology market by developing different products, equipped with their constantly improving intelligent assistants. These are often being used with other devices in the household in order to control for example lights, heating or even the coffee machine. Due to the increasing amount of functionalities and the efficient accessibility of voice controlled assistants, they might have a strong impact on the way we communicate with intelligent machines in everyday life. Current market forecasts can support these assumption, suggesting that the majority of households in the US will own a virtual home assistant (Perez, 2017).

Given that technology is evolving quickly and machines consequently become more intelligent over time, trending virtual assistants will be used in many different fields of application. Besides, they will be used on many devices, like for example smartwatches. However, until today voice technology is rarely being used in wearable devices. Designers and developers for coaching technology currently hardly take the possibilities to implement voice assistants in wearables into account. There is a vast number of emerging areas, where voice interaction can be very useful, since it makes conducting certain tasks more efficient and users will consequently save their most valuable resource, time.

Especially in the area of health and behaviour science its is very likely that there will be opportunities for voice technology to improve the users quality of life. In recent decades the interest in personal developemnt is becoming increasingly popular. Although, changing one's behavior can be very challenging. In the past ten years much progress in understanding the

neurological and psychological foundations of habits have been made (Duhigg, 2014, p. 224). It is known now, how much impact habits can have on individuals and how they can be changed. By identifying routines we can influence how much we eat, get us to do more sports, work more efficient or eat healthier (Duhigg, 2014, p. 224). Valuable data can be retrieved through voice interaction and in combination with modern technology, intuitive interactions and latest insights on habits, it might become way easier to achieve our goals by changing our habits for good.

The challenge of this project is to make use of the microphone in smartwatches in a beneficial way in order to assist the user by changing bad habits, which leads to the research question: 'How should voice interactions be designed for wearables in the context of habitual change.' In the first chapter of this report the current state of the art on the topic of this project will be explored. Whereby, the focus will be on voice interaction technology, voice interaction design, wearables and the science of behaviour change. Consequently, the the design and project execution will be documented. In the end there will be a evaluation with an outlook on possible future work.

2. State of the Art

The first section of this chapter will discuss the state of the art of the following domains: Voice interaction technology, voice interaction design, wearable technology and habitual change. The investigations will help to understand what kind of technology is available today and what has been done in these fields already. Subsequently the findings will be discussed in the second section. In the third and final section of this chapter the findings will be used as a foundation for the idea generation of the project (chapter 3).

2.1. Background research

2.1.1. Voice Interaction Technology

In order to get a better understanding of the broad subject of voice technology, it is essential to get an overview of what kind of voice interaction technologies are already out there and how they are being designed. Voice interaction technology is here referred to as the ability of a technical device using natural language to communicate with humans. Voice technology that most people nowadays are familiar with are interactions with intelligent virtual assistants integrated in smart home devices or smartphones, such as apple's Siri, amazon's Alexa or the Google Assistant. Most of these assistants work in similar ways and combine several technologies. A voice recognition technology, a text to speech technology, a natural language processing engine and a backend service for processing capabilities and accessing large amounts of data, as stated by Matic (Matić, 2017, p. 1). Reehal (2016, p. 2) highlights that a typical interaction can be divided in the following five steps: [1] The voice command by the user is being recorded as an audio file and then [2] gets sent to a data centre, where it gets converted from speech into text. [3] The data gets analysed on powerful computers (instead of on the mobile device with limited resources). The data can be used to continuously improve the service. [4] Keywords are being identified and the system tries to guide the conversation in a specific directions in order to achieve the users goal. [5] In the final step, after the request has been processed, the voice assistant converts the result into text that is then spoken to the user. A process that is also being called natural language generation. Summarized, voice interaction

technology works in a way where a request by the user gets send to the cloud, is them being processed and an response given by the assistant to the user.

Various fields of applications of voice technology can give valuable ideas of how this technology might be used in the the future: M.J. Callagahan for example made use of the assistant alexa in an electrical engineering laboratory, which can be used to guide students, provide teaching resources and help controlling instrumentations. However, if this technology will be useful in practice did not become clear yet. Only suggestions for further explorations have been made by the author (M.J. Callagahan et al., p. 670, p. 670). Attempts have also been made in order to help elderlies, suffering for example from dementia. ‘Robin’ is a skill developed for the virtual assistant alexa that is meant to enable independence for individuals with cognitive disabilities by helping with their routines. “Robin will alleviate the pressure on caregiver to be continuously amiable for day-today assistance” (Carol et al., 2017, p. 51). From a study conducted by the MIT Media Lab with several kids of different ages it turned out that the children perceive virtual assistants differently than adults do. During the study the participants believed that they could teach the agents and learn from them (Druga, 2017, p. 5). This leads to the question, how these voice controlled virtual assistants can help children as learning companions in the near future. Unfortunately, for all these mentioned examples, the practicality of the applications is doubtful.

As Susan L. Hura mentions, one field of application where voice interaction technology is becoming widespread is in vehicles (and several other niche markets like medicine, warehouse logistics, etc.) (Hura 2017, p. 213). Also Chun-Cheng Chan states that voice is increasingly being used in vehicles. A major issue with voice control in cars are occurring imprecations such as response time delays and speech recognition errors, which increases the cognitive workload for the user and therefore can cause distraction (Chang, 2017, p. 43). Fang Chen (2010, p. 198) agrees with this statement when pointing out that '(...) it may increase the mental workload of the driver(...)'. Although, Chen further mentioned that the vehicle industry is experiencing a 'fast-growing interest in speech technology'. While driving, eyes can be kept on the street and hands on the steering wheel, therefore it is considered to increase driving safety. Besides, it would save a lot of space, since only a noise cancellation microphone would be needed (Chen et al., 2010, p. 196). Another important point being made is that there is no

standard methods to measure the usability of interactive speech-based systems in vehicles yet (Chen et al., 2010, p. 199). Applications in cars seem to be promising, although it is important that the interactions become more human-like and machines more intelligent in order to make interactions effective. A proper visual display could be given in order to make complex speech navigations easier. Thereby, the cognitive load for the users will be lowered and the interactions be more natural and less frustrating. If these challenges will be tackled, proposed concepts might find effective use in practice.

In times where social robots are a rapidly-expanding category of robots, the role of voice interaction technology should not be neglected. Soon, robots might have to potential to help us in our everyday lives. Already today they are partly being used in education, training and in healthcare (Markowitz, 2017, p. 2). Scientists try to find out how to increase the acceptance of social robots by making them more human-like. Therefore, the communication via voice is an important skill to take into account. What has been found out is that social robots that have a voice which sounds human-like and besides sounds similar to the gender of the user itself, have a higher acceptance (Eyssel et al, 2012, p. 126). Another important insight is the impact of affect in social robots. Within a study by Scheutz et al. (2006) it has been shown that if a robot is affective and allowed to express urgency, the team performance with a human was better. It can be seen that voice technology, is also in the field of social robots a feature that is highly in demand. It is a valuable asset that is relevant in order to get a higher acceptance for robotos and to integrate them in our everyday life. Research in this field provides useful cues in order to design voice user interfaces. Still, it seems like the highest advantages in the field of voice have been made in the field of virtual assistants.

Conducting literature research on voice interaction technology, it has mostly been looked at automatic speech recognition within this report so far. It should be noticed that also progress in the field of speech emotion recognition is being made and that it is getting increasing attention. Understanding the emotional state of the speaker can be very valuable and be applied for example in call centers, computer tutorial applications, cars (Schuller, 2004, p.1) or as a diagnostic tool for therapists (Landau, 2008, p. 835). However, the technology still seems to face many difficulties before being applied effectively. As mentioned by Chul Min Lee (2005, p. 293) "One main difficulty comes from the fact that there is a lack of complete understanding of

emotions in human minds" and further "Agreement among researchers is a prerequisite to satisfaction in attempting to build an effective machine for the task of automatic emotion recognition. (Chul Min Lee, 2005, p. 293)". Despite all challenges, speech emotion recognition seems to be an exciting and new field that will ad a lot the to the performance of computers understanding the most natural way of human communication.

2.1.2. Voice Interaction Design

Voice Interaction Technology is one part that is important in order the enable natural conversations with machines. Another important factor is the voice interaction design. How will a virtual assistant be perceived? What are the sentence structures? What paths can a conversation take? Those are design questions that should be answered. The design process for voice user interfaces differs very much from the one of a regular user interfaces. UIs that are for example screen based, always have limitations (there is a limited amount of options available that the user can choose from) while the possible voice inputs from users seem endless. Therefore, the challenge is to design dialogues in a natural and for the user beneficial way, offering many different ways of dialogues. In general there are a few guidelines to follow in order to understand when it makes sense to use voice interaction. As Susan L. (2017, p. 213) states in “eyes-busy” or “hands-busy” situations voice technology can be very beneficial, since a user usually needs hands and eyes in order to interact with a regular UI. Further, it is being stated that “The aim is to avoid using speech technology for its own sake, and instead look for opportunities in which speech provide distinct value to the user […] (2017, p. 216)”. This statement implies that there should always be a goal before designing an interaction. Farzaneh Nasirian emphasised with a conducted study, how important interaction design for voice assistant systems is, in order to create trust between user and the system: “[…] we found that interaction quality is the most important factor of quality which builds trust in users, and as a result they intend to use the VASs“ (voice assistant systems) (Nasirian, 2017 p. 8). There are many skills available that assistants provide, but they depend on the personal assistant implantation and its purpose (Lopez, 2017). When designing VUIs it is highly important to consider, where the interaction is taking place, in what context and on what device. One more challenge that might also occur is that skills from different domains are needed to design the interactions. Speech system developers and also human factors professionals are needed.

“Speech system developers are often not accustomed to working with human factors professionals” (Hura, 2018, p. 2014). Besides, when testing current voice interaction systems, Aseffi et al. suggests that delay, affected by packet loss and jitter can occur in recent systems. A problem that should be solved in order to enable real time cloud speech recognition applications for “more critical tasks (Aseffi et al, 2015, p. 6)“. From this different perspectives on voice interaction it can be conclude that technological advanced still need to be me made in order to design more natural interactions for different devices such as home assistants, smartphones or smartwatches.

2.1.3. Wearable Technology

Before looking into the field of wearable technology it should be clear, what exactly a wearble is. According to the oxford dictionary this is: “Denoting or relating to a computer or other electronic device that is small or light enough to be worn or carried on one's body.”[1] There are many fields of application, where a wearable technology can be beneficial in combination with voice assistants. For example: If a user wants to interact with an assistant while holding a child's hand and walking down a busy street, it would be desirable to access the voice assistant via a wearable like for example a smart watch instead of a smartphone (Kubo, 2017, p. 1). A smartwatch is a mobile device with a touchscreen display, designed to be worn on the wrist.[2] There it needs to be taken into account that a cross-device interaction, for instance with the smartphone screening additional large amount of information is not possible. Besides, there are different kinds of data available that can be valuable for the interaction. A smartwatch for example is often equipped with numerous sensors (Kubo, 2017, p. 2) that could be used to track the users arm posture or other context information. As Vinciarelli (2015) states, there is a growing number of applications that can "track user activity, sleeping and eating habits and covert and overt signals such as blood pressure, heart rate, skin temperature, speech, location, movement […]” (Vinciarelli, 2015, p. 11). Also the article of Pantelopoulos (2010) gives a comprehensive overview of how wearable technology can be used in the context of health monitoring. Pantelopoulos came to the conclusion “WHMS (wearable health-monitoring systems) have the potential to revolutionize healthcare by providing low-cost solutions for

[1] Source: https://en.oxforddictionaries.com/definition/wearable
[2] Source: https://en.oxforddictionaries.com/definition/smartwatch

ubiquitous, all-day, unobtrusive personal health- monitoring and are expected to enable early detection and better treatment of various medical conditions as well as disease prevention and better understanding and self-management of chronic diseases. (Pantelopoulos, 2010, p. 9)“ However, there are several technical and design challenges that need be solved. In conclusion it can be said that there is much potential for wearable technology to improve our lives in very different situations. Two main fields of applications have been identified: Firstly, whenever it is more convenient to use a wearable instead of a non-wearable technical device such as for example smartphone. Secondly, in the context of health, whenever biosensing can help to track the patiences health.

2.1.4. Habitual Change

Another field worth to look into is habitual change. Wearables allow easy and efficient communication between a user and a virtual assistant and therefore enables new ways of assistance. Helping the user to track or even change habits would be one example. Habits can have a high impact on society and the lives of individuals. Nervous habits for example are psychological, which are often highly resistant and can cause medical problems, as stated in a study by N.H. Azrin (Azrin, 1973). Azrin also clarifies within an experiment, how important it is to identify habits, in order to help people eliminating them. Results show that there is a simple method available in order to treat habits. For 12 clients participating in the study with different nature of habits (fingernail-biting, head shaking, gumsucking, etc.), the habit of each participant has been reduced by at least 90 percent. Research conducted by Caroline Free, showed that several interventions have been made in the field of technology-based health behavior change. In total there were 14 interventions been stated, but many of them did not result in a statistically significant effect of change (Free, 2013, p. 28). In another study conducted by Verplanken (1999, p. 601) it has been demonstrated that implementation intentions (here more general denoted as habits) to eat healthier were effectively be established. There are concrete methods available to change human behaviour, habits specifically. Many attempts have been made already, but findings provide mixed evidence for the effectiveness of interventions. Attempts have also been made in developing smartphone applications that help the user to track their

habits. Examples are the applications ‘Momentum Habit Tracker’[3], ’Productive - Habit Tracker’[4], ’Strides: Habit Tracker’[5] or ‘Done: A Simple Habit Tracker’[6]. These apps all work in similar ways and often include goal setting and reminder functionalities. However, the do not make use of advanced psychological methods to change them. In conclusion it can be said, that several frameworks to change habits exist, although it is not always clear in what situation they will effectively help the user. Besides, functionalities of current apps can track habits but have a lack of functionality that actually help to change them.

2.2. Discussion and Problem Analysis

The state of the art and the future of voice technology seem to be very promising for designing interactions for many different fields of application. When getting more familiar with the subject, it becomes clear that virtual home assistants such as Amazon Alexa and the Google Assistant are the latest trends in the world of voice interaction technology. They are easy to use and can be extended by third party services, often referred to as skills[7] or actions[8]. Research shows that there are many concrete examples of application for voice interaction technology. However, literature on voice interaction between users and wearables, such as smartwatches is rare, which is not surprising, since it is a very new field and voice interaction technology only recently gained a lot of popularity. Although, it might take some time, till voice interaction with machines will reach a state of enabling natural human-like conversation, there are clearly some fields where current voice interaction technology can be handy. So far it seems that there are many attempts being made to push this technology, in order to discover situations and ways to make the technology beneficial. Not only the technology itself (speech processing, language understanding, etc.) is crucial for a great user experience, but also the way voice user interfaces are being designed. In order to create a great user experience, it is essential to design the interaction as smooth and effective as possible so that they help to acquire the user's goal. Within this project that will be the main challenge.

[3] Source: https://itunes.apple.com/us/app/momentum-habit-tracker-routines-goals-rituals/id946923599?mt=8
[4] Source: https://itunes.apple.com/us/app/productive-habit-tracker/id983826477?mt=8
[5] Source: https://itunes.apple.com/us/app/strides-habit-tracker/id672401817?mt=8
[6] Source: https://itunes.apple.com/us/app/done-a-simple-habit-tracker/id1103961876?mt=8
[7] Source: https://developer.amazon.com/alexa-skills-kit
[8] Source: https://developers.google.com/actions/

There are a some benefits of interacting with a smartwatch over interacting with a smartphone. One outstanding advantage is that smartwatches are very easily accessible, especially when using voice. While having to operate smartphones using hands and often also the eyes, a voice controlled smartwatch would more easily be accessible. Today not many tools or guidelines are available in order to design and test voice user interfaces, which makes the design process not only more challenging, but also very valuable for future research in the field of HCI. So far, mainly documentations by companies inventing virtual assistants provide useful guidance in order to design voice user interfaces. Although, a few guidelines could have been identified that can be beneficial for the application of voice interaction technology: Firstly, voice interaction works very well in situations where the user can not use its hands or eyes to interact with a machine (in healthcare, people with disability, people conducting practical work). Secondly, it is more efficient, if the demanded cognitive load is small and the interaction faster than with a traditional UI (i.e. typing messages using a virtual keyboard).

While the subject of personal development and hence habitual change is generally getting more popular, having a virtual assistant that can help the user by forming or changing habits might be very beneficial. For example losing the habit of chewing fingernails, or forming the habit of going running each morning. But these are just a couple of very specific examples where a voice controlled coaching technology might be helpful. Books like 'The power of habit'[9] by Charles Duhigg give a great overview on studies that have been conducted in the field of habit change and serve as a good foundation to design interactions between humans and smartwatches in the specific context of behavioral change. From the current standpoint, it can be investigated how these technologies and methods can be applied in order to create the initial design idea for a project.

2.3. Conclusion

After the conduction of research on the state of the art of the previously discussed domains, some relevant information have been discovered. The gained insights help to make a conclusion about the benefits and limitations of latest voice interaction technology and design in

[9] Source: http://charlesduhigg.com/the-power-of-habit/

the given context. When looking at voice technology, there are many applications out there and many advances being made in the field. Still, systems are not perfect and have a long way to go to enable fluent and natural conversations. How practical current applications are, is often doubtful, since virtual assistants are simply not intelligent enough to communicate in a way humans do. Although, when looking at the progress being made, it can be assumed that this is a problem that might get solved soon, at least to a certain extent. Speech emotion recognition seems promising still too limited in order be be relevant for the topic of this project. Research methods for designing voice interactions are still very rare. Especially, when it comes to designing voice user interfaces for wearables, like for example smartwatches, there has not been much research being done yet. However, situations have been identified, where the interaction with wearables, like smartwatches can be very useful (i.e. hands-busy or eyes-busy situations). Methods from psychology for behaviour change have not yet been applied in wearable devices like smart watches yet.

Based on previous discussion in the different domains, the following research question has been developed: 'How should voice interactions be designed for wearables in the context of habitual change.' The goal will be to find out, how current voice technology can be applied in order to design a friendly voice user interface for a smartwatches in the context habitual change, that helps to improve lifestyle and therefore the quality of the users health.

3. Ideation

3.1. Introduction

Based on the identified research question “How should voice interactions be designed for wearables in the context of habitual change?” the process of designing the envisioned interaction will begin. All the steps that have been conducted in order develop the product idea will be stated within this chapter. Afterwards, initial user and system requirements will be defined. Based on the findings, new user and system requirements will be developed and a lo-fi prototype for a first user test be created.

3.2. Creative Idea

After comparing different tools in order to design voice interactions, there was one option that seemed to be most promising in order to efficiently achieve the goal of designing the interaction and building a prototype: The is tool called ‘Google Dialogflow'[10]. The method used on the smartwatch to help the user changing his or her habit is based on the book ‘The power of habit’ by Charles Duhigg. Duhigg explains that in order to change a person's habit, it is important to identify its underlying neural loop. This loop consists out of three different elements: A cue that triggers the user to follow his or her habit, a routine and a reward (Duhigg, 2012, p. 336). In order to change a habit, it can be helpful to first isolate the cue of the it and therefore ask the user a few questions about his situation when currently having the desire to follow his or her habit. For this project these questions will be asked by an assistant on smartwatch (instead of pen and paper being used by the user). Below an overview of the steps that need to be conducted in order to change a habit is given:

1. Finding out what the reward is that you are looking for (by experimenting with different ones)
    a. What is the actual desire? Experiment with different rewards, by changing routine a little.
    b. If you try 4-5 different rewards, write them down and search for patterns like feelings or thoughts. Check 15 minutes later if you still have the desire (the writing forces attention).
    c. Find out what the REAL desire (and reward) is.

2. Isolating the cue
    a. Often in one of the five categories: Location, time, emotional state, other people or a preceding activity. So everyday you should answer these questions:
        i. Where are you?
        ii. What time is it?
        iii. How do you feel?
        iv. Who else is with you?
        v. What did you just do?
    b. After a view days testing it should be clear what's the cue is.

3. Making a plan

[10] Dialog Flow is an end-to-end development suite for building conversational interfaces for websites, mobile applications and popular messaging platforms.

a. Change the routine (add alarm, to make sure you do not forget it)
b. Usually will feel better when it works
c. Becomes a habit over time

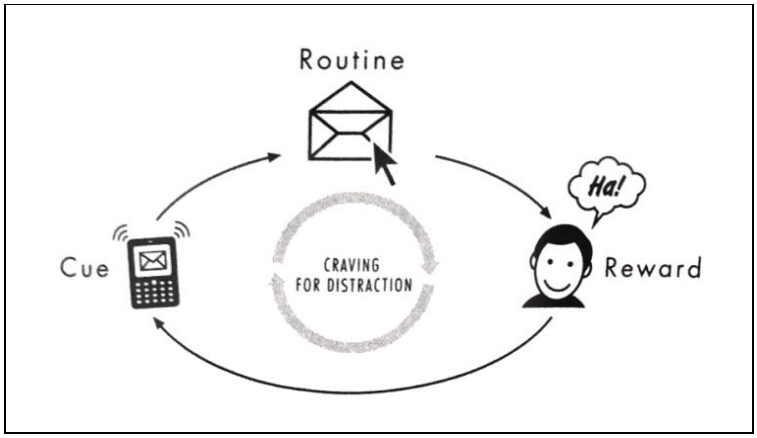

Figure 2: Charles Duhigg's Conception of the pepsodent habit loop[11]

3.3. User and System Requirements

When designing a new product it is important to understand, who the main stakeholders will be. In the context of this project this is basically everyone, who tries to change a specific habit and seeks help by an assistant. Typical habits could be for example exercising frequently, meditating every morning, starting a healthy diet or reading a book for 30 minutes a day. Of course, users might also want to use the product in order to change negative habits, like for example biting fingernails, following an unhealthy diet, smoking or drinking alcohol. No matter what the habit might be, the goal is to assist the user in achieving his or her goals and improve his quality of life. Another important point is that the interaction should be efficient. The interaction with the assistant on a smartwatch should be more convenient and let the user save time, compared to for example getting help by using a notebook, a real life assistant (i.e. therapist, coach, etc.) or a text based assistant on a smartwatch.

[11] Source: http://www.slate.com/articles/arts/culturebox/2012/02/an_excerpt_from_charles_duhigg_s_the_power_of_habit_.html?via=gdpr-consent

It is important that the system can be used by as many diverse people as possible. No matter what age, gender, familiarity with new technology and so on. The technology should be self explaining. Potential users should understand, how to interact with the assistant on the smartwatch. Different tools are available on the market that offer the possibility to easily design the system. The two most popular tools are Google Dialogflow (as mentioned earlier) and the Alexa Skills Kit[12]. These tools do not only help to design voice interactions, but also allow to develop and deploy them. A less complex and therefore more user friendly alternative would be Sayspring[13], a tool to create voice interfaces for Amazon Alexa and Google Assistant. While writing this report, many new tools are in the making and will probably be available soon, PullString[14] is one example. After comparing different tools, it initially was the plan to work with Sayspring for creating a frist lo-fi prototype. However, after getting more familiar with the platform, it turned out that in terms of functionalities, this program has been too limited in order to effectively create voice user interfaces and fulfill the system requirements. Consequently, Google Dialogow has been a better choice for creating VUIs, since there are less limitations and many options to test the interactions. Further requirements that the system should fulfill are the ability to store the information given by the user, to understand what the user is saying and also make the user feel comfortable interacting with the assistant.

3.4. Lo-Fi Prototype

3.4.1. Description

In this section it will be investigated how the user will interact with the a virtual assistant using latest voice interaction technology. Therefore, a simple flow diagram[15] (section 3.4.2) has been designed first, in order to help to visualize the idea of possible interactions. After that a lo-fi prototype has been designed. The goal for the prototype is to conduct a first, quick user study. The interactions will be designed using Google Dialogflow. The Google Dialogflow project will then be deployed to a smartphone and be accessed by the Google Assistant smartphone

[12] Source: https://developer.amazon.com/alexa-skills-kit
[13] Source: https://www.sayspring.com/
[14] Source: https://www.pullstring.com/
[15] Source: https://www.oreilly.com/ideas/basic-principles-for-designing-voice-user-interfaces

application[16]. The users will have to answer a few simple questions asked by a virtual assistant, in this case the google assistant. The questions that are being asked are relevant for the context of habitual change. The methods for changing habits will be based on the one used in the book by Hopkins Duhigg mentioned earlier.

3.4.2. The Flow Diagram

In order to design voice user interfaces it can be helpful to first visualize all the possible paths that can be taken through a VUI system. Below it can be seen how an interaction with an assistant on a smartwatch in the context of habitual change could look like. The flow diagram includes dialogues of different stages of the procedure to change a habit (more details are being provided in the next chapter). Possible prompts that can be used by the assistant and by the user are shown in the diagram as well.

[16] Source: https://itunes.apple.com/de/app/google-assistant/id1220976145?l=en&mt=8

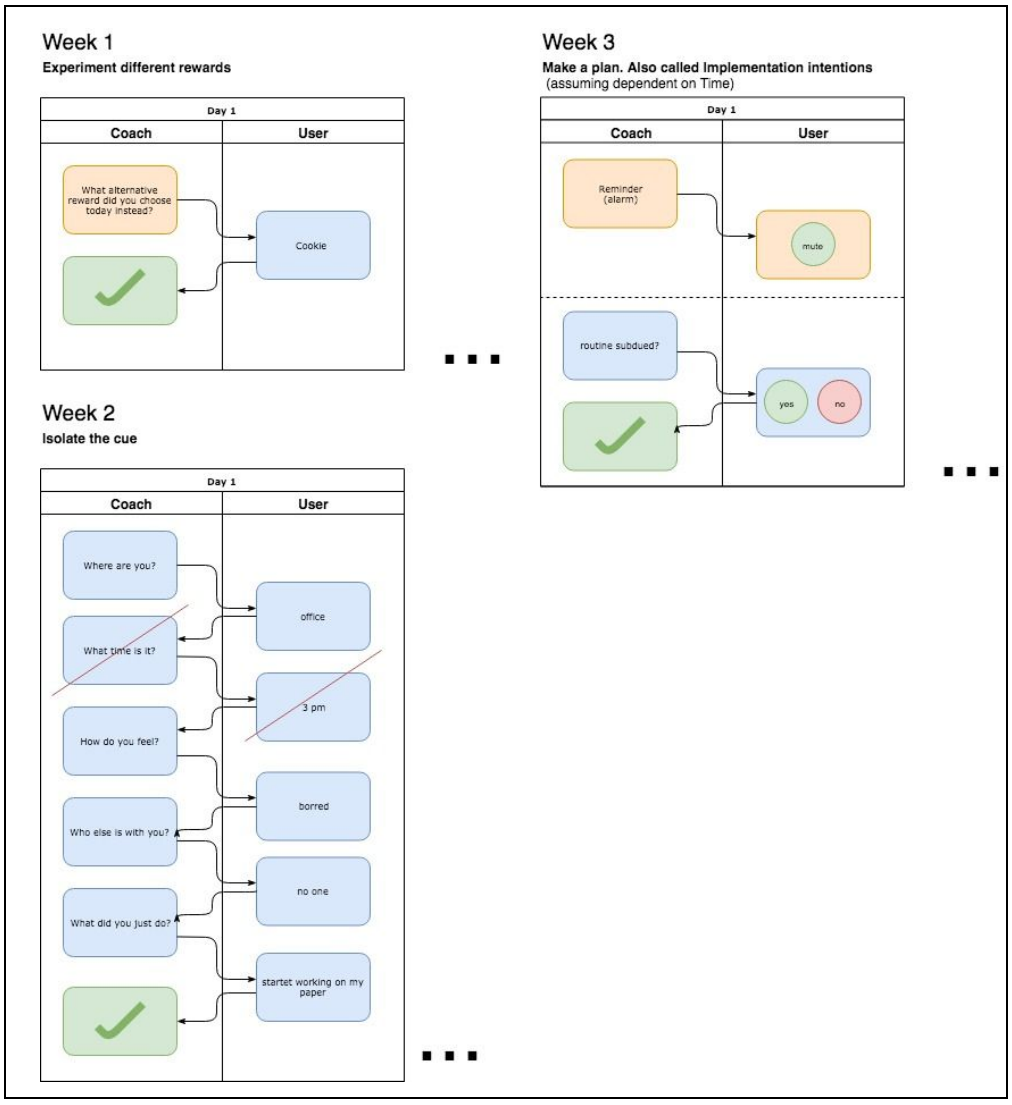

Figure 1: Flow diagram for the voice interaction with a smartwatch.

Designing a flow diagram turned out to be a very helpful method in order to efficiently visualize, how the interactions with a virtual voice assistant could look like. It shows what kind of text responses can be given by the assistant, prompts the user might use and in what paths are possible, if interactions become more complexe. Another advantage is that it can be shown how visual feedback of a device could like like. Also, it is illustrates how the regular UI on the screen of the smartwatch could look like (i.e. week three - the 'yes' and 'no' buttons).

3.4.3 Setup

There are many different ways, to formulate one and the same question. Therefore, it will be explored what voice dialogue strategie might work best by comparing two different dialogue strategies within a user test. One strategy will be designed in a way that makes the interaction very direct and efficient, using only a small amount of words. The other version will be designed in a more human-like way, containing more so called “conversational markers” (Pearl, 2017, p. 39) that make the conversation more engaging. Besides, the assistant will always acknowledge answers by the user with short phrases, instead of jumping straight to the next question. In the table below the strategies will be used, can be seen:

Prompts for Strategy 1	Prompts for Strategy 2
- “Where are you?” - “How do you feel?” - “Who else is with you?” - “What did you just do?” - “Bye.”	- “Hi. I’ll be asking you a few questions about your current situation: First, where are you right now?” - “All right...and how do you feel?” - “Is there anyone else being with you right now?” - “Got it...Last question, can you tell what you are doing?” - “Okay. That’s it for now...talk to you soon!”

Table 1. Prompts for two different dialogues strategies.

3.5. User Study

3.5.1. Goal

The goal of the study is to find out, how well the voice interaction will work in general and what problems might occur. This first round of validation will help to see where and how deeper exploration, analysis, and testing should be conducted. The Two different dialogue strategies will be used in order to find out, which of the two designs works best in the given context. Will users will know how and when they can speak to the assistant? How might users react to errors in the interaction and recover from them? The results gathered from the study will then help to update the system requirements and user requirements.

3.5.2. Procedure

The chosen environment, where the study will be conducted is a quiet room, so the user does not get distracted by surroundings like for example other people or background noises. The user will be seated on a chair with a table in front of him and interact with the agent on a smartphone. This will help to make comparisons later on in order to identify what is specifically important when designing VUIs for a smartwatch. The phone will be placed on the table in front of the user, ready to get interacted with. The conductor of the user study will sit at another location in the room, in order to observe but not distract the user. A camera will be used to gain more insights about the user behaviour during the interaction. The following steps will be conducted during the user study:

1. Before the test begins, the user will be asked to sit down on a chair and to sign a general consent form before participating. After that all necessary preparations for the interaction will be conducted. The user will be provided with following information:

   "This study is about interacting with a virtual assistant that helps the user to change their behavior by adopting good habits or loosing bad ones. The goal is to find out how to make the voice interaction with the agent that will be used on a smartwatch as efficient and intuitive as possible. Your participation will help to achieve this goal. In this session you will be working with a simple prototype in order to test different dialogue strategies. You will be asked to perform tasks that typical users of the assistant might have to conduct, such as answering questions about your current location or mood. After the user test you will be asked some general questions by the researcher about how the interaction went."

   Besides, there will be some additional general information about the study be provided, which can be found in the appendix of this report.

2. As soon as the user starts to interact with the prototype, the prompts that have been defined will be used to enable a dialogue between smartwatch and user in form of an

dialogue. Everytime the assistant recognizes an answer given by the user, it will respond with a follow-up questions till every question has been successfully answered or the user got stock at a specific step of the conversation.

3. After the interaction has been completed, the user will be asked a few interview questions by the conductor of the experiment. The questions will be asked in order to get more valuable feedback from the user on how he felt during the interaction and what his thoughts are. This is especially important when it comes to testing user interfaces, since strategies like the 'think out loud' strategy are difficult to apply, if a virtual assistant is listening. The following questions are the ones that will be asked:

    - "How old are you?"
    - "What is your profession or study?"
    - "How did the overall interaction go?"
    - "Was it clear to understand what the assistant wanted you to do?"
    - "Was there something specific challenging?"
    - "What do you think about the questions, that you have been asked?"
    - "How did the assistant react, if it did not understand you?"
    - "Was it complicated to get back on track if the interaction failed?"
    - "Do you think the assistant was friendly?"
    - "Did something annoy you?"
    - "Did you notice a difference between the two strategies?"
    - "If yes, what way of interaction would you prefer and why?"

3.5.3. Participants

For the user study five participants have been randomly selected and asked to participate. Five users should be enough in order to find most of the usability problems[17]. It can be assumed that all the participants all have a general understanding about technology. The age range of the participants is from 21 to 23 years and all of them are students in Netherlands.

[17] Source: Source: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

However, there nationalities and gender differs. The interaction with the assistant will take place in english.

Participant	Age	Gender	Study / profession
1	23	female	Creative Technology
2	23	female	Creative Technology
3	23	female	Communication Science
4	21	female	Psychology
5	22	male	Financing and Accounting

Table 2: Demographics of participants that participated in the ideation phase user study.

3.6.4. Results

3.6.4.1. Usability Test

The results of the user study are listed in the table below. It shows, how each of the five users recovered from errors during the interaction. Within this usability test everytime the interaction does not work correctly it has been referred to as an error. For example: After the assistant did not recognize the answer of participant number one during the test of the first dialogue strategy, the user simple repeated the answer. Besides, some more general problems that occurred during the interaction are mentioned in the column on the very right.

Participant	Error recovery	General remarks
Dialogue Strategy 1		
1	User gives same answer several times.	-
2	User gives same answer several times.	-
3	User phrases answer differently.	The final question has not been understood by the assistant.
4	User phrases answer differently.	The final question has not been understood by the assistant. Besides, the user has been insecure and wondering if the assistant would understand the answer.
5	User phrases answer differently.	
Dialogue Strategy 2		

1	Assistant leaves conversation.	When the participant answers the question 'Is there anyone with you right now?' The answer was a simple 'yes', which does not give detailed information.
2	Assistant leaves conversation at final question.	-
3	The user is not sure about, how to indicate that he will start talking again, after the assistant stopped listening.	The assistant left the conversation and started googling information.
4	-	The user gets confused, when the assistant did not understand the answer and simply stopped having a conversation.
5	Assistant leaves conversation at final question.	The user knows exactly how to indicate on the UI that he will speak, after the assistant stopped listening (even though not being familiar with the assistant).

Table 3: Results of first user study.

The usability test helped to gain many new insights on how users behave when interacting with a virtual assistant. There are a few issues have been revealed that should be solved in order to make the interaction more fluent and less frustrating for the user. The user study showed, how users react and what errors occurred during the interaction therefore should be improved.

3.5.4.2. Interview

From the conducted interviews there were more relevant information gained by the answers of the participants. The most important ones are stated in the table below (table 4):

Participant	Feedback
1	- It was not possible to get back on track of the conversation - Assistant has been perceived as friendly - There was no big or relevant difference between two types of strategies
2	- It was not awkward to communicate with the assistants - The Intentens by assistant were clear - "Should I say the same again?" or "Should I rephrase" - The feedback on an error was clear - Communication might be annoying on daily bases - No difference noticed between two dialogue strategies
3	- It was complicated to get back on track - The assistant was not friendly - "I did not know what to do at all" - The short introduction by second dialogue strategy very valuable for the given context

4	- The context of the questions not so clear - Not sure what to say, if assistant did not understand - Its annoying, if assistant does not understand reply - Second dialogue strategy: 'More context, more likable' - Still, context of questions unclear - "I want to say …, but is it going to understand me?"
5	- Not easy to get back on track of the conversation - Friendly assistant - Not focused on different style of asking questions, therefore not relevant

Table 4: Relevant feedback by user study participants during the interview session.

3.6. Conclusion

The ideation phase helped to develop the idea for the envisioned interaction based on the research question defined earlier. The goal of the study was to find out, how well the voice interaction will work in general and what problems might occur. A first simple prototype has been built using google dialog flow in order to conduct a first explorative user study. Two different dialogue strategies were used within in user test in order to find out, which of the two designs works best in the given context. As it can be seen in the results section, the study helped to get an understanding of the technical challenges that can occur. Several new and unexpected events happened during the interaction of the participants with the prototype. By the conducting the interviews, many points for possible improvements have been mentioned by the user, that help to understand the main concerns and challenges when it comes to interacting with the a virtual assistant. The results of this user study help to define new user and the systems requirements that will be discussed in the next phase of the design process (chapter 4: specification) and lay the foundation for the design of the next prototype.

4. Specification

4.1. Introduction

In the specification phase the results from the previously conducted user study will be analyzed and used in order to define updated user and system requirements. These requirements will be the foundation for the realisation of a hi-fi prototype that is being developed during the next phase of the design process (chapter 5).

4.2. Updated user and system requirements

One of the design problems that should be solved is that the assistant sometimes tends to leave a conversation. This should not be an option in the given context. When talking to the assistant, only the input by the user matters. The user should finish answering all questions or explicitly indicate that he wants to end the conversation by using a prompt like for example “Stop asking me questions”. From the interview questions, it became clear that if once a user gets of track of the conversation, it was not possible to get back on track. This problem will be solved with the next prototype. A possible solution would be to make sure, the assistant will be more capable of understanding the user. This will be accomplished by giving the user more freedom in giving answers. The assistant will acknowledge and save any kind of answer without depending on specific keywords (words, the assistant is taking action on) that the user needs to trigger. After the user gave an answer, the assistant will simply go to the next question.

One user has been hesitating when giving an answer to the assistant, because it has not been clear, if it was still listening or not. Often, this is indicated by the assistant through colorful visuals on the screen, animated according to audio signals received from the smartphones microphone (see image below).

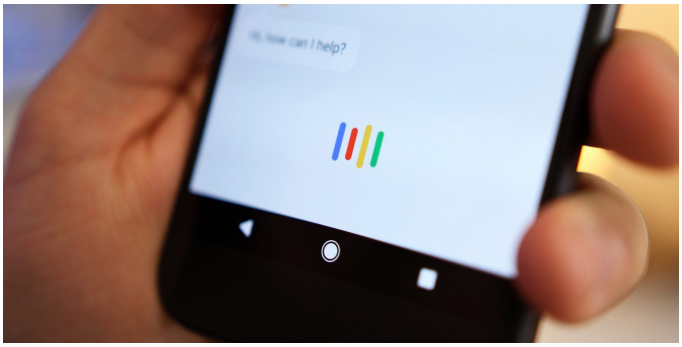

Figure 3. Google Voice assistant indicating that it is listening.[18]

Although, the indication seem to be clear to most of the users, some users might not be familiar with voice interactive technology. Therefore, an easy fix would be to simply increase the amount of time, the assistant is actively listening, even though the user is not replying immediately.

Another problem to point out is that if the goal of an interaction with an assistant is to gain precise answers from the user, simple ‘yes’ or ‘no’ questions should not be used. When asking the question “Is there anyone with you right now?” one of the users simply answered ‘yes’ without further stating, who is with him. The question will be rephrased for the next prototype. Another question that caused problems was the question ‘Where are you?’. Some users tend to answer this question very general by for example saying “I am in the Netherlands”. Even though this answer is not wrong, it is still way to broad for the given therapeutical context. For the next study the question could simply be rephrased by adding the adverb ‘exactly’. This might help to get a more precise answer to the question.

Often, users were confused, why the assistant did not understand them. Especially, when the words said by the user were detected correctly by the assistant (as indicated on the smartphone screen). Consequently, the users were not sure if they should simply repeat there answer, as requested by the assistant, or phrase them differently. The reason why the assistant in this prototype sometimes did not understand the answers by the user, is because it was necessary to define keywords for possible answers. For Example: If a user says “I am at the university” but the word “university” has not been defined in google dialogue platform as an

[18] Source: http://www.androidguys.com/news/google-assistant-is-coming-to-a-phone-near-you-soon/

intent, in order to be recognised by the assistant. To fix this issue, the plan is to design the assistant for the next prototype in a way that it simply accepts any answer by the user, without having to understand the meaning of each recognized word in the given context.

While 2 out of 5 users clearly stated that the difference between the two dialogues was not relevant or important to them and one did not even notice it at all, still two out of the five users mentioned that the second strategy gave ‘more context’ and was more ‘likeable’. The only question that provided more context was the first one “Hi. I’ll be asking you a few questions about your current situation: First, where are you right now?”. However, the assistant also gave more feedback on if he understood a question (by for example saying ‘Got it [...]’) it can be concluded that a given context seems to be very beneficial to begin the dialogue with. Although, the context in a real world scenario would be more clear beforehand, it can be suggested to provide the user more context information at least for the first few interactions with the assistant until he gets more familiar to the questions (the user will have to go through the same questions each day for several weeks). The plan for the next prototype is to provide more information at the beginning of the interaction and orientate more on the style that has been used in the second dialogue strategie.

4.3. Conclusion

In this phase several relevant points of the lo-fi prototype that can be improved have been mentioned and been used in order to define updated system and user requirements. The updated requirements will help to improve the interaction with the next prototype. For the next prototype the recognition of the users answers will be improved and the way questions are being phrased as well. Besides, more context information will be provided for the user.

5. Realisation

5.1. Introduction

After building the lo-fi prototype, conducting the first user study and specifying new user and system requirements, the next step will be to develop the hi-fi prototype based on these requirements. The hi-fi prototype will then be used for a second user study in order to find further usability problems in the interaction between the user and the virtual assistant that could be improved.

5.2. The Hi-Fi prototype

5.2.1. Prototype Description

The hi-fi prototype will be an improved and extended version of the lo-fi prototype. Same as for the lo-fi prototype, Google Dialogflow will be used as a tool to design a interaction between the user and the virtual assistant. The following things will be improved for this prototype:

- The dialogue strategy used for the prototype will be improved by using a combination of the two strategies used in the ideation phase. However, it will mainly be based on the second strategy, since that one has been preferred by most of the users.
- The VUI will be extended in order to test the complete interaction necessary for the method of changing habits as stated in the book by Charles Duhigg. For the previous prototype the assistant only asked questions that are necessary to identify the cue that triggers a specific habit. This time, the interaction will be designed in a way that the assistant can also help the user to find out, what the reward of a habit is and help to make a plan to implement a new routine.

- User Information that are provided while the user interacts with the system will be stored, so they can be accessed and analyzed at a later point. This is not only helpful for the user of the envisioned system in order to analyze the progress, but also for a user study that will be conducted in the following section.
- The prototype will be used within a usability test. The interaction will therefore take place on three different devices: A smartphone and two different smartwatches, one silent watch and one with speech output. More details about user study are provided in the section 6.1.

5.2.2. Voice User Interface

In a real world scenario the user would experiment with different rewards for a few days or weeks, until it is clear what the actual desire is that the user follows. Everytime the user noticed that he or she is about the follow a specific habit, a different reward should be tried out and be tracked by using the assistant on the smartwatch. After the actual desire of the user has been identified, the cue that triggers the habit will analyzed. Therefore, different questions about the current situation of the user will be asked. Once, the desire and the cue of the habit have been identified, a plan will be made, so the user will get reminded to implement a new routine. If the habit is depending on a specific time of the day for example, the user could always set a reminder, in order to get reminded by the smartwatch to follow the new routine. The following prompts will be used for the assistant in order to help the user going through the mentioned process:

Reward experimentation	• "Hey, how are you? I am helping you to identify the desire that is responsible for your habit. How did you change your routine this time?" • "Got it. How are you doing right now? Any particular thoughts or feelings?" • "All right, thank you!"
Isolating the cue	• "Hi. I'll be asking you a few questions about your current situation: First, where are you right now?" • "All right. And how do you feel?" • "Who is being with you right now?" • "Got it. Last question, can you tell me what you are doing?" • "Okay. That's it for now...talk to you soon!"

Making a Plan	• "At what time would you like me to remind you, following your new routine?" • "Reminder set."

Table 5: Prompts used for the hi-fi prototype.

5.2.3. Designing the Interaction Using Google Dialogflow

5.2.3.1. Intents

As mentioned earlier, the tool Google Dialogflow is being used in order to build the conversational experience with the assistant. Within the tool, different intents can be defined for the interaction. An intent can be seen the action the user wants the assistant to conduct, like for example checking weather information. As defined by google "An intent represents a mapping between what a user says and what action should be taken by your software."[19] Therefore, the three dialogs "reward experimentation", "cue isolation" and "making a plan" have been used as intents in the context of this project. Besides, a default fall back intent will be used, that will be triggered, if the assistant did not understand correctly, what the user said. For each intent different training phases have been entered into the system. These help dialogflow to understand, how the user might express their intent. For "cue isolation" these phrases are for example "start isolating the cue" or "help to isolate the cue for my habit".

[19] https://dialogflow.com/docs/intents

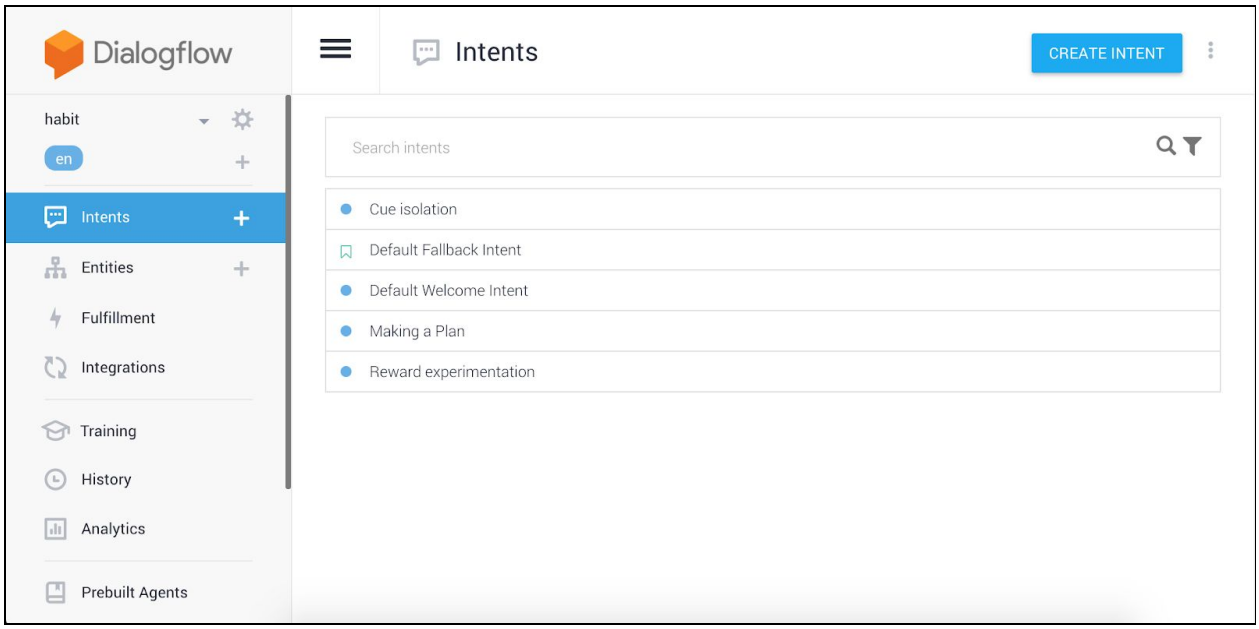

Figure 4:. User Interface of dialogflow for defining intents.

5.2.3.2. Entities

For each intent, different so called ‘entities’ have been defined. Entities are used as specific parameter values that can be identified from a natural language input. If for example the user triggers the intent "cue isolation", it can be defined that before ending the conversation, all parameters like the users location, mood or current activity have had to be mentioned during the interaction. If those parameters are not recognized by the assistant, it will ask for them in descending order (see example below).

Figure 5. Parameters for the cue isolation in dialogflow.

As soon as all parameters have been detected, the assistant answers to the user with a response that is defined as the end of a conversation. At the example of the cue isolation this will be the prompt "All right that's it for now. Goodbye." Entities offer a great alternative to the procedure used earlier in the lo-fi prototype, where each question has been defined as a new (follow-up) intent. The hi-fi prototype works more in a way of 'checking the boxes', which makes designing the dialogues way more simple and intuitive.

5.2.3.3. Integrations

While designing the interactions in Dialogflow they can simultaneously be tested within the web browser. Alternatively, an integration function can be used in order to make the application available on all different devices that support the google assistant and are used with the same google account that is being used for dialogflow project. These devices can be smartphones, wearables, earphones or even in cars.

5.2.4. Setup

For the user study with the hi-fi prototype in the next chapter, the VUI will not only be used on a smartphone, but for this time additionally on two different smartwatches. For the study a within subject design has been chosen, which means that all participants are going to test the interaction on all three devices. The devices that the application for the google assistant will be running on are the smartwatch Moto 360 (2nd generation)[20], the Huawei Watch 2 and a smartphone, the iPhone 6s[21]. In the table below it is being stated what software versions have been used:

Device	Technical Specification	User Interface	Screen size
Apple iPhone 6s	Model: MKQJ2ZD/A OS: iOS 11.4 App: Google Assistant, 1.3.2110	Input: speech Output: text and speech	4.7 inches (11.94 cm)[22]

[20] More information: https://www.motorola.com.au/products/moto-360
[21] More information: https://www.apple.com/iphone-6s/specs/
[22] Source: https://www.apple.com/iphone-6s/specs/

Moto 360 (2nd generation)	Model: Moto 360 OS: Android 7.1.1 Wear OS: 2.12.0.197555195 App: Google 8.8.12.25	Input: speech Output: text	1.37 inches (35mm)[23]
Huawei Watch 2	Model: LEO - BX9 OS: Android 7.1.1 Wear OS: 2.12.0.19755195 App: Google 8.8.12.25	Input: speech Output: text and speech	1.2 inches (30.4 mm)[24]

Table 6: Devices that the interaction has been tested with.

While in practices it could take a few weeks in order to change the users routine, in the experimental setup the participant will simply answer the questions on the prototype one after another, since the focus of the study lies on the interaction and the user experience. Speech output enables a more human-like interaction. It is closer to the natural way of how humans communicate. However, this does not necessarily mean that the user will prefer this way of interaction. Users might for example feel uncomfortable having a device talking to them. This is why two different watches will be used and the interaction compared. Besides, a smartphone will be used for the interaction. The phone will help to make a comparison of using the virtual voice assistants on different devices (phone and wearable).

5.3. Conclusion

The concept of designing voice interaction is very different to designing for example regular interfaces. For many designers this can be a very new and different way of designing interactions that might take some time to get familiar with. However, after getting a understanding of the key concepts, tools like dialogflow can be very helpful to designers and developers in order to test and develop conversational experiences. Especially, the possibility to make use of google's machine learning plattform is very valuable, since it helps the assistant to understand answers by the user even though they are not completely identical with the phrases within the application. Still, it has to be kept in mind that the system is not always as smart as

[23] Source: https://www.smartwatchspex.com/comparison/motorola-moto-360-2nd-gen-42mm-vs-46mm-spces/
[24] Source: https://www.gsmarena.com/huawei_watch_2-8585.php

expected by the user and that the way, users express their intents can highly differ. Therefore, it it can be suggested to frequently test the dialogs with potential users while designing the them and to train the system with phrases and synonyms. In the next section another user study will be conducted using the hi-fi prototype.

6. Evaluation

6.1. User study

6.1.2. Goal

The purpose of the hi-fi prototype is to use it within a user study and to find what device the user prefers for the interaction. The interaction will be tested on two different smartwatches and a smartphone. One smartwatch will have a speech output, which means that it is actually talking to the user, the other watch will not speak and simply display text messages. The smartphone will as well have a speech output. It will help to find out what is specifically important when designing voice interactions for a smartwatch and how it can be improved. Another thing that would be interesting to find out, would again be the errors that might occur during the interaction. Also, do people speak differently, using different devices? Is it more convenient to talk to a smartphone (handheld device) instead of an smartwatch (wearable) because the users are more familiar with phone? Once the the user interacted with the assistant on the different devices, a few general questions will be asked in order to gain more information by the user on how the overall interaction went. Based on the findings of this study using the hi-fi prototype, the results will be discussed in the next chapter.

6.1.2. Procedure

The testing procedure will be qualitative, because most of the questions that are tried to be answered are about the user experience that can be observed by the conductor during the interaction an by interview questions. There are no quantitative measurements like time compilation time or number of errors. Besides, the qualitative procedure will offer the possibility to gain more insights by the user trough asking more detailed question and being more flexible. According to the Nielsen Norman Group, 7 - 8 participants are often enough in order to find most of the usability problems[25]. For this study a total of 10 participants will be used. The environment, in which the study will be conducted will be simple, so the user does not get

[25] Source: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

distracted by surroundings like for example other people, but focuses on the interaction. The user will be seated on a chair with a table in front of him and interact with the agent on a smartphone and on a smartwatch. The devices will be placed on the table in front of the user, ready to get interacted with. The conductor of the user study will sit at another location in the room, in order to observe but not distract the user. The order of the interaction with the different devices will be balanced between the user. A camera will be used to gain more insights about the user's behaviour during the interaction and to document answers to interview questions.

1. Before the study begins, the user will be asked to sit down on a chair and after that read and sign a general consent form before participating. After that, all necessary preparations for the interaction will be conducted. Further, the user will be provided with following information:

   "Imagine the following scenario: You are suffering from a habit that you would like to change. The process to change it can take up to several weeks. A virtual assistant will help you to change your habit. Therefore, three different conversations will be held on regular basis. During this user study you are now going through all the three conversations.

   Conversation 1:
   Part of a method that will be used in order to help you, is by changing your routine a little in order to identify your actual desire. For example: After having the desire to smoke a cigarette you ate a gum instead. Please imagine for a moment that you just had the desire to follow your habit, but chose a different reward instead.

   Conversation 2:
   In the second dialog, the assistant will ask you a few questions about your current situation. These will help to isolate the cue that triggers a habit. Please simply answer the questions.

   Conversation 3:

Please assume the following: After getting a better understanding of your habit, you now made a plan to change it. Please answer the questions by the assistant."

2. As soon as the user wears the watch on his wrist, the interact with the prototype will begin. The prompts that have been defined will be used to enable the dialogue between smartwatch and user. Before each of the three dialogues (as mentioned in the section above) the user will be provided with additional information in order to understand the context of the questions that are being asked. Everytime the assistant recognizes a given answer by the user, it will respond with follow-up questions until all questions has been successfully answered, or the user got stock at a specific step in the conversation.

3. After each of the three interactions with the different devices, the user will be asked questions about the the interaction. The questions are being asked in order to get more valuable feedback from the user on how he experienced the interaction. The following questions are being asked:

    - How would you say the overall interaction with the device did go?
    - What did you like about the interaction?
    - How did you perceive the assistant and why?
    - What errors occurred and what could be improved?
    - Was it complicated to get back on track, if the interaction failed?
    - How did you feel, talking to the device?

    After all the interactions have been completed on all three different devices, the user will be asked comparison questions and some on his or her demographics:

    - How familiar with voice interaction technology in general are you?
    - Which device did you prefer for the interaction and why?
    - Did you prefer the text or speech output by the assistant and why?
    - Were there some significant differences between talking to a phone and to a smartwatch?

- Could you imagine to use a voice controlled virtual assistant on a smartwatch in your everyday life?
- Do you see benefits of using voice interaction particular on a smartwatch?
- In what situations could you imagine to use it and where wouldn't you use it?
- How old are you?
- What do you study?
- Where are you from?
- What's your gender?

6.1.3. Participants

For the user study 10 participants will be randomly selected at a university campus and asked to participate in the study. There were no specific criteria when it came to selecting participants, it only has been tried to keep balance variables like gender, nationality and educational background. Like for the previous user study, it can be assumed that all the participants have a general understanding of using technology. The age range of the participants is 20 to 30 years and all of them are students in Netherlands. Their nationalities and gender can differ. All interactions will take place in english language.

Participant	Age	Gender	Study	Nationality	Familiarity with voice interaction technology
1	23	male	Creative Technology	German	Known from the Smartphone, but not used regularly
2	23	female	Social Work	German	Not familiar
3	23	male	Creative Technology	Indian	Not familiar
4	21	female	Creative Technology	Bulgarin	Not familiar
5	21	female	Communication Science	German	Known from the Smartphone, but not used regularly
6	22	male	Creative Technology	Dutch	Relatively familiar
7	25	male	Psychology	German	Not familiar
8	21	male	Psychology	Dutch	Not familiar
9	22	female	Creative Technology	German	Not familiar

10	26	male	Social work	German	Known from the Smartphone, but not used regularly

Table 7: Demographics of participants that participated in the user study with the hi-fi prototype.

6.2. Results

From the conducted user studies some interesting discoveries have been made. The answers to the interview questions have been collected in a spreadsheet. All the relevant results are stated in this chapter and will hopefully help to eliminate future usability problems in the field of voice interaction on wearables.

6.2.1. Interaction with the smartwatch without speech output

When asking the participants, how they would say the overall interaction with the device did go, many users (6 out of 10) stated that it worked not so well. From observing the interaction it became clear that the speech recognition did not work as good as on the other two devices. Besides, one user mentioned the delay that occured when the spoken word by the user have been processed. When asking the users, how they felt, talking to the device, they used terms like ‘awkward’, ‘stupid’, ‘lonely’, ‘weird’’. There were positive no positive responses to this question. However, there was one positive feedback this type of interaction got: "It is good that it does not always speak". Another, issue that occured by two users, was that they were not sure, if the device was listening or not. They have not been sure in the beginning how to indicate that they were going to speak again. One user mentioned that it would be better "(...) if you can have a real conversation, so (the assistant) speaks to you".

6.2.2. Interaction with the speaking smartwatch

For the interaction with a smartwatch that was speaking to the user, instead of just giving text feedback, most participants reacted positively to the question on how the overall interaction went. The interaction has been described as ‘smooth’, ‘very well’, ‘pretty well’ and ‘good’. 9 out of 10 people gave a positive response. One user mentioned that the latency could be improved. Same as on the other two devices the virtual assistant on this watch has been perceived in a

positive and friendly way. Speech recognition errors still occurred sometimes. When asking participants, how they felt, talking to the device they some descipred their feelings as ‘strange’, ‘weird’ or ‘awkward’, others said that it was okay. Recovering from errors has not been a problem. For one user there were some minor errors in the recognition. He said that it was not difficult. If the conversation failed "I could simply try again". One user further noticed that the assistant is not listening while it is talking. One user mentioned, that timing was a problem. At some points the watch stopped waiting for an the users answer.

6.2.3. Interaction with the smartphone

When asking participants, how the overall interaction with the device did go, they responded positively, although two participants mentioned the performance of the speech recognition still as a negative point. Three participants mentioned that they like the conversational, messenger-like layout which helps to read the past interactions (prompts that have been used by the user and by the assistant). Same as for the other two devices, the reaction the the question about how the assistant has been perceived was positive (10 out of 10 users). One user however mentioned that the assistant sounded ‘robotic’ and another one said that it sounded ‘choppy’. One user said that it was difficult to get back on track, if the conversation failed. Another participant mentioned that the assistant moved on to the next question, even though the the recognition failed (the recognized sentence did not make sense). When asking about how the user felt during the interaction, only one out of ten users stated that he felt awkward.

6.2.4. Device comparison and general remarks by the participants

Only in some specific situations users would prefer to not have a speech output. Mainly in public places, like for example the bus, or locations where it could distract people, i.e. in a library. Most (9 out of 10) users preferred the speech output on the devices. According to the users, this would make the interaction more ‘personal’, as stated by one of the users and also easier to understand what the assistant wants. 6 out of the 10 users that were being interviewed, said that they would prefer the smartphone over the smartwatch when it comes to voice interaction. The assistant on the phone would worked better, was perceived faster, being

more efficient and gave the possibility to read along with the conversation. Two users stated that their familiarity with a phone might have an influence on their choice. However, there were also some clear statements being made on why the interaction with the with smartwatch would be beneficial: The smartwatch is light weight, easier to access and therefore the interaction sometimes more efficient (i.e. tasks like setting reminders). One user even stated that the smartwatch feels like a part of the body, while a smartphone is perceived more as an external device. Some specific situations where the watch can be very beneficial is during sports, in emergency, for setting reminders, making grocery lists, setting alarms, when the user is in the rush or if the user has a disability that makes accessing a smartwatch easier than accessing a phone.

Some participants (4 out of 10) mentioned that they would not use the device in public spaces, because they might feel awkward talking to a watch when other people are around or they would be concerned about the functionality due to background noise. However, 8 out of 10 users stated that they could imagine to use a virtual assistant on a smartwatch in their everyday life, under the conditions, that it works well and getting used to the way of interaction.

Even though the error rate of the speech recognition is relatively low and improves over time, it is still not advanced enough to be as good as humans. Sometimes it happens that assistant processes an answer by the user, even though one or more words have not been recognized correctly. Since, the user is limited in understanding the context of the conversation, it does not realize that some words in the sentence simply make no sense at all.

6.3. Discussion

The results from the observative user study, provided some interesting insights on what users like about voice interaction in general, what could be improved, what technical errors and also what design errors can occur. In the following the main points will be discussed.

Surprisingly, not many users liked the interaction with the silent smartwatch. Even though it seems beneficial to have a watch that is not always speaking, especially in public places, many participants had a negative response concerning the quality of the interaction.

Since users mentioned that they felt "awkward" or "lonely" during the interaction it can be assumed that the interaction felt not as natural as the interaction with an actual human-like voice (as on the other two devices). Besides, it can be assumed that it also takes more effort to understand what what the smartwatch wants. Instead of just listening to the spoken answers, the user had to pay attention to the tiny text strings appearing on the smartwatch screen, reading them, and sometimes even scrolling through the text. At that time it was also important that the user was not reading the prompts by the watch out lout, since this could irritate the watch.

Interacting with the the speaking smartwatch seems to be much more comfortable to the user. As mentioned earlier, 9 out of 10 gave a positive response to the question about how the overall interaction with the device did go. This strengthens the argument that users like the smart watch that is actually speaking to them and gives a clear answer to the research question mentioned previously, where the question was, what device and kind of interaction the user would prefer.

By letting the users also interact with the assistant on a smartphone, the advantages of interaction with an assistant on a phone became clear very quickly. It is the user interface that simply offers way more possibilities in order to show further information about the interaction. For the users it is easier to read a long to the conversation. They can see how the assistant is recognizing the words by the user and if it is recognizing the words correctly. Besides, the answers by the assistant can be read and might be recognized due to a bigger font-size even more easily. Another advantage is the conversational like design. Users can scroll through the past conversation and see what the assistant has asked and what the user answered to the questions.

When comparing the interaction on the three different devices, it becomes clear what the requirements by the user are. They want to be able to have conversation with the assistant using their voice. Furthermore, they want to feel good and not awkward during the interaction, which is why is one reason, they would not use it in public places. Other people could feel distracted and might wonder what the exactly the user is doing, since they are not familiar with this new concept of human machine interaction. One problem that has been mentioned by some

users about the interaction with the assistant on different devices is the latency. Sometimes it takes the assistant an uncomfortable amount of time to process the responses that the user has spoken to the device. These answers are being sent to the cloud, processed, send back the device and ultimately spoken or displayed to the user. Depending on the answers the user is giving, the length of the answers, the processing power of the watch and also the connection to the internet are responsible for the smoothness of the interaction. This problem is not new and has also be stated in similar previous research that has been conducted (like for example the application of voice interaction technology in cars).

What is also interesting, is that 8 out of the 10 interviews said that they could imagine to use an assistant on a smartwatch on a daily basis, given the condition that it works well. This is a result that is very motivating to continue conducting further research in the fields of voice interaction technology and wearables.

7. Conclusion

7.1. Summary

In the previous chapters it has been shown, how from a simple technological topic: ‘Adding voice interaction to a wearable’, background research in different domains and an ideation phase, an interesting research question has been developed. On the basis of this research question, an iterative process has been used, in order to design an interaction between a virtual assistant on a smartwatch in the context of habitual change. There have been two prototypes and two user studies being conducted in order to improve the design and to define new user and system requirements. In the context of the user studies, participants interacted with the prototypes and afterwards answered several interview questions. Answers from the participants have been recorded, analyzed and been used for the design process and for discussion.

7.2. Discussion

There are some relevant findings that have been made within this project: It has been shown that users prefer to have a human-like, conversational interaction with a virtual assistant. Many users could imagine to use an assistant on a smartwatch in there everyday life, given that advances in the fields of speech recognition, language processing and voice user interface designs are being made. Designing VUI’s turned out be a very new field and therefore, more research and design guidelines on this very specific way of designing would be pleasant not only for designers but also for developers of voice applications. From the user observations having a virtual assistant on a smartwatch seems to be a promising tool for efficient interactions. It is faster to take a smartphone out of the topic, navigating through apps and typing commands. When advances will continue to be made in the these fields of technology, there seem to be no barriers that could prevent us from interacting with machines in this today often very futuristic perceived way of interaction. Even though the conducted experiments have been taken place in the context of habitual change, findings of this study can also be applied to other contexts where voice interaction technology is being applied on wearables devices.

In order to also look at the limitations of this research, the following three points should taken into account: Firstly, the time frame in which the users interacted with the assistant was very short (around 10-15 minutes). This is of course not much time in order to get to know the assistant and to experience the interaction in detail. It might be the case that when interacting with the assistant in the given context on a daily basis the interaction for each of the devices might work even better, or worth (depending on the circumstances). Some of the questions are being asked very frequently and the user would get used to them. Consequently, it would be possible for the user to give more clear and quick answers. During the interactions it could have been observed, that sometimes the speech recognition was still problematic. For example: If the user thought about what he or she wanted to say, the watch stopped listening after a while, same if the user made a longer break within a sentence. The second point is that during the second user study, each user had to interact with all three devices one after another. The users sometimes might have been better prepared to answer the questions when interacting a second or third time. However, this probably did not strongly influence the possibility of identifying usability errors. The third and final point is the performance of the two smartwatches. Even though the same operating system is running on the two watches and the same voice assistant is being used, it might be that the two watches perform differently. The positions of the microphone, quality of the microphone or the connectivity to the wifi network might differ and influence the results.

7.3. Future Work

It can be recommended that further research should be conducted in the field of voice user interface design, in order to enable fruitful discussions on voice user interface design and virtual assistants on smart devices. It would be interesting to investigate, what exactly causes the latency when interacting with the device and what options are available in order to improve it. It might be that for example not only the internet connection, but also the battery status, processing power of the device, or length of answers the watch has to process affect the latency.

Another promising way to test and improve the system would be to conduct a user study for a longer time frame and to see how users would interact and experience the assistant after several weeks or month. Users might get tired over time, being asked the same questions in the same way over and over again. It could be investigated how the users would react if the assistant asks questions in different ways. Possibly after interacting for several weeks, they users might prefer to get asked questions like in the first dialog strategy that has been used in the first of the two user tests, since the interaction is faster and the user does not get interrupted from his current activity.

An interesting next step would also be to investigate, how the data that is being collected by the watch can be stored in a database, how relevant keywords be identified and the information later on be visualized for the user, so that the user can easily detect correlations draw conclusions concerning his or her habit.

Besides, it would be interesting to find out, how the interaction would work, using different tools in order to design the interactions. One option for example would be to develop an a complete application for the watch instead of relying on the given google assistant. This might offer a few more possibilities to design the user interface. It might be easier to adjust specific parameters, like for example the time span for the assistant until it stops listening, if the user does not talk. Furthermore, this would offer more ways to design the graphical user interface, for example by making use of visually appealing animations to motivate or reward the user. Besides, it could be made use of other sensors and tools available in the watch that could help to measure information valuable for the context (time, location, noise, heart frequency, etc.). Depending on this information the watch could try to identify for example, if the user is in environment where he does not want to talk and prefers to enter information on the smartphone.

Appendix A: Information Brochure - Lo-Fi Prototype User Study

This study is about interacting with a virtual assistant that helps the user to change their behavior by adopting good habits or loosing bad ones. The goal is to find out how to make the voice interaction with the agent that will be used on a smartwatch as efficient and intuitive as possible. Your participation will help to achieve this goal.

In this session you will be working with a simple prototype in order to test different dialogue strategies. You will be asked to perform tasks that typical users of the assistant might have to conduct, such as answering questions about your current location or mood. After the user test you will be asked some general questions by the researcher about how the interaction went.

The investigator will sit in the same room, quietly observing the experiment. A camera will record the session in order to gain more insights on the interaction. The video material will only be used in the context of the project and deleted permanently within 6 months after the conduction of the user test.

Anonymity of your participation in the research will be guaranteed and data will not be disclosed to third parties without your permission. There will not be any risks or discomfort to experience. It is only being investigated what aspects of the user interface are confusing or do not work well, so they can be improved. The participation remains voluntary and without giving any reason, you may refuse to participate in the research at any time and also afterwards (within 48 hours) decide if the data may be used in the research or get deleted.

After the test has been completed you are welcome to ask more questions concerning this study and the field of research.

This research project has been reviewed and approved by the EEMCS Ethics Committee of the University of Twente.

Appendix B: Information Brochure - Hi-Fi Prototype User Study

This study is about interacting with a virtual assistant that helps the user to change their behavior by adopting good habits or loosing bad ones. The goal is to find out how to make the voice interaction with the agent that will be used on a smartwatch as efficient and intuitive as possible. Your participation will help to achieve this goal.

In this session you will be working with a simple prototype in order to test a dialogue strategie. You will be asked to perform tasks that typical users of the assistant might have to conduct, such as answering questions about your current location or mood. After the user test you will be asked some general questions by the researcher about how the interaction went.

The investigator will sit in the same room, quietly observing the experiment. A camera will record the session in order to gain more insights on the interaction. The video material will only be used in the context of the project and deleted permanently within 6 months after the conduction of the user test.

Anonymity of your participation in the research will be guaranteed and data will not be disclosed to third parties without your permission. There will not be any risks or discomfort to experience. It is only being investigated what aspects of the user interface are confusing or do not work well, so they can be improved. The participation remains voluntary and without giving any reason, you may refuse to participate in the research at any time and also afterwards (within 48 hours) decide if the data may be used in the research or get deleted.

After the test has been completed you are welcome to ask more questions concerning this study and the field of research.

This research project has been reviewed and approved by the EEMCS Ethics Committee of the University of Twente.

Appendix C: User Study Consent Form

Researcher
Matthias Codes
Creative Technology - University of Twente
Haaksbergerstraat 311, 7545 GJ Enschede
+49 15787371779

Ethics Committee University of Twente
J.M. Strootman - Baas (Anja)
Supporting staff
+31534896719
5, Drienerlolaan, 7522 NB Enschede

Statement of Informed Consent

With this consent form I confirm that I have read the information brochure[26] about the user study and fully understand the information and tasks that I will have to conduct during the study. I voluntarily agree to participate in the study and accept that it will be recorded and the retrieved data from the interaction and the interview be used for research purposes.

Please tick the appropriate boxes	**Yes**	**No**
I have read and understood the study information dated [__/__/___], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves audio or video recordings, which partly might be transcribed as text. All video recordings will be destroyed within 6 weeks.	□	□
I understand that information I provide will be used in a report.	□	□
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	□	□

Agreement

I have read the information provided above thoroughly and understand what data will be collected, what it will be used for and how it will be treated if I choose to end my participation. I provide my consent for the participation in this experiment and interview.

[26] Handed to the participant together with the informed consent form before participating in the study.

_____	_____	_____
Name	Date	Signature

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Name	Date	Signature

Study contact details for further information:
Matthias Cordes*, m.cordes@student.utwente.nl*

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente by ethics-comm-ewi@utwente.nl or the supervisors of the project Khiet Truong and Randy Klaassen by k.p.truong@utwente.nl or r.klaassen@utwente.nl.

Appendix D: Hi-Fi Prototype User Study - Context Information

Context information for conversations with the virtual assistant

Imagine the following scenario: You are suffering from a habit that you would like to change. The process to change it can take up to several weeks. A virtual assistant will help you to change your habit. Therefore, three different conversations will be held on regular basis. During this user study you are now going through all the three conversations.

Conversation 1:
Part of a method that will be used in order to help you, is by changing your routine a little in order to identify your actual desire. For example: After having the desire to smoke a cigarette you ate a gum instead. Please imagine for a moment that you just had the desire to follow your habit, but chose a different reward instead.

Conversation 2:
In the second dialog, the assistant will ask you a few questions about your current situation. These will help to isolate the cue that triggers a habit. Please simply answer the questions.

Conversation 3:
Please assume the following: After getting a better understanding of your habit, you now made a plan to change it. Please answer the questions by the assistant.

References

Aseffi, M. (2015). An Experimental Evaluation of Apple Siri and Google Speech Recognition. Retrieved April 02, 2018, from https://www.semanticscholar.org/paper/An-Experimental-Evaluation-of-Apple-Siri-and-Google-Assefi-Liu/42e0a535049b2183f167235fc79f0deace4f11c3

Azrin, N., & Nunn, R. (1973). Habit-reversal: A method of eliminating nervous habits and tics. Behaviour Research and Therapy, 11(4), 619-628. doi:10.1016/0005-7967(73)90119-8

Chang, C., Boyle, L. N., Lee, J. D., & Jenness, J. (2017). Using tactile detection response tasks to assess in-vehicle voice control interactions. Transportation Research Part F: Traffic Psychology and Behaviour, 51, 38-46. doi:10.1016/j.trf.2017.06.008

Chen, F., Jonsson, I., Villing, J., & Larsson, S. (2010). Application of Speech Technology in Vehicles. Speech Technology,195-219. doi:10.1007/978-0-387-73819-2_11

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it OK if I eat you?" Proceedings of the 2017 Conference on Interaction Design and Children - IDC 17. doi:10.1145/3078072.3084330

Duhigg, C. (2014). The Power of habit: Why we do what we do in life and business. New York: Random House Trade Paperbacks.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., Ruiter, L. D., & Hegel, F. (2012). If you sound like me, you must be more human. Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI 12. doi:10.1145/2157689.2157717

Hura, S. (2008). HCI beyond the GUI: Design for haptic, speech, olfactory and other nontraditional interfaces. Amsterdam: Elsevier/Morgan Kaufmann.

Kubo, Y. (2017). Exploring Context-Aware User Interfaces for Smartphone-Smartwatch Cross-Device Interaction. Retrieved April 02, 2018, from https://dl.acm.org/citation.cfm?id=3130934

Landau, M. J. (2008). Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk. Vanderbilt Undergraduate Research Journal, 4(0). doi:10.15695/vurj.v4i0.2783

Lee, C. M., & Narayanan, S. (2005). Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing, 13(2), 293-303. doi:10.1109/tsa.2004.838534

López, G., Quesada, L., & Guerrero, L. A. (2017). Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. Advances in Intelligent Systems and Computing Advances in Human Factors and Systems Interaction, 241-250. doi:10.1007/978-3-319-60366-7_23

Markowitz, J. A. (2017). Speech and Language for Acceptance of Social Robots: An Overview. Voice Interaction Design. doi:10.18411/d-2016-154

Matic, M., Stefanovic, I., Radosavac, U., & Vidakovic, M. (2017). Challenges of integrating smart home automation with cloud based voice recognition systems. 2017 IEEE 7th International Conference on Consumer Electronics - Berlin (ICCE-Berlin). doi:10.1109/icce-berlin.2017.8210640

Milanesi, C. (2016, June 03). Voice Assistant Anyone? Yes please, but not in public! Retrieved March 02, 2018, from http://creativestrategies.com/voice-assistant-anyone-yes-please-but-not-in-public/

Nasirian, F. (2017). AI-Based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives. Retrieved April 02, 2018, from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1032&context=amcis2017

Pantelopoulos, A., & Bourbakis, N. (2010). A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(1), 1-12. doi:10.1109/tsmcc.2009.2032660

Perez, S. (2017, November 08). Voice-enabled smart speakers to reach 55% of U.S. households by 2022, says report. Retrieved March 02, 2018, from https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report/

Reehal, S. (2014). Siri –The Intelligent Personal Assistant. International Journal of Advanced Research in Computer Engineering & Technology. Retrieved April 02, 2018, from http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-6-2021-2024.pdf

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The utility of affect expression in natural language interactions in joint human-robot tasks. Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction - HRI 06. doi:10.1145/1121241.1121281

Schuller, B., Rigoll, G., & Lang, M. (n.d.). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. doi:10.1109/icassp.2004.1326051

Vinciarelli, A., Esposito, A., André, E., Bonin, F., Chetouani, M., Cohn, J. F., . . . Salah, A. A. (2015). Open Challenges in Modelling, Analysis and Synthesis of Human Behaviour in Human–Human and Human–Machine Interactions. Cognitive Computation, 7(4), 397-413. doi:10.1007/s12559-015-9326-z