Today's Voice-Based User Interfaces: Analysis of Six Paradoxes

University of Oulu
Faculty of Information Technology and Electrical Engineering / OASIS
Master's Thesis
Mikko Rouru
29.3.2019

Abstract

This master's thesis was about voice-based user interfaces, including the history of voice as an interactive medium, acceptance of the medium among users, and modern implementations of the interfaces. Initial problems of early implementations of voice-based user interfaces were about human-machine communication, as natural language of humans is ambiguous and highly interpretable, causing difficulties to machines when interpreting the given queries. Other topics of interest with the interfaces included implementing human-like cues to machines in order to make them more appealing to users, and how humans see inanimate objects when communicating with them. The problems of machines' understanding of humans have been reduced greatly, and modern voice-based conversational virtual assistants emerged in the early 2010s. These assistants were examined: what they are, what they have achieved and what does the future hold for them. To evaluate voice as an interactive medium and the assistants, conceptual framework of six distinct paradoxes was used. Prior research, from history to today's issues, was used to reflect against the conducted research.

Keywords
Voice as an interactive medium, voice-based virtual assistants, voice-based user interfaces, six paradoxes, qualitative research

Supervisor
Professor and Dean Harri Oinas-Kukkonen

Foreword

I would like to thank Professor and Dean Harri Oinas-Kukkonen, for presenting the opportunity and guiding me through the master's thesis – the idea already sparked back in the course "Emerging Technologies and Issues" in late 2017, and the work on thesis started a year later. The topic has been interesting, as new technologies which are on the brink of a breakthrough have always interested me. As an overall experience, conducting own research for the first time has also been a good experience, as the process was rather unknown before starting the master's thesis. Learning something new is always thrilling for me! Also, I would like to thank all my friends and family for the support – a lot has happened during the writing process, but with the constant help and support, the work is now done and the next challenges wait. Thank you!

Contents

1. Introduction

The purpose of this master's thesis is to examine modern voice-based user interfaces in several formats and device types, such as conversational virtual assistants in smartphones and smart speakers. The history of voice as an interaction medium in human-computer interaction (HCI) goes far beyond the modern examples of voice-based virtual assistants: for example, the topic has been examined in the late 1980s when voice's potential and restrictions with human-machine communication were recognized (Delogu, Paoloni & Pocci, 1991).

Although voice's features and characteristics in human-computer interaction have already been researched for decades, the consumer adoption rates of voice have been on the rise as late as in the 2010s. This is due to the modern era of voice-based virtual assistants, started by technology giant Apple's Siri in 2011 in their smartphone iPhone 4S. Following Apple's footsteps, Google, the search engine giant, introduced their own virtual assistant in the Android mobile operating system, and also other competitors such as Amazon and Samsung did introduce their alternatives. Initially, the voice's potential was used and seen in smartphones, but as Amazon introduced their Echo smart speaker system in 2014, three years after Siri's emergence, voice's use cases were expanded rapidly beyond typical ones in smartphones. (Kinsella & Mutchler, 2018)

Motivation for this research emerges from the aforementioned quick consumer adoption of these particular interfaces. Although the history of this interaction medium is relatively long, the interesting part is to examine what are the success factors contributing to the relatively late success. For example, around 62 percent of smartphone owners in the US use voice-based virtual assistants monthly. Another example of rising adoption rates is very recent: in January 2018, the smart speaker audience reach was 47,3 million in the US, but later in September of the same year, the number had risen to 57,8 million. (Kinsella & Mutchler, 2018)

Before diving into the modern era of voice as an interactive medium in the 2010s, the prior research of the topic has especially tackled the issues and differences between human and machine voices. Humans are ambiguous: they do not use explicit language and the tone, pitch and other variables change a lot when they speak. Humans adapt to different situations, for example voice's attributes change a lot when speaking to different persons. Machines on the other hand are quite the opposite, especially when considering the explicit factor. (Brennan, 1991) Initially, the voice's use cases with computers were mainly speech recognition and voice activation systems. Vocabulary and system training issues were also present in the early implementations: for example, acoustically similar but meaning-wise completely different words and software's need to be trained before using it caused issues with voice-based user interfaces. (Delogu et al., 1991; Rogoff, 2001)

After the advancements of the technologies, many aforementioned issues were resolved or at least reduced, causing the use cases of voice to be expanded from one-sided to conversational interaction. The systems are nowadays not only taking commands from the user, but they also respond and maintain conversations with them. Artificial intelligence (AI) has a big role in this – AI has been the main accelerator of the voice-based virtual assistants. (Guzman, 2019) Conversationality has brought anthropomorphism to the spotlight in voice-based user interfaces – this meaning

implementing human-like features into non-human objects. Everyday life is getting more and more into interacting with non-human objects with human-like features, and these design cues most likely will be essential ones in the future when designing easy-to-understand systems. (Epley, Waytz & Cacioppo, 2007)

Social cues of machines are profoundly researched in the field – social behaviour is essential in human-human interaction, which sets the question whether these cues are preferable in machines, too. Gender stereotypes, personalities, and tweaks of machine's self-referential cues either in 3rd or 1st person are examined with machines' context. (Nass, Steuer & Tauber, 1994) Social cues of human-human interaction and anthropomorphic characteristics implemented in the modern voice-based user interfaces are at the core of the modern conversational voice of the machine, and these features can affect the acceptance and efficiency of technology (Guzman, 2019).

The main research question of this thesis is surrounding these new, modern and conversational voice-based user interfaces: what are today's voice-based user interfaces in modern web and mobile environments, reflected against to the case of six paradoxes? Concept of six paradoxes is a conceptual framework by Oinas-Kukkonen and Oinas-Kukkonen (2013), initially reflected against social web issues. These six paradoxes are privacy, identity, credibility, friend, filter and value paradox. These are further disclosed and discussed in the prior research section, and these also work as the base of the qualitative research itself.

Another deeper research question is concerning the most common use cases with modern voice-based user interfaces. In prior research, users have stated that the voice's potential is in activities which require either hands, eyes or both to be occupied. Out of the top five of use cases where voice is preferred interactive medium, four are activities where the users are otherwise occupied on the activity itself, restricting the physical interaction by other interactive mediums, such as by using a touch screen. For example, driving and household chores are among the top activities when voice-based user interfaces are used. (Kinsella & Mutchler, 2018) Overall, ubiquity of voice has high potential in these particular tasks, and also in accessibility issues (Jeong & Shin, 2015).

The ubiquity leads to the third research question: what are the success factors of voice-based user interfaces and how these interfaces could be improved in order to achieve wider consumer adoption? Prior research indicates that the voice's ubiquity, meaning medium's presence everywhere without major physical restrictions might be the reason for adoption rate rising, as tasks can be done hands- or eyes-free (Jeong & Shin, 2015). Accessibility issues are also one highly potential area with these interfaces: disabled people and elderly can have easier everyday life, due to ubiquity of these interfaces with no major restrictions in physical presence (Portet, Vacher, Golanski, Roux & Meillon, 2013).

To recap and summarize, the three research questions are:

RQ1. What are today's voice-based user interfaces in modern web and mobile environments, reflected against the six paradoxes?

RQ2. What are the most common use cases with modern voice-based user interfaces?

RQ3. What are the success factors of voice-based user interfaces and how these interfaces could be improved in order to achieve wider consumer adoption?

To answer these research questions, first, a look to the prior research is taken, including the history of voice as an interactive medium, acceptance of voice as an interactive medium, modern voice-based user interfaces, and the framework of six paradoxes. These issues and topics form the base for the research. Second, qualitative research as a research method is generally introduced, along with document usage as a data gathering method. The documents will be based on US-based electronics retailer Best Buy's product reviews about voice-based smart speakers. The gathered data from the product reviews will be reflected against the framework of six paradoxes and also other supportive questions formed from prior research.

Goal is to answer all the research questions and paradoxes' issues extensively, by combining the prior research and conducted research together – for example, in the research, friend paradox can be applied to source orientation issues by Guzman (2019), by examining the product reviews to see how the users see modern voice-based virtual assistants. Thus, the main contribution is to find out whether the conducted research agrees with the prior research, and if there are any new issues found out in the research, such as potential hindrances or success factors of voice-based user interfaces.

The structure of the thesis is following: first, prior research is introduced including the history and acceptance of voice as an interactive medium, modern voice-based user interfaces and the conceptual framework of six paradoxes by Oinas-Kukkonen and Oinas-Kukkonen (2013). Then, qualitative research as a research method is generally introduced among the main goals and concepts of it, including topics such as case study as a qualitative research method and documents as a method to gather qualitative research data. The document sources and assessment criteria are introduced, and after the research, results are analysed, discussed and concluded, including the reflection against the research questions.

2. Prior research

This section of the thesis examines the prior research of the topic. This includes several topics and themes, such as the definition of voice-based user interfaces, their history, comparison against other mediums to interact with interfaces, modern voice-based user interfaces and the conceptual model of six paradoxes, which is used to reflect the examined user interfaces against.

In the 2010s, voice-based user interfaces have achieved mainstream popularity due to the launch of Apple's Siri (2011), Microsoft's Cortana (2015) and Amazon's Alexa (2015) (Luger & Sellen, 2016). These recent emergences do not however mean that voice as an interaction medium has not been around before – there is a relatively long history with these particular interfaces, and the section of prior research starts by examining the history of voice as a medium in human-computer interaction and what have voice-based user interfaces achieved during their existence.

2.1 Voice as a medium in human-computer interaction

The basis of voice-based user interfaces is the interaction of human and computer: the human actor uses the voice to control a computer or other elements in the surrounding environment (Rogoff, 2001). In the 1990s, there were several speech recognition systems available in commercial markets. During the time, these systems had the percentage of 95 to 97 when measuring the performance, the recognition rate of the queries. (Delogu et al., 1991)

In the late 1980s and early 1990s, the research has tackled the challenges of voice input/output systems in human-computer interaction: evaluation criteria and standards were needed to achieve the full potential of the systems. During the time, paying attention to human factors in voice was considered to be the essential factor in order to reach larger markets. (Delogu et al., 1991)

Humans do not act in the same way as the computers when having conversations – they adapt to the different situations by forming the conversation in natural language. There are several hindrances against successful human-computer conversation. Humans do not always use grammatically correct and explicit sentences, causing ambiguous indirect messages which do not explicitly successfully tell the right message. Humans also tailor their speech to different conversation partners: for example, with friends or strangers, the tone, pitch and other voice attributes vary. (Brennan, 1991)

These largely varying situations are not the only problems when considering voice as a medium in HCI. Delogu et al. (1991) continue with several other problems, such as vocabulary, speakers and noise levels in the environment. Vocabularies, the collections of used words, form two big problems: first, vocabularies always contain acoustically similar but by meaning otherwise very different words (for example, "beam" and "bean"), and second, the size of the vocabularies usually is not very small. These are also connected: when the vocabulary's size grows, the acoustically similar confusing word pairs also grow, leading to understanding problems. (Delogu et al., 1991)

In the early 2000s, Rogoff (2001) presented some use cases for voice activation, such as giving simple commands to different "smart" objects. For example, car's headlights could be activated with a simple "headlights on" command. This is continued with the problem of variance with different voices: pitch can be high or low, and gender also can affect the results of a query. At the time, a lot of time was needed to even get started with voice activation: the software must have been trained by reading big amounts of text to the computer. (Rogoff, 2001)

In the early speech recognition systems, the count of supported words varied from 40 to 30 000 (Delogu et al., 1991). Brennan (1991) states that although during the time in the early 1990s significant progress with forming mappings of database queries was made, the human language use is so complex that there lies a pessimism whether there will ever be useful natural language interface technologies. These databases are examined further in Delogu et al.'s (1991) research: several corpora (entities of languages and text) were developed to help the evolution of voice input/output systems.

To summarize, the early main struggles of the early voice-incorporated interfaces were based on the complexity of natural language, meaning the variance of speakers and how do they interact with each other. This complexity caused problems when implementing it to computers, which rely heavily on explicit language. (Brennan, 1991; Delogu et al., 1991)

Dubiel (2018) presents some observations on the topic almost 30 years later – what has changed, or has anything changed at all? The context is one of the main problems: today's computers still have struggles to distinguish the meaning and intent of human actor. This causes that even if there are successful queries which somewhat satisfy the intent, the human-machine conversation can either be lacking, or worse, longer than it should be, due to system's need to ask further clarifying questions. (Dubiel, 2018)

There have been several, constantly evolving problems with successfully using voice as an interactive medium, and also the shift of usage has happened. Rogoff (2001) saw that voice activation is the future of user interface, meaning that entities could be activated with queries, without the conversational aspect between human and machine. However, the shift has been going towards two-sided communication, not one-sided. Artificial intelligence plays a big role in this: the systems are part of the conversation, not just taking commands with no conversational aspects. For example, these virtual assistants have evolved from car navigation systems to more extensive, conversational agents with more human-like features. (Guzman, 2019)

2.2 Acceptance of the voice as medium to interact

Dubiel (2018) has listed some reasons for voice interaction's limitations: the lack of conversational flow and the constant query clarifying of the machine in order to achieve the intent of the user. One constantly debated question is the machine's output when considering voice-based user interfaces: does the system need to be more human-like with skills to maintain conversations, leading to interaction which reflects more human-human like dialogue (Dubiel, 2018)?

HCI has started to take voice more into account in recent years, due to the recent popularity of conversational agents such as Apple's Siri, Amazon's Alexa and Microsoft's Cortana (Luger & Sellen, 2016). These agents are considered to be smarter than previous attempts of speech technologies, as they are enabled by artificial intelligence, making them more human-like with more communicative, two-sided conversation flows (Guzman, 2019).

According to recent conversational voice assistant consumer adoption report by Kinsella and Mutchler (2018), the main problem with acceptance of voice as a medium of interaction is the lack of understanding the user. Up to 47 percent of consumers report that they would use more voice-based virtual assistants in a smartphone if they understood better the user. One third of the consumers hope for more features. Interesting observation is that one user out of five would use voice as interaction medium more, if it is prompted. As the medium is not habitually established, many users are not even aware of the option. What the consumers do like with voice is that it is hands-free, and also some feel it is more convenient and faster than touching and interacting physically with the device. This does not mean that the "touch and type" interaction is going away – voice is felt useful in certain situations as a complementary user interface, thus it is not considered as the main one. These numbers of the consumer adoption are based on survey's results of 1 040 US citizens. (Kinsella & Mutchler, 2018)

Interesting human-machine study conducted in 2003 by Shechtman and Horowitz examines how the experiment's participants behave in different conversational situations. The participants had structured conversation with a computer – the main difference between the situations was that the participants were told that they were interacting with either the computer itself or a real person. As a result, the participants who believed that they interacted with another person became more emotional and used relationship statements, such as influencing and hostility. (Shechtman & Horowitz, 2003)

Interesting observation is also that although human-human interaction simulation is usually a desired outcome with human-machine interaction, there sometimes are benefits when trying not to achieve human-like properties. The participants who believed to be interacting with another person put more effort into communication, which can hinder the task's effectiveness. Another aspect is the emotional side: negative emotions can hinder the effectiveness, and machine's intervention to this could be more suitable than human-like responses. (Shechtman & Horowitz, 2003)

Social behavior is considered to be essential part of human-human relationships, but is it essential when considering human-machine interaction? Nass et al. (1994) have implemented social factors to five human-machine situations, to find out what is to be learned from social rules in these kinds of interaction. Rather interesting situation is the engendering of computers: does it matter if the voice of the computer is gender-based? Based on the study, the 48 subjects' results showed that gender stereotypes in human-human interaction apply with computers, too. (Nass et al., 1994)

Another conversational agent question, the perception of the agent itself by humans, is tackled in the study. The participants used a program with slight tweaks on how the computer refers to itself – either in 3rd person or in 1st person. Overall, it was concluded that the self-referential behaviour of the computer did not affect the results on social influences when using the computer. (Nass et al., 1994)

Nass et al.'s (1994) research is rather old but offers an exciting pavement for the future. This experimental framework of five social factors has become to be known as CASA (Computers Are Social Actors) framework, meaning that humans apply social rules and other relevant social aspects to computers. This is further elaborated in later research by Nass and Moon (2000), where a concept of mindless responses is examined further and their role in social rules with computers. Mindless responses are social cues in different contexts, which make humans act accordingly and take the attention away from other information. In order to achieve these social responses, there must be enough cues to be presented in order to elicit social responses. (Nass & Moon, 2000)

Individuals rarely consciously acknowledge computers as social actors, due to typical computer's drastic look dissimilarities to humans. Mindless responses do provide some explanation why social rules are still applied to computers, but the exact precision on how to create these responses to elicit different types of responses are unknown. The variance whether humans see the computer as a tool or possibly as a social actor relies on this rather unknown mindless response theory. One presented possibility is that embedding heavily socially related interaction types, such as voice, generates more social behaviour towards the computer. According to the research, rather unknown still are the exact social rules which are applied and how they are emerged. (Nass & Moon, 2000)

The modern conversational agents such as Apple's Siri are much later examined after the emergence of CASA framework, and the focus has been especially on the social features of these particular virtual agents. These mobile agents have extended the mobile device's main purpose: before, they were used to listen and speak to another human voices, but for now the voice has extended to include the phone itself. How do people see this new voice? (Guzman, 2019)

Source orientation is essential when determining how people feel about the voice coming from the device itself. This means determining how people see the source of the message and how does it affect when interacting with the source. Conversational agents are much more complex than previous attempts of voice-based user interfaces: they have several layers, such as vocal abilities and artificial intelligence – but their location is vague. (Guzman, 2019)

Some participants felt that they are communicating straight to the device, whereas some felt that the voice coming from the device is transmitted through the device – which makes the device a medium, not the source itself. This distinction makes the assistant separate from the device. Overall, the perceptions vary a lot when considering conversational agents, from the device itself to a separate assistant, and one participant even said that she feels how the phone has "like a little person" inside it. (Guzman, 2019)

Guzman's (2019) research shows various perceptions on how the modern conversational agents are seen, and this could be explained by anthropomorphism. This concept is about how humans have tendency to see or embed human-like characteristics into non-human agents or objects. These cues are seen to make the non-human agents of social connection more powerful, especially when the other levels of human connection are low. Also, these cues can make the technology more efficient when users are learning how to use it. (Epley et al., 2007)

Epley et al. (2007) also state that anthropomorphism has a big impact on how the human-computer interaction is shaped in the future. This is due to the fact that everyday life is more and more based on interacting with life-like virtual agents and technologies – anthropomorphism should be the base for designing agents which are easily understood and learnt. The goal is to create social connections in human-computer interaction, just like in human-human interaction. (Epley et al., 2007)

Recent study by Forster, Naujoks and Neukum (2017) has put anthropomorphistic cues into practical case to find out whether trust is increased when adding these characteristics into conditionally automated driving functions. 17 participants performed two simulator drives, one with speech output and one without. After this, a questionnaire was filled to find out whether anthropomorphistic cues of speech in the human-machine interaction increased trust. It turns out that the speech helps with building the trust: the automation is felt safer when there is a speaking agent telling

explicitly what is going to be done next in the simulator. These descriptions also made it clearer for the participant about what is going to happen, this leading to increased trust. (Forster et al., 2017)

Study by Cowan, Branigan, Obregón, Bugis and Beale (2015) examines the anthropomorphism in voice and how it affects choices in a certain task. First, the participants were asked to evaluate several voices, from robotic to extremely anthropomorphic. The more anthropomorphic the voice was, the more advanced, flexible and competent it was considered to be. The task itself was a picture/name matching task – the main difference was that it was performed in three variances: with a human partner, robotic computer or anthropomorphic computer. (Cowan et al., 2015)

After the experiments, it was concluded that conversational partners adapt their language even in human-computer dialogue. This however was not said as the exact truth, as the environment was a highly controlled laboratory experiment. The researchers' hopes for the future are outside of the laboratory, with real-world examples of how the voice affects the communication between a computer and a human. (Cowan et al., 2015)

2.3 Modern voice-based virtual assistants

In the prior research, several modern voice-based virtual assistants are mentioned, such as Apple's Siri, Amazon's Alexa and Google's own services (Guzman, 2019; Luger & Sellen, 2016). These services and devices provide presence in ubiquitous way, making them very accessible for different types of persons. Ubiquity allows the users many times more convenient and user friendly way to interact, as tasks can be done hands- or eyes-free. (Jeong & Shin, 2015) Figure 1 shows at a glance the timeline of modern voice-based virtual assistants with the introductions and launches in US markets of Apple's Siri, Amazon's Alexa and Google's Assistant. During the time, the launches have also expanded to other countries, such as UK, Germany, France, Japan and India. ("Voice Assistant Timeline," 2018)

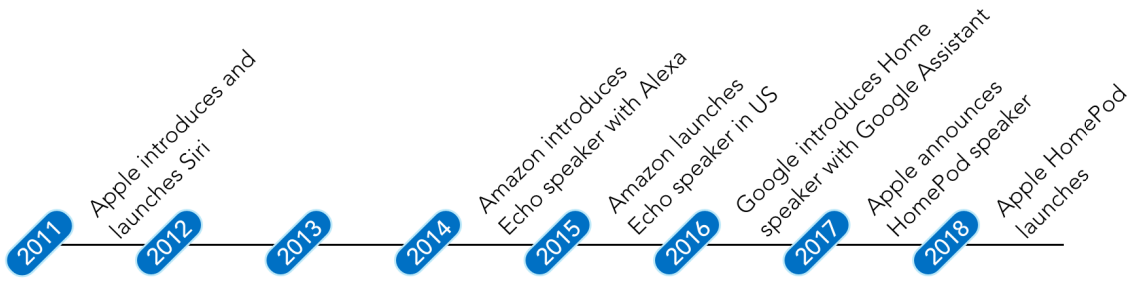

Figure 1. Timeline of three main voice-based virtual assistants (based at "Voice Assistant Timeline," 2018).

Kinsella and Mutchler's (2018) report on voice assistant consumer adoption divides the era of modern voice-based virtual assistants into two phases. The first phase started in 2011, by Apple's iPhone 4S smartphone and the introduction of Siri, Apple's own voice-based virtual assistant. Apple was quickly followed by Amazon's Alexa and Google's Google Assistant. The first phase evolved quickly to several device types to utilize voice, such as smart speakers, headphones, cars and other appliances. Phase two is ongoing – while the first phase's focus was to get users familiar with voice as an interaction medium, phase two is about making the voice persuasive with even more capabilities and multi-device integration. (Kinsella & Mutchler, 2018)

Table 1 shows the US market shares of voice-based virtual assistants in smartphones and smart speakers in 2018. By November 2018, Apple's Siri owns up to 44 percent of voice assistant market shares in smartphone markets in the US. The second place goes to Google's Google Assistant with around 30 percent, and third place is for Amazon's Alexa with around 17 percent. The smart speaker market numbers were considerably different to smartphones: Amazon is the dominant leader with 64,6 percent market share, followed by Google (19,6 percent) and Apple (4,5 percent). (Kinsella & Mutchler, 2018)

Table 1. US market shares of voice-based virtual assistants in smartphones and smart speakers (Kinsella & Mutchler, 2018).

	Apple Siri	**Amazon Alexa**	**Google Assistant**	**Others**
Smartphones	~44%	~17%	~30%	~8%
Smart speakers	4,5%	64,6%	19,6%	11,3%

The drastic variance on Apple's and Amazon's smartphone and smart speaker market shares is explained by the emphasis differentiation between the companies: Amazon has been around much longer in smart speaker markets than Apple, which entered the smart speaker markets in February 2018 (Kinsella & Mutchler, 2018). As per these statistics, the top leaders are Apple, Amazon and Google, thus these companies and their voice-based user interfaces are examined in the next sections in order to gain insight about today's modern voice-based user interfaces.

Although the main devices with voice-based virtual assistant use cases are smartphones and smart speakers, voice has been integrated into other platforms as well. Cars especially benefit from voice as a medium of interaction, due to requirements of focusing on driving and voice offering hands- and eyes-free way of interaction. Another example of voice's benefits is with the smartwatches: due to small screen sizes, navigation and other functionalities can be difficult to implement but voice as an interaction medium can reduce these difficulties. (Kinsella & Mutchler, 2018)

Voice queries' accuracy with modern voice-based virtual assistants is examined very recently by Munster and Thompson (2018). In the test which was concluded in December 2018, the smart speakers by Apple, Amazon and Google were asked total of 800 questions each. The test had two metrics: first, did the speaker understand what was said and second, did it deliver the correct answer. Questions were in five categories: local surroundings, commerce, navigation, information and command. (Munster & Thompson, 2018)

The understanding of the queries varied between 99 to 100 percent, but the correct answer percentage was between 73 to 88 percent when comparing the three biggest players in the market. Interesting observation with the almost perfect understanding is that the misunderstood queries had a proper noun, such as a town or a restaurant – this emphasizes that the natural language processing has come far from its early days and most of the queries are understood with no problems. In only a 9 to 12-month period, the virtual assistants have improved their correct answer scores by 7 to 22 percentage points, due to natural language processing advancements. The most drastic improvement (22 percentage points) comes from Siri, due to enabling more domains where queries could be asked. (Munster & Thompson, 2018)

Due to the difference of use cases between smart speakers and smartphones, separate yet similar test was concluded with smartphones earlier in July 2018 also by Munster and Thompson (2018). For example, Apple's Siri on the phone can be used in very different situations than Amazon's Alexa on Echo speaker in the kitchen. The understanding varied between 98 to 100 percent, and the correct answer percentage was between 61 to 86 percent. (Munster & Thompson, 2018)

The numbers do not differ drastically between smart speakers and smartphones, but one interesting observation is the Amazon Alexa's scores: the dominance of smart speaker use cases is clearly visible due to better score in that category (73%) compared to smartphones (61%). Although separate testing was conducted, both of these researches come to the same conclusion, no matter what the device type is: with constantly evolving technology and correct answer scores of 80 to 90 percent, the assistants are allowing more and more queries to be performed with voice (Munster & Thompson, 2018; Munster & Thompson, 2018).

2.3.1 Apple Siri

Apple's voice-based virtual assistant Siri was introduced in 2011 along the company's smartphone iPhone 4S. The phone's marketing heavily relied on Siri's features, depicting that voice as an interaction medium would be "the next thing". Initially, the assistant was rather limited due to language, feature and performance issues. As these early stage issues were resolved, Siri has achieved the market leader's place in smartphones' virtual assistants. (Kinsella & Mutchler, 2018)

Since 2011, Siri's support has expanded from smartphones to a large range of devices, such as headphones (AirPods), smartwatches (Apple Watch), Mac computers, cars, and most recently, smart speakers (HomePod). Siri can be accessed in various ways, for example by pressing a certain button on the device, or accessibility-wise better, saying "Hey Siri" without any physical interaction with the device. ("Use Siri on all your Apple devices," 2018)

Apple emphasizes the hands-free and on-the-go features of Siri, for example with making phone calls or sending text messages without any device-tied physical distractions while driving. Another showcased use case category includes everyday tasks, such as alarm setting, traffic information, reminders and calendar schedules. For example, a timer could be set by saying "Hey Siri, set a timer for 20 minutes". ("Siri," n.d.)

Due to increased maturity of artificial intelligence, the smarter context understanding of the virtual assistants has improved (Guzman, 2019). This shows for example in the following music-related queries of Siri: "Hey Siri, play me something I'd like" and "Play the latest Beach House". Siri gets the personal music taste from user's listening history in Apple's music streaming service and also understands the word "latest" without explicitly stating the artist's latest album's name. ("Siri," n.d.)

Siri's capabilities have also expanded outside the device itself, forming larger entities such as home peripheral activation by saying "Hey Siri, turn on the lights in the living room" or "Hey Siri, warm it up in here". Machine learning and personalization are highly used in the service: Siri can suggest new tasks and activities based on older ones, and for example, Siri can be taught who are the user's family members. This way, instead of saying "Call [sister's name in contacts]", the user can say "Call my sister". ("Siri," n.d.)

Siri is considered somewhat lacking feature-wise when compared to rivals such as Google Assistant, but it still remains as US market leader of virtual assistants in smartphones with the market share of around 44 percent. This is due to the fact that the Apple's device base is simply larger in the US and more controlled than with Google. (Kinsella & Mutchler, 2018) This market dominance also shows in the prior scientific research: both Guzman's (2019) and Luger and Sellen's (2016) research state that Siri has overrepresentation among the participants, and future research should address this by gathering more diverse virtual assistant base for the research.

2.3.2 Amazon Alexa

Amazon's solution Alexa has a different approach compared to Apple's Siri: whereas Siri pioneered with smartphones and stays dominant in that area, Amazon Alexa reigns in smart speaker markets, due to Amazon's pioneering introduction of Echo in 2014, their own smart speaker system. This shows in the US market shares: in smartphones, Amazon Alexa has only around 17 percent share, whereas in smart speakers the share is dominant 64,6 percent. The rise of smart speakers has been incredibly rapid in the US: in January 2018, 47,3 million US adults used a smart speaker and just in nine months later in September of the same year, the number had risen to 57,8 million. (Kinsella & Mutchler, 2018)

Amazon offers variety of Echo smart speakers with different price ranges for different needs, and all of them have Alexa as a virtual assistant. The use case categories highlighted by Amazon are similar to Apple's ones: Alexa can be used for example in entertainment, communication, productivity, smart home and information gathering purposes. ("All things Alexa," n.d.)

Lopatovska et al. (2018) have examined specifically Alexa's uses among users. In a span of four days, several types of interactions were determined, for example checking weather, finding facts and listening to news. From the interaction types, weather and music inquiries were the most popular ones. The research states that these might not still be the essential features of a smart speaker, as users discover the strengths and weaknesses of the service and when the needs vary in time. Time variance plays a big role in adoption of smart speakers: those who had owned a device with Alexa less than three months, had average uses of 2,75 per day. Participants who had owned the device for more than a year used the device only once a day. (Lopatovska et al., 2018)

The main difference between a smartphone and a smart speaker is the mobility: speakers are for home use, whereas the smartphones are on-the-go solutions. The average number of smart speakers in household has risen from 1,8 to 2,2 devices between January and September 2018, due to low price of the speakers and making the voice available in multiple rooms in households. (Kinsella & Mutchler, 2018) In Lopatovska et al.'s (2018) research, three room types were detected for most common placements for Amazon's Echo speakers: living room, kitchen and bedroom. Due to a stationary device, the placements most likely reflect to common activities performed in a certain room, for example weather inquiries when dressing up accordingly in the bedroom (Lopatovska et al., 2018).

Interesting connection is with Kinsella and Mutchler's (2018) phase two declaration of voice-based virtual assistants (persuasion with advanced capabilities and multi-device integration with the power of voice) and participants with advanced skills with Alexa in Lopatovska et al.'s (2018) research. These advanced users controlled other devices with the help of Alexa five times more frequently than other users, and also these users were more frequently understood by Alexa. This sets an interesting setup: what unlocks the

full potential of Amazon Alexa among the "basic" users so that they become more advanced users? There is no clear answer, but the research states that it is unclear whether the system has the vitality for long-term adoption. The lack of vitality at least is shown in the decline of average uses per day in long-term usage, but this requires more profound research and more participants in order to state this more concretely. (Lopatovska et al., 2018)

While Lopatovska et al. (2018) have more general approach for Amazon Alexa, Pradhan, Mehta and Findlater (2018) have examined Alexa with more specific approach: accessibility with disabled people. The authors have examined product reviews of Echo smart speakers, by filtering the reviews with a list of keywords to find reviews which mention users with disabilities. Overall, the reviews were rather positive, for example with highlighting the assistant's companionship with disabled persons and helping with everyday tasks. The voice-based assistant is especially helpful with visually impaired people – the voice as a medium makes tasks much easier, as no visual cues are needed. (Pradhan et al., 2018)

There are however problems – for example, the commands can be difficult to be memorized and sometimes they are difficult to pronounce and deliver correctly. One distinct improvement point emerging from the research is the ecosystem of Alexa – it does not work on its own, as it requires a device ecosystem. For example, setup and troubleshooting is handled by a smartphone application, which can restrict the accessibility as the benefits of voice-based user interfaces are hindered by other mediums. (Pradhan et al., 2018)

2.3.3 Google Assistant

Although Apple and Amazon were the first ones in virtual assistant markets in 2011 and 2014, Google has caught up the competition a lot after introducing their own solutions in 2016. Although Google's Android mobile operating system's ownership share is almost the same than Apple's, Google Assistant has only around 30 percent market share of smartphone voice-based virtual assistants compared to Apple Siri's around 44 percent. In smart speakers, Google has market share of 19,6 percent, which is only around one thirds of Amazon Alexa's share but still four times more compared to Apple Siri's. (Kinsella & Mutchler, 2018)

As a technology giant, Google has taken several different device types into account when designing the voice-based virtual assistant. The company offers both smart speaker (Google Home) and smartphone solutions, and also smartwatch, computer, TV, car and smart display solutions. Google also emphasizes the extensive support of the service for 3rd party applications, such as music streaming service Spotify and media streaming service Netflix. The categories of use case possibilities are similar to Apple and Amazon's offerings, ranging from information gathering to smart home peripherals. ("Google Assistant," n.d.)

2.4 Conceptual framework: six paradoxes

In order to approach the research question and the research itself, the conceptual framework of six paradoxes presented in Oinas-Kukkonen and Oinas-Kukkonen's (2013) book are examined. These paradoxes are originally presented against the social web: for example by examining how privacy is taken into account in social web, or how does the social web provide value to its users (Oinas-Kukkonen & Oinas-Kukkonen,

2013). These paradoxes are presented in the table 2 below, and furthermore opened in their own following chapters.

Table 2. Six paradoxes by Oinas-Kukkonen and Oinas-Kukkonen (2013).

Paradox	**Example evaluative questions**
Privacy	What should or should not be shared? How much personal information is needed to use the service or device? Where goes the boundaries of user's privacy, for example regarding health details and web browsing history? What is the user's responsibility on privacy when using the services or devices?
Identity	How much is known about the user? How does user's behavior change when acting anonymously? How does the identity in real life mix with the identity on social web or other services?
Credibility	How credibility and trust is created and earned in the services? How to know what information is trustworthy?
Friend	Is the quality of friendships quantitative or qualitative (the number of friends vs. the quality of these friendships)? What is the effect of social web relationships on deeper relationships? How does the narrowness of service's communication possibilities affect relationships?
Filter	How to find meaningful information? How to filter out irrelevant information?
Value	How do the services or devices provide value to users? How to avoid producing non-valuable information?

The paradoxes vary a lot and thus offer a great guidance for the research questions, offering an extensive view for voice-based user interfaces. For example, by the paradoxes, privacy aspect can be evaluated through the views on constantly "hearing" microphones of the devices, or value paradox can be evaluated with the view on accessibility and value-providing aspects for disabled people. Also, the improvements and the future can be evaluated through the paradoxes – if there are some issues in a certain paradox, it can work as an improvement point for the future's voice-based user interfaces.

2.4.1 Privacy paradox

Privacy paradox handles issues about user's boundaries. Today in order to be used, many services require registering and quite a lot of personal information about the user. Users are many times blind when sharing sensitive data: they believe that as the data is electronic, it is not going to last forever. The constant debatable question is about the use of the data: is the data really needed to use the service or do the service providers have deeper true interests for data collection which are not disclosed to the user? Especially with new technologies and services, deep examination on possible misuses of the innovations are needed in order to avoid them. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

Recent study by Blue, Vargas and Traynor (2018) examines privacy issues in voice-based user interfaces. In the research, security issues considering the authentication of the voice and distinguishing of human and electronic speakers are examined. Most of the modern voice-based user interfaces do not have user authentication, leading to security vulnerabilities. Electronic voice does have a certain distinction compared to human voice, called "sub-bass over-excitation" phenomenon, meaning low-frequency elements in audio. This feature of electronic voice is fundamental to every electronic speaker, no matter how good the quality otherwise is. This feature can be detected and thus, it is used to distinguish whether the query giver is human or an electronic device. (Blue et al., 2018)

Although the responsibility of respecting user's privacy boundaries is many times on the service provider, the user itself is also responsible for sharing information. For example, considering a situation where the user shares not so formal picture of themselves, the consequences can go quite far although the user might not realize it. Possible new employer might find it later and make assumptions just based on the picture. Privacy is a vast and complex matter: no matter how careful the user itself might be, other people and service providers also have the power to ruin this particular user's privacy. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

Lau, Zimmerman and Schaub (2018) have approached the privacy issues of smart speakers, by interviewing users and non-users of these devices. For example, non-users many times were highly concerned about giving companies access to their homes, thus the trust on these interfaces was many times lacking. On the other hand, users believe that the devices record voice only when needed, due to the need of extremely demanding computational power for storing and analyzing all the data if everything was recorded. One distinctive difference between the users and non-users was that the users were comfortable to trade their privacy for convenience, whereas non-users were strongly against this. (Lau et al., 2018)

Mukhopadhyay, Shirvanian and Saxena (2015) have studied voice and stealing it to create voice impersonation attacks to fool humans and machines. Today's advanced speech synthesizers are able to create someone's voice with extremely limited number of samples. Voice samples are everywhere, as voice itself is one of the most fundamental features of communication enabling, making it easy to create artificial malicious voice actors. This sets services which use voice biometrics very vulnerable, and also can make human-human communication extremely harmful as the artificial voice can basically say anything the malicious actor wants, leading for example to reputation loss or safety issues. (Mukhopadhyay et al., 2015)

2.4.2 Identity paradox

Mukhopadhyay et al.'s (2015) voice impersonation study tackles the privacy issues on gaining the voice samples from the victim, but furthermore the study is about faking someone's identity. These two paradoxes (privacy and identity) are strongly related to each other (Oinas-Kukkonen & Oinas-Kukkonen, 2013). The study took two very famous celebrities, Oprah Winfrey and Morgan Freeman as the voice sources. There are tons of their voice samples on the Internet, and some of them were used to convert the samples into an artificial, new voice. In addition to this, female and male participants of the study were asked to mimic these celebrities' styles of speaking and emotion – from the participants, the best impersonators were chosen to record the samples in non-noisy environment. These samples were also converted into a new artificial voice. (Mukhopadhyay et al., 2015)

In order to find out whether the artificial voices can fool humans, three different types of celebrity voices were presented: the authentic ones collected from the Internet, artificial one created based on the authentic ones and artificial one created based on the human impersonators. The results showed that the original speaker was identified correctly with accuracy between 89 to 92 percent. The impersonator conversion voice was clearly recognized to be a fraud with accuracy of 95 to 98 percent. What makes the results quite worrying is the recognition percentage of converted artificial voice based on authentic voice samples – only around 50 percent of these were recognized correctly. (Mukhopadhyay et al., 2015) Combining this to the statement that most of the current voice-based user interfaces do not authenticate their users makes an alarming example of misuse possibilities (Blue et al., 2018).

Creating persuasive artificial voices just based on a few samples is alarming, making the identity paradox important to be examined with voice-based user interfaces (Mukhopadhyay et al., 2015). From a social web point of view, identity paradox is about how the representation of user varies between the web account and real life. These different representations of a person and their distinctive boundaries can become blurred, causing trouble to each other. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

Furthermore, identity in voice-based user interfaces also can be examined from another point of view: instead of the user's identity, what is the identity of the machine's voice? In Guzman's (2019) research, the perception of identity of the machine varied a lot – some interviewees felt that the voice assistant is a separate entity from the device itself, whereas others felt that the device itself is speaking without distinctive assistive entity. Also, the identity paradox extends to recognizing the speaker – most of the current voice-based user interfaces do not authenticate their users and the distinction of human and machine voices can be difficult to distinguish (Blue et al., 2018).

2.4.3 Credibility paradox

Believing everything from every source creates the credibility paradox. In today's massive amounts of information, it is essential to recognize reliable sources and the right, needed information. In credibility, gaining and maintaining the trust is essential. What makes gaining the trust tricky is the variance of people: some people easily believe everything for example on the web, whereas some people are extremely cautious. Credibility maintenance is extremely sensitive, as one big, or even small mistake can collapse the organization's reputation, which is many times very hard to remedy. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

Credibility and machine-generated voice is examined by Nass and Lee (2000). In the study, human voice characteristics of personality were implemented in a machine-generated voice. Personality is affected by several aspects: for example, gender and age have roles in forming person's personality. Personality however is not formed only by these attributes, as it is more complex entity. (Nass & Lee, 2000)

Is it possible to convince users better (i.e. be more credible) with certain settings in machine-generated speech? In the study, introverted and extroverted students were recruited as participants to an experiment where different text-to-speech voices read book descriptions, and the machine-generated speech was adjusted according to participants' personalities, either introverted or extroverted. The participants were constantly reminded that the voice is not by a human, but still, they were more convinced if the voice reflected their own personality. This leads to some conclusions: for example, machine-generated voice for match-making services should be extroverted, as the audience is most likely more extroverted. By this evidence, credibility is higher

when personalities match between the user and the machine, no matter if the personality is artificially made in a machine-generated voice. (Nass & Lee, 2000)

2.4.4 Friend paradox

Friend paradox is more applicable to social web than voice-based user interfaces, as it is about the quality of friendships on the web. Many people today are obsessed with the quantitative, number-based aspects of friendships, while forgetting the qualitative, quality-based attributes. This might lead to an unbalanced situation, where the deeper, meaningful relationships are neglected. Communication can become brief and lose its meaningfulness when mixing numbers with the lack of quality. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

The capacity of handling deeper connections does not rise when acquiring "new friends" online – it stays the same, although some might believe they are extremely social and having meaningful connections when handling hundreds of relationships online. Thus, the capacity stays the same whereas the number of connections is constantly growing, which can fool some people to think that they are fully socially happy. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

What is the role of voice in relationships, especially when considering human-machine interaction? Anthropomorphism is one aspect which affects the persuasiveness of the voice: making it more playful can make the voice more "human" and thus, more persuasive. This can however be dangerous for voice interface designers: the expectations go higher for the interaction and user experience if the voice starts to be too human-like, with for example humorous responses. (Luger & Sellen, 2016)

In Pradhan et al.'s (2018) research on voice-based user interfaces with disabled people, many users describe the assistants as "new best friend" or "someone to talk to". One user even mentions how most humans do not understand the user, but the device does (Pradhan et al., 2018). These observations are also present in Moussawi's (2018) research – one user mentions how Siri is very helpful and friendly and sometimes cheers the user up when feeling sad. Another user also considers Siri as a friend. The users many times feel that they are grateful towards the assistants and mention also how they have made the user laugh, thus indicating that friend paradox could be applied to voice-based user interfaces. (Moussawi, 2018)

2.4.5 Filter paradox

Today's flood of information on the web and other platforms has created the filter paradox – how to find meaningful information among the massive amounts of it? On the web, one solution is the friend network: person's network members share something which most probably is meaningful for them, thus, it might also affect the receiver of the information. This however is paired with friend paradox: with massive networks of loosely connected friendships, the shared content can also be very low quality. Filtering can also lead into creating own habitats with connections with same interests – this can either be a positive or negative thing to happen. For example, radical groups can easily organize over the web and share their bad intentions. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

Voice as a medium itself somehow avoids the filtering problem, due to the rather explicit requirements of the queries. This however has been investigated and developed further, as natural language is more and more integrated to the machines. In order to

avoid the massive amount of information and query possibilities, machines many times rely on domains, these meaning areas of particular topic which define the boundaries of the queries. By this way, the input possibilities are restricted to the domains, making the designing of the voice-based user interfaces easier. (Brennan, 1991)

Ambiguity of human language, which leads to requiring explicit queries is especially a voice-related problem. Human language is largely ambiguous, and the devices can have hard time to interpret the query as there can be several possibilities. (Mankoff, Hudson & Abowd, 2000) The ambiguity leads many times into query suggestions, these meaning follow up questions which clarify the user's intention. Without these, the filtering and handling of the initial query has a big chance of failing. Clarification questions itself do not hinder the user experience, in fact, they might sometimes have a positive effect on it. (Kiesel, Bahrami, Stein, Anand & Hagen, 2018)

Apple's Siri supports several domains for voice-based interaction, thus restricting the possibilities of the service in order to eliminate ambiguity as much as possible. These domains include for example fitness, messaging, payments and list management. In domains, there are actions to be performed, known as intents. Siri itself handles the language processing to turn the queries into intents. ("Domains and Intents," n.d.) Table 3 provides some examples of Siri's domain areas and the intent possibilities in them. ("Domains and Intents," n.d.)

Table 3. Examples of filtering via domains and intents in Apple's Siri ("Domains and Intents," n.d.).

Domain (topic area for intents)	**Intents (actions user can perform in the domain)**
Audio and video calling	Initiate calls, search the call history
Fitness	Start/pause/resume/end/cancel workout
Lists and notes	Create/search to-do lists and items, mark as complete, create reminders/notes, search/modify notes
Messaging	Send/read/search for messages

To increase the accuracy and reduce the ambiguity problem, developers do have the responsibility to define custom vocabulary if needed. For example, if a fitness application has certain specific names for different workouts, these are defined in the custom vocabulary by the developer. Pronunciation cues can also be provided, if for example application's name pronunciation varies between users. ("Domains and Intents," n.d.)

2.4.6 Value paradox

The base for value paradox is the intent of obtaining value from using any kind of information system. Value paradox is connected to the filter paradox: some information is just not valuable, causing only "web pollution" to the services. Having the plethora of information available is not always a blessing – in fact, tons of information make no sense at all and offer zero value to anyone. Value paradox is also connected to identity paradox: one might believe that the information they have produced is valuable for their audience, whereas in reality it is not. This false believing might in extreme cases lead into identity distortion. (Oinas-Kukkonen & Oinas-Kukkonen, 2013)

In order to achieve the full potential in information value, skills for finding it are needed (Oinas-Kukkonen & Oinas-Kukkonen, 2013). Domains restrict the area of voice-based user interfaces in order to reduce ambiguity and interpretation possibilities as much as possible (Brennan, 1991). In order to find out the value of voice-based user interfaces, motivations for the use can be examined. One reason for using these interfaces is multitasking when hands are otherwise occupied, such as while cycling to work, while looking after children, or while driving. Time saving also is mentioned to bring value: some tasks are quicker to be done with voice rather than typing. (Luger & Sellen, 2016)

Voice-based user interfaces are great solution when task needs to be performed hands- and eyes-free, making the solution suitable for various environments where other interaction mediums are not possible. Value is easily born in these interfaces as voice is natural to humans as a communication method, and it is used in a daily basis. (Jeong & Shin, 2015) Although the performed tasks can be similar compared with other mediums of interacting, voice's value comes from the naturality itself. The main difference to other mediums is that the intent of query is expected to be understood correctly from the context, meaning that the natural interaction of the voice is understood correctly. (Kiseleva et al., 2016)

According to Kinsella and Mutchler's (2018) report on voice assistant consumer adoption, the top five common use cases are while driving (62 percent), relaxing at home (38 percent), doing household chores (26 percent), cooking (24 percent) and walking somewhere (21 percent). Out of five categories, four are activities which require hands, eyes or both to be occupied on the activity itself. The fifth activity, relaxing at home, shows that in this activity the voice acts a convenience at home rather than necessity due to restricted use of hands or eyes. (Kinsella & Mutchler, 2018)

Voice's value is particularly high in one area: accessibility. People with disabilities and elderly can benefit voice interaction greatly, as it does not force the user necessarily to use the system in a certain physical location. Voice-enabled smart homes can provide autonomous, but still frail persons more fulfilling life by making everyday tasks easier. Interesting take is about the design of these systems – some persons who tested the system considered that they can provide too much value. This means that if the system does too much for the user, such as coffee making or all the other mundane tasks, they are taken away from user's daily routines and thus making the user "useless" when taking away the tasks which the user would be eager to do. (Portet et al., 2013)

In Pradhan et al.'s (2018) research, people with disabilities have mentioned several advantages and value-inducing aspects of voice-based user interfaces. For example, controlling devices without visual cues is considered a huge advantage for visually impaired people. Also, the interfaces are giving possibility to be more independent due to the possibilities without having to use any physical effort. The independence has taken away some tasks from the caregivers, taking also some pressure off the user as they do not feel being constantly annoying, for example when asking the same questions many times. (Pradhan et al., 2018)

Accessibility is radically different in speech interaction compared to more traditional motor skills requiring ones. Initially, voice as an interactive medium has been used to supplement "more traditional" touch-based interactions in mobile devices, but now the accessibility is examined more profoundly, as voice interaction can be done completely hands-free. As the interfaces have previously had primarily supplementary status, the current voice assistants such as Apple's Siri and Google's Google Now do not fully support complete hands-free interaction, which is one thing to be considered in the future of voice-based user interfaces in order to provide maximum value for people with disabilities. (Corbett & Weber, 2016)

3. Research methods and the study

This section of the thesis introduces qualitative research as a research method in general, and also sheds light on the to-be-conducted research itself. The following sections will cover the goals of the research, qualitative research's definition and what are the suitable use cases for this kind of research. In a deeper level of qualitative research, case studies and documents as data gathering method are examined – how they affect the research and what are the goals with these particular topics? In conclusion, the documents' contents and sources and their assessment criteria for the research are examined, for example by presenting deeper details how the product reviews work as documents and how the data from them is evaluated.

3.1 Goals of the research

Rogoff (2001) has stated in the early 2000s that "voice activation is the user interface of the future". This has come to be true, at least in some extent when looking at the rapid growth of voice-based user interfaces in various device types such as smartphones and smart speakers. These devices have worked as a platform for this "new" way to interact and the interaction medium has its own place and use cases in many people's daily lives. (Kinsella & Mutchler, 2018)

The rather quick adoption rising of these interfaces give much potential to research community: for example, the reasons of people for using these interfaces can be examined, or how can them be improved in order to achieve even further audience. Previous studies have tackled the phenomenon from various aspects, such as source orientation (the perception about the machine voice's origin) and expectations when using these voice agents in daily lives (Guzman, 2019; Luger & Sellen, 2016).

The main goal of this research is to gain insight why or why not these voice-based user interfaces are adopted, what are the most common use cases and what should be done in order to gain the full potential from them and thus causing a bigger breakthrough in adoption rates of the interfaces. Kinsella and Mutchler's (2018) voice assistant consumer adoption report has dived into the reasons of quick adoption rates, stating that although there are wishes for improvements for example in understanding the user better, the consumers still feel that these interfaces are constantly improving.

The conceptual framework of six paradoxes by Oinas-Kukkonen and Oinas-Kukkonen (2013) is used to form the base structure to evaluate against the collected data from product reviews – the framework offers six distinctive paradoxes which are very well suitable to examine rather new phenomena, like these interfaces. As there are six of the paradoxes, the results most likely will offer broad perceptions of voice-based user interfaces and not just a few insights with rather narrow results. To make the research even more robust, in addition to the framework of six paradoxes, the base for evaluating the gathered data is extended with other prior research. The prior research works as a platform for forming further categories and questions surrounding the voice-based user interfaces.

One particular problem mentioned in previous studies is the overrepresentation of Apple's Siri – this is due to its market share domination, and most favourable outcome

of the research would be to achieve much larger interface base in order to gain extensive results (Guzman, 2019; Kinsella & Mutchler, 2018; Luger & Sellen, 2016). As the data is collected from a large base of product reviews, the overrepresentation problem will be avoided by selecting Amazon's Echo product family device and Apple's HomePod as devices to be evaluated. Although Apple's Siri is highly represented in prior research, Amazon's Alexa has the largest market share in smart speaker segment, making it easier to narrow the overrepresentation's gap by selecting smart speakers as the base for data gathering.

3.2 Qualitative research

Qualitative research's main goal is to emerge understanding of people and their actions under social and cultural contexts. At the emergence of qualitative research, the need for this kind of understanding was recognized – thus, qualitative research methods were developed to study social and cultural phenomena. The context plays a big role in qualitative research: why does someone act or decide in a certain way in a certain situation – context is used to explain these aspects. Usual questions in qualitative research start with "what", "why", "how" and "when". (Myers, 2013) Table 4 shows and summarizes some of these aforementioned question attributes, their usual surrounding questions and the result goals for this research (Lanamäki, 2017).

Table 4. Qualitative research's questions and research's result goals (Lanamäki, 2017).

Question attribute	**Usual questions to evaluate**	**Result goals and answers for this research**
What?	What puzzles me? What do I want to know more about or understand better? What are my key research questions?	The needed and/or already achieved success factors of voice-based user interfaces. The hindrances of these interfaces. The most common use cases and why especially these are suitable for voice interaction.
Why?	Why will this be of enough interest to others? Can the research be justified as a contribution to knowledge?	Understanding the success and hindrance factors of voice-based user interfaces.
How – conceptually?	What models/concepts/theories can I draw on to answer my research questions? How can these be brought together into a basic conceptual framework to guide my investigation?	Qualitative research as the research method, more particularly case study – documents as data gathering method. Conceptual framework of six paradoxes and other issues from prior research to form the evaluation base for the gathered data. The data evaluation base works also as a base for the overall research.
How – practically?	What investigative styles and techniques shall I use to apply my conceptual framework? How shall I gain and maintain access to information sources?	Product reviews of different smart speakers from Best Buy, the popular US-based electronics retailer. Reviews are available on the retailer's web site.

Qualitative research is suitable in many situations – one example is when depth in subject is needed. Social, cultural and even political sides of people and different organizations can be examined in qualitative research. (Myers, 2013) The debate on HCI's voice interaction and its "humaneness" supports qualitative research as a suitable method for this research, due to the social aspects of machines such as emergence of conversationality in modern voice-based virtual assistants and other anthropomorphistic cues (Epley et al., 2007; Guzman, 2019).

In qualitative research, there are several methods for gathering data. The selected one for this research is examining documents and extracting the needed data from them. The selection of this particular method was due to large number of available documents. Other data gathering methods include for example interviewing, observations, and questionnaires about the subject. All these methods have one thing in common: the data is a record of someone's sayings, i.e. qualitative data. Qualitative data can help understand people and their actions, and these can also put into a larger context to form even further assumptions of the big picture. (Myers, 2013)

The main disadvantage of qualitative research is the lack or difficultness of generalization to larger population. The samples, such as selected documents can be generalized, but when measuring statistically in quantitative (number-based) terms, rather small sample sizes do not have big impact. Thus, samples do not very reliably generalize into a population, whereas in quantitative research large sample sizes can be more easily generalized into a population. (Myers, 2013)

3.2.1 Case study as a qualitative research method

Case studies are used in various contexts – for example, cases can be used in teaching environments to illustrate phenomena or theories in a practical context. In research, cases are used to contribute towards a new theory or exploring an existing one. Research cases are formed by first citing the prior research surrounding the topic, preferably as latest research as possible. The formed case must also be applicated against the topic – the goal in the case is to either discover or test out features which might apply in similar situations. (Myers, 2013)

When defining a case, the simple definition is that the examined phenomenon is an example of a more general, larger category. One main characteristic is that cases are based in real world context. For example, empirical evidence can be gathered from real people in real organization, and thus establish contribution towards knowledge. Case as a qualitative research method thus tries to convince new knowledge into existing one. (Myers, 2013)

Good case is one which is interesting – the main measurement for this is that the case tells something new for the research community. One particular misconception of cases is the number of them – some researchers believe that the bigger the number of cases, the more valid the research is. Even from rising the number from one to four cases does not provide big enough sample to generalize the results into a population. Still, the case must provide sufficiently evidence, to contribute in a solid way to researcher's argument. Cases must also consider alternatives and downsides – as the cases are based on real life, everything does not go always in the same way or as expected. Different theories, views and disagreements have to be taken into account. (Myers, 2013)

As for this research, a case study is conducted with the help of gathering qualitative data from documents. The case itself comes from the conceptual framework of six paradoxes by Oinas-Kukkonen and Oinas-Kukkonen (2013) – in this particular setting, the

framework is the case which represents the larger phenomenon, voice-based user interfaces. Like Myers (2013) states, case studies are used to either discover or to test out new theories: in this research, the prior research of voice-based user interfaces are reflected against this research's results. The results can this way either support or not the existing research, for example, by agreeing with the most common use cases of voice-based user interfaces, and possibly also bring new contributions to the knowledge.

3.2.2 Documents as qualitative data gathering method

Documents are one way to gather qualitative data for research, although it is not as popular as for example interviewing. Although the method is not as popular as some other methods, it can provide richer picture than the others. Documents can be almost anything – they record what someone has said or what has happened. Examples of documents are emails, web pages, records and newspapers: there are no strict restrictions what can be considered as a document. For example, one definition for a document is "anything that can be stored in a digital file on a computer". (Myers, 2013)

Sometimes documents are the only way to gather information, for example in historical research. If a person is dead, they cannot be interviewed but the possible written records by this person could be examined. Also, documents can work as an independent actor: for example, a will of a person "represents" the person after the death. (Myers, 2013) Documents can be public, private and personal – for example, public ones can be governmental publications, private one examples can be ones from private sector businesses and personal one examples can be photo albums or medical records (Mogalakwe, 2006).

For this research, one particular document type was chosen as a data gathering method. The prior research is heavily based on experiences of voice-based user interfaces users, and one way to discover the experiences are product reviews about the devices. As the available number of these reviews is high, they can offer a wide picture of these interfaces, their advantages, disadvantages and most common use cases. For example, in a study by Pradhan et al. (2018), similar data gathering method was done with Amazon's customer product reviews of Amazon's Echo smart speaker devices – the number of available reviews for the study was over 60 000.

Documents can be extremely useful and make the research much easier than for example interviewing, but they have several disadvantages and problems which are easily encountered. One problem is the accessibility to documents – sometimes the researcher might know that there are valuable documents but accessing them is difficult. For example, emails or documents in a distant archive can form an access problem. Another problem is the assessment of authenticity, credibility, representativeness or meaning of the document – if these cannot be assessed correctly, the value of the document might be worthless. (Myers, 2013)

3.3 The research's documents: product reviews and their sources

The product review source for this research is Best Buy, a big US-based electronics retailer. Amazon, another big US-based retailer was also considered as the source of product reviews, but the lack of diversity in sold voice-based user interfaces excluded it as a data gathering source: Amazon sells its own Echo smart speakers, but not Apple's Siri-based HomePod, the rival of the Echo devices. As the goal of the research is to cover voice-based user interfaces in general and not just one manufacturer and its

interface, several manufacturers and interface platforms must be examined in order to achieve reliable results. Thus, Amazon as a source was excluded.

Best Buy is highly based on Internet-based eCommerce, meaning that the customers can shop products online. The product reviews are visible to everyone at the Best Buy's web site and they help the customer when they are pondering whether the examined product is worth the purchase or not. These product reviews can contain almost anything, as long as it is about the reviewed product. For example, in the Apple's HomePod reviews, sound quality, setting up the device, Siri's capabilities and build quality are constantly mentioned. Customers can also filter the reviews by topics, such as by distinct advantages and disadvantages which are constantly mentioned by the reviewers. ("Customer Ratings & Reviews," n.d.) Overall, as time passes and the number of reviews grow, the reviews start to show congruence with the opinions on what is working and what is not – these are shown to the customer as filtering options.

The initial problems with product reviews as a document source for data gathering are the authenticity and credibility, issues mentioned by Myers (2013). How to assure that the product reviews are authentic and credible as a data source? This issue does not only hinder the research, but also the retailers' themselves, as the product reviews can be written by anyone with no possibilities to trace the authenticity of the review's contents.

The issues of authenticity and credibility of reviews have been resolved with one simple verification by the retailer itself: when the reviewer writes and posts their review, they must log in to Best Buy's web site. This account to which the reviewer logs in, also stores all the purchases the reviewer has done at Best Buy – by this information, Best Buy can verify that the customer really has bought the product and is eligible for writing a "Verified Purchase" review at the product's page. Best Buy also allows writing a review even if there is no purchase history with that particular product, but the review does not then have the "Verified Purchase" status. ("Customer Ratings & Reviews," n.d.) The non-verified reviews will be filtered out in this research.

As stated before, several manufacturers and their interfaces for voice-based user interfaces are needed to be examined in order to reach reliable results. The goal is to examine the phenomenon in general, not just one interface. Thus, two biggest players on the markets are selected to be examined through product reviews: Amazon's Alexa and Apple's Siri. Best Buy offers both of these platforms – Amazon has several smart speaker types in several price ranges for different needs, such as Echo, Echo Dot and Echo Plus. Apple has only one smart speaker, the HomePod. From these, the most popular Amazon device, Echo Dot, was selected with Apple's HomePod to be examined through product reviews. This is done so that the number of devices between manufacturers is balanced.

What is especially interesting between Alexa and Siri is that they indeed have the biggest market shares, but in two different sections: Siri dominates in smartphones with around 44 percent market share, whereas Alexa has the clear dominance in smart speakers with around 65 percent market share. This is due to the fact that Apple entered the virtual assistant markets in 2011 with emphasis to smartphones, whereas Amazon started in 2014 with emphasis to smart speakers. Apple entered to the smart speaker territory rather late, in February 2018. (Kinsella & Mutchler, 2018) Apple's advantage is the longer experience with the platform and thus having the possibility to iron the initial kinks out before launching the smart speaker, but Amazon's initial high emphasis on smart speaker markets might still be tough to beat by the recent Apple's HomePod.

3.4 Gathering and assessing the data

As there are two manufacturers and their devices to be examined, adjustments for equal balance between the platforms are made. As Amazon has three main devices and Apple has only one, the numbers of reviews are initially imbalanced. The total number of verified reviews is around 14 300, with the dominance of Amazon's Alexa (three devices in total, ~12 500 reviews, ~87,4% of the total reviews) as Apple has only around 1 800 reviews with its one device (~12,6% of the total reviews). In order to balance the ratio between manufacturers, the number of devices and examined reviews will be fixed. Thus, the most popular Amazon device, Echo Dot, was selected to represent Alexa in this research.

One particular issue with Best Buy's product reviews is the pagination – the product page shows only 20 reviews at a time. This makes the analysis much slower if some kind of data mining techniques are used to highlight the potential reviews, as all the reviews are not on the same page. However, the data mining techniques are out of the question, as Best Buy's terms and conditions state the following:

> You may not use or attempt to use any engine, software, tool, agent, data or other device or mechanism (including without limitation browsers, spiders, robots, avatars or intelligent agents) to navigate or search any Best Buy Property other than the search engine and search agents we provide and generally publicly available browsers. ("Terms and Conditions," 2019)

This leads to a situation where the only way to analyze the reviews without violating the terms is by reading them manually via web browser, as scripts or other automated data mining techniques are forbidden by Best Buy's terms and conditions. As the number of reviews is very high to be read manually, a sample of reviews is taken from both devices. First starting point is selecting the relevant reviews – Best Buy can filter the most relevant ones by assessing the reviews' "helpfulness, recency, pictures and other traits" ("Customer Ratings & Reviews," n.d.). These reviews are listed as the first ones, meaning that the first review fulfills best the given traits, and the last review does the job worst. To get the best results, the Best Buy's relevance filtering is used.

As Apple's HomePod is in clear minority state in number of reviews when compared to Amazon's devices, the assessment of number of the reviews to be read is started with it. Thus, 500 most relevant reviews of HomePod will be read, as the number represents a high share of HomePod's total number of reviews, offering most likely enough relevant data about the device. To balance the manufacturers, the same number of Amazon devices' reviews are read, particularly Echo Dot's reviews as it is the most popular Amazon device.

As stated before, the customers can filter out the reviews according to their interests, as the reviews can contain almost anything, as long as it is about the reviewed product. For this research, the base for product review data evaluation comes from case of the framework of six paradoxes by Oinas-Kukkonen and Oinas-Kukkonen (2013), thus the paradoxes' topics form the main interest point for product reviews' contents. As the research questions do not only cover the voice-based user interfaces through the paradoxes, in addition to the framework, prior research is used to form certain constructions of interest – for example, other common themes about the use of voice-based user interfaces. The figure 2 summarizes the relationship between the gathered data and the base of six paradoxes and other prior research for evaluating the data.

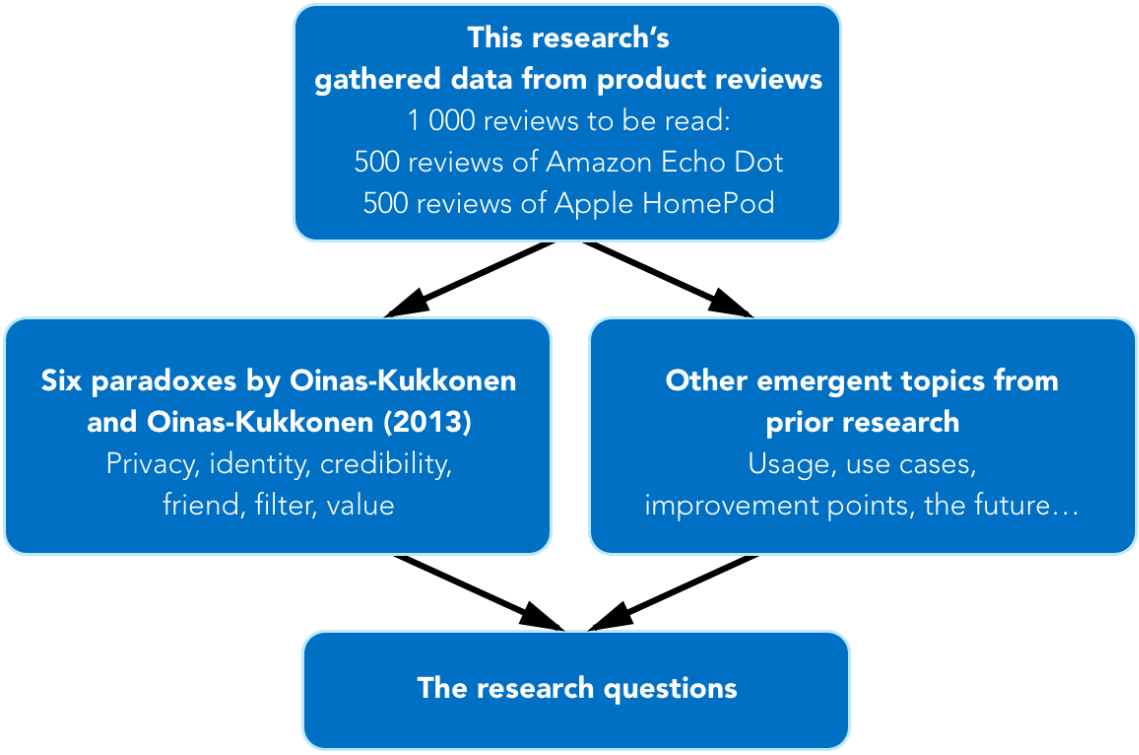

Figure 2. The gathered data's evaluation base with six paradoxes and other issues from prior research.

Improvement points are examined in order to find out possible answers for the third research question about the future of voice-based user interfaces and how they could be improved in order to achieve wider audience. All the reviews which show connection to the selected topics, will be coded according to that particular topic. For example, if the review is talking about privacy issues, it could be used to assess the questions surrounding the privacy paradox. Included as an Appendix A is a table containing the complete base for data evaluation with evaluative questions from prior research.

4. Results

This section introduces the results of the research conducted against the product reviews. In total, 1 000 product reviews were read, 500 of them being about Amazon's Echo Dot and another 500 being about Apple's HomePod. These reviews were collected and then categorized by the main categories of Appendix A. For example, if the review was talking about identity aspects of voice-based user interfaces, it was placed under the second category (six paradoxes). The six paradoxes had clearly the biggest amount of reviews, as it is extensive by its six rather distinctive topics. Results of the categorization of the reviews by Appendix A's categories are presented in table 5 below.

Table 5. Results of the product review categorization.

	Amazon Echo Dot (Alexa)	**Apple HomePod (Siri)**	**Total between devices**
1. Six paradoxes and their connection to voice-based user interfaces	322, ~72,4% of the device's accepted reviews	170, ~46,4% of the device's accepted reviews	492, ~60,7% of the total of accepted reviews
2. Other common themes and use cases of voice-based user interfaces	53, ~11,9% of the device's accepted reviews	33, ~9% of the device's accepted reviews	86, ~10,6% of the total of accepted reviews
3. Improvements points and the future of voice-based user interfaces	27, ~6,1% of the device's accepted reviews	78, ~21,3% of the device's accepted reviews	105, ~12,9% of the total of accepted reviews
Reviews having multiple categories	43, ~9,7% of the device's accepted reviews	85, ~23,2% of the device's accepted reviews	128, ~15,8% of the total of accepted reviews
Total/accepted/ unaccepted reviews	500/445/55 89% accepted 11% unaccepted	500/366/134 73,2% accepted 26,8% unaccepted	1000/811/189 81,1% accepted 18,9% unaccepted

From the 1 000 read reviews, 811 were accepted to be eligible for the research. Exclusion criteria include review duplicates and reviews which were focusing on sole technicalities or comparisons between different generations of devices. For example, the examined device to represent Amazon's Alexa was Echo Dot's 3rd generation, and some reviews only handled sound quality improvements between the current and previous generations. Thus, if the review did not handle voice as a medium or use cases for devices, it was excluded. The results are examined by their main categories and the findings of them in the following chapters. Along with other analysis in each category, meaningful quotes from the reviews are provided by presenting the number of review and the device, for example in this way: "R3 of Echo Dot mentions that the device provides value by its hands-free features".

4.1 Six paradoxes

Six paradoxes, the main research interest, formed the biggest main category with 492 reviews out of 811 accepted ones. To be eligible for this category, the review must only handle issues surrounding the paradoxes. In addition to this category, there were also reviews with multiple labels over categorization boundaries – this rises the total of reviews mentioning at least one paradox to 614 out of 811 accepted ones. Device-wise, the difference was rather big: Echo Dot had 363 reviews out of 445 accepted ones mentioning at least one paradox, whereas HomePod had 251 out of 366 accepted ones. However, this can be explained as the number of accepted reviews between devices has a rather big difference, as Echo Dot had 445 accepted ones and HomePod had 366 ones.

In order to find out the distinctive paradoxes, the reviews were also allocated with a suitable paradox – for example, if the reviewer was referring to Alexa or Siri by she/her pronouns, the review was labelled with identity paradox. The same review could have also been labelled with multiple other attributes, if suitable ones were detected. Many times one review handled multiple paradoxes, for example issues surrounding identity, filter and value paradoxes were all sometimes mentioned in one review. The distribution between paradox appearances in reviews is summarized in figure 3.

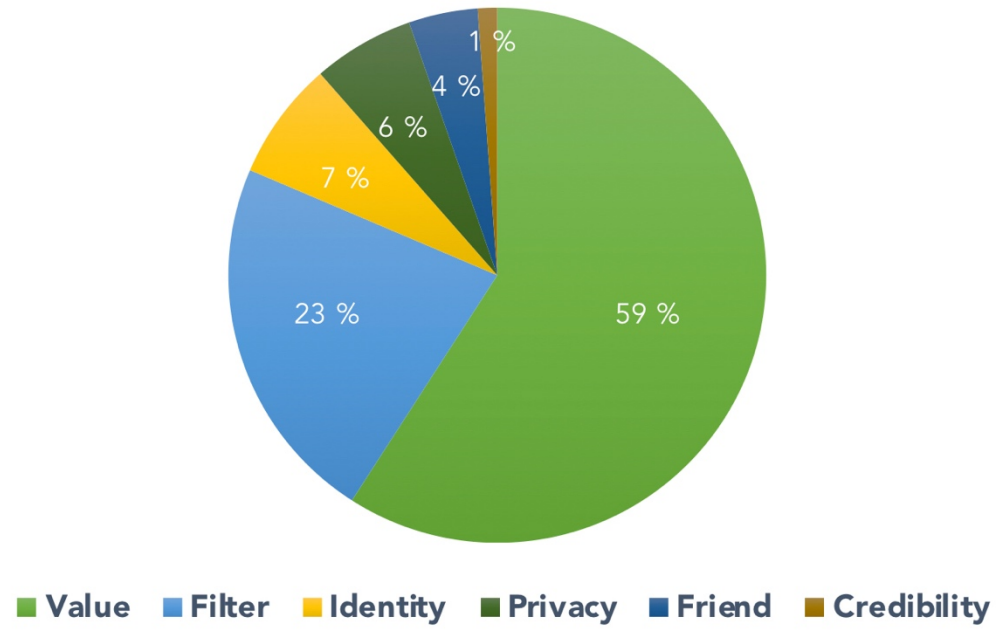

Figure 3. The distribution between paradox mentions in reviews.

Each paradox is handled in its own chapter, in which the numbers of the paradox are broken down and analysed against the Appendix A's template questions, while also providing meaningful quotes from the reviews to support the paradoxes' findings. The paradoxes are presented in descending order: value paradox is the first one as it had the most mentions in reviews with 455 reviews, whereas rather surprisingly, credibility had the lowest amount with only nine mentions. Further discussion on the relationship between numbers behind the paradoxes, the research's connection to prior research and implications to the future is presented in the fifth chapter.

4.1.1 Value

Value paradox was the most common paradox with mention in 455 out of total of 811 accepted reviews. Supporting questions for categorization and evaluation included issues surrounding the most important qualities of the voice-based user interfaces, satisfaction on the interfaces and overall usefulness (Kinsella & Mutchler, 2018; Lopatovska et al., 2018; Pradhan et al., 2018). Overall, the big number of mentions with value paradox came from the use cases – the reviewers many times listed the devices' use cases which bring value to reviewers' lives. This most likely is part of the product reviews' nature, as their job is to provide information to other customers for justifying the purchase, for example by telling where the product shines and where is the room for improvement.

Qualities-wise as an important aspect, the reviewers mentioned the ubiquity by the voice – for example, R3 of Echo Dot mentioned how "Alexa is very reliable and it is great with the hands free features", which was continued by R10 of Echo Dot: "asking questions while studying without having to reach for my phone feels ~oh so good~". Sometimes the reviews even mentioned how the voice as an interactive medium bring "the future" from sci-fi movies – this mainly emphasizing that the ubiquity convenience of the voice is the main quality of the interfaces. The reviews listed a lot of use cases where the convenience can be considered as important quality of the device – for example, in the R18 of HomePod, the reviewer mentioned how the device is "convenient for checking weather, time, telling jokes and so on".

The convenience aspect was emphasized in R31 of HomePod, where the reviewer told how controlling smart home devices is easier: "I'm a lazy lazy man, and I hate light switches. I just want to tell someone to turn off the lights for me, and since my wife is just as lazy as I am, Siri is the best." R34 of Echo Dot mentioned how "life is so much easier now, especially with our hands full holding baby" and "we don't know how we ever lived without it". Thus, use cases which were previously done in another way, were now providing convenience as they can be done by just using voice without physical effort. Another review also mentioned how the device has made life easier when having kids around, as controlling devices hands-free is possible (R286 of Echo Dot).

Accessibility is one distinct value-providing aspect with these interfaces, due to once again, the ubiquity: R45 of Echo Dot mentioned how the reviewer's father is "physically disabled … and a product like this makes his daily living much easier with just a voice command that enables him to live more comfortably". One reviewer bought an Echo Dot for their 76-year-old mother, who uses the device to control lights. The physical switch of the lights was located in a difficult place, which made the reviewer concerned if she tripped in the dark when reaching for it. Now, as the lights can be controlled without reaching for the switch, the reviewer has "much less worry now" (R70 of Echo Dot).

One ubiquity-related use case is enabled when the user had multiple devices in the house: for example, one reviewer stated how "it's also nice to be able to use the drop in feature so I don't have to yell across the house" (R135 of Echo Dot). This was continued in R405 of Echo Dot: the reviewer was injured and going up and down the stairs in the house is difficult. The drop in feature provided accessibility: when the reviewer's son is upstairs playing online games and ignores the injured parent's texts or calls, the parent has managed to contact him by getting his attention with Echo Dot's lights which indicate that there is a message coming via the drop in feature. This way, the reviewer did not have to go through the unpleasant stairs just to say something.

Accessibility can show in rather surprisingly ways: in R299 of Echo Dot, the reviewer stated how their husband had a stroke a while ago, which led to a normal post-stroke situation where he constantly asked the same questions again and again. This was embarrassing for him, but this embarrassment was reduced after acquiring the Echo Dot, as he can ask the assistant the same questions over and over again without feeling constantly bothering.

In addition to the accessibility's value when considering the ubiquity of controlling other devices, one reviewer also emphasized the feature in which the user can create a list of contacts in cases of emergency. R367 of Echo Dot mentioned how this feature was the main reason for acquiring the device, in case the reviewer's 85-year-old mother falls: "primary motivation was that, if she falls, she can call out to Alexa to make a phone call to me". In R445 of Echo Dot, the same ubiquitous call feature was emphasized as the device was acquired for elderly parents.

In the review about mother's possible fumbling in the dark by a concerned person, the reviewer stated also that later the mother's use cases extended to asking news, weather and the time (R70 of Echo Dot) – this was also one common thing with the reviews, as the users constantly discovered new utilities. Many times, especially with Echo Dot due to its low price, the reviewer had received the device as a gift or as a freebie along another product. Initially, the reviewer had not seen any value with the device, but with some time, they have started to use it and discovered constantly new value-providing features. For example, in R7 of Echo Dot, the reviewer stated how "I didn't think I'd get much use of it, but I was surprised". This is continued with a list of use cases the reviewer has found out to be useful.

Sometimes the reviews showed that initially the users might have seen the value of the device, but they rarely had the initial knowledge how to unleash it, like in R22 of Echo Dot: "I had absolutely no idea what the Echo Dot was, honestly!" In these cases, the learning process also showed in the reviews as the reviewers started listing their initial use cases followed with ones found later. Also, surprise was one factor when learning about the capabilities: "I don't believe that I didn't own one before having first interaction only 45-days ago" (R184 of Echo Dot). Some reviewers did feel that the assistant can be a novelty, but overall the constant learning along with discovering the capabilities made the reviewers more committed to the devices.

The differences between Echo Dot and HomePod, or particularly Alexa and Siri, showed in the reviews. For example, one reviewer mentioned how HomePod "isn't like Google or Alexa … Homepod's focus isn't on being a massive online search engine" (R6 of HomePod). Overall, the reviews many times stated that the HomePod and Siri's strengths are in music and as a speaker, whereas Echo Dot and Alexa's strengths are in offering more broad range of functionalities and general use cases. For example, R51 of HomePod stated that "Alexa integrates with more things than Homekit [Apple's platform for smart home appliances]". Another review of HomePod condensed this common opinion into one sentence: "the audio quality is bar none, but without better features, your money would be better spent elsewhere" (R35 of HomePod).

Overall, the value paradox's presence in the reviews was constant with the ubiquitous aspects. The reviewers seemed to benefit from voice's features, as it does not physically restrict the user unlike many times the alternative ways to perform a task do. This convenience showed in the reviews as satisfaction and higher ratings for the devices. To summarize the results of value paradox in the product reviews, R351 of Echo Dot stated it clearly: "I just use my voice".

4.1.2 Filter

Filter paradox was represented in 172 reviews, being the second largest paradox among the reviews. Supportive questions were by Kiseleva et al. (2016) and Lopatovska et al. (2018), concerning the assistant's abilities to understand the user's commands and the successful completion of the tasks. Unlike in value paradox, which could be approached from various angles, there was one distinct topic among the reviews with filter paradox: how well do Alexa and Siri catch and understand the user's commands? The reviews many times handled this issue by stating if the device is able to hear the user from distance, or by handling the recognition capabilities at the same time when music is playing from the device.

Getting the message successfully to assistants sometimes worked, but sometimes not. For example, R18 of Echo Dot stated how "when there is a song or audio already playing … the wake word response is very low probability unless you are really loud". On the other hand, R23 of Echo Dot stated similar experiences of lacking wake word recognition, but sometimes the reviewer also felt that the device is too sensitive: "these issues seem sporadic, sometimes non-existent and then giving trouble the next". One reviewer stated that the room must be silent before using the HomePod: "have to ask everyone in the room not to talk and mute the TV before I attempt to speak to it, because it can't seem to separate out that type of background noise" (R38 of HomePod).

The enunciation and accent cues were also mentioned in the reviews. R74 of Echo Dot stated how "you have to say Alexa, and pause, and then ask something", making the reviewer wish for more intuitive query inputs. R208 of Echo Dot stated accent problems: "with my accent, it doesn't understand me very well". This was continued in R289 of Echo Dot: "she needs to learn to interpret my Oklahoma twang". The influence of speech impediments was mentioned in R383 of Echo Dot: "It has much more trouble understanding my husband than me. He has a slight lisp."

As sometimes the reviewers felt that the filtering is lacking when catching commands, thus hindering the advantages of ubiquitous assistants, there was also another side of the issue. In R24 of HomePod, the reviewer gave the device only two stars out of five, due to sensitive microphones: "if I "Hey Siri" a question to my iPhone on the table in front of me, HomePod upstairs will answer". The reviewer hoped for a future update where the user could toggle the sensitivity of HomePod. One reviewer used very colorful language to describe Siri's capabilities to catch anything: "Siri can hear you from across the ocean with 20 elephants trumpeting right beside it" (R237 of HomePod).

Interesting point with the sensitive microphones came from one reviewer who stated how making phone calls through the speaker can be troublesome. "Callers told me I sounded far away", although the user was only six inches from the device (R37 of HomePod). One Echo Dot user stated how it is not possible to say "Alexa" at all in the house unless the user was actually talking to the device, as the activation always goes on when saying it (R191 of Echo Dot).

With filtering, it is also important to examine whether the assistants are capable of understanding the user when the command itself is successfully caught by the device. R40 of HomePod stated how it is needed to study the commands for controlling music and still sometimes Siri was "hit or miss". On the other hand, R93 of HomePod told how "Siri is also a bit more forgiving when it comes to saying the exact name of lights", when comparing the needed accuracy of queries of Siri against Alexa's. R118 of Echo Dot stated that "the more precise you are with your queries the better results you will receive". One user had problems with Siri, "but, that's probably because I'm not good at phrasing my requests" (R120 of HomePod).

4.1.3 Identity

Identity's share of paradoxes was 55 reviews, placing it to the third place of all paradoxes. Guzman (2019) has studied the assistant's location and existence before, and the supportive evaluation questions were about assistant's form and location. In the examined reviews, the focus was on pronoun usage of the users. Many times, the users used feminine pronouns she/her to describe the assistants, as the voice of the assistant is female. Thus, the users placed anthropomorphistic cues into non-human objects. Location of the assistant was a tricky question to be evaluated through product reviews, due to the nature of this type of documents. It seems that the users are not interested in going deep with the assistant's nature and location, as the reviews mostly focus on the value-providing use cases of the devices.

The reviewers many times used the feminine pronouns to describe the assistant and the capabilities. R26 of Echo Dot told how "she can meet a lot of needs that you didn't even know you had. … I just can't live without Alexa anymore". One reviewer humanized the assistant even further: "it seems that this girl is hard of hearing" (R58 of Echo Dot). Another reviewer referred to Alexa as "a great hostess at parties", thus giving the assistant a job title with feminine cue (R77 of Echo Dot).

Some reviews did go deep enough to examine the borderlines of technology, for example in R133 of Echo Dot: "this seems more like a personal assistant than technology". Another reviewer also stated how "it is like having a personal assistant right on my nightstand" (R206 of Echo Dot), continued by R247 of Echo Dot: "a little weird, like having an extra person in the room". One reviewer referred to Siri as "an inanimate object" (R253 of HomePod).

Problem with Echo Dot was the naming identity: R50 of Echo Dot stated how many smart devices come with Alexa's functionalities, without possibility to change the name for each device. This reviewer particularly was frustrated when the Alexa-enabled thermostat and Echo device activated when saying the wake word. R306 of Echo Dot continued this by telling that "you'll want to avoid having friends named Alexa or talking about her on the phone" – this implicating that the accidental triggering is constant and hopes for name changing are present.

4.1.4 Privacy

Privacy's appearance in reviews was almost as high than identity's, with 47 reviews. The reviewers had concerns on "always hearing" microphones of the devices, but also many compared companies' overall privacy strategies with each other. The supportive questions came from Lau et al. (2018) and Luger and Sellen (2016) – in their research, the topics surrounding privacy were about users' privacy concerns on data handling and if the privacy is or can be compromised with these interfaces.

Overall, many reviewers thought that Apple has advantage when it came to privacy – R6 of HomePod told how "Homepod doesn't feel creepy. Siri doesn't feel like she is always listening to you, and that makes me happy." R14 of HomePod continued with this, as the reviewer stated how trust for Apple and Siri is much higher than Amazon or Google. Common with these reviews was that many times the reasoning for these beliefs were not disclosed any further – usually no solid evidence for these privacy concerns was not offered. One reviewer mentioned how the HomePod was chosen due to "way they handle data requests" (R36 of HomePod), and this was continued in R88 of HomePod where the reviewer stated how "Apple's listening and encryption protocols are significantly better than Amazon's or Google's".

Reviewers were very aware of the microphones and talked rather often about them but did not see them as a dealbreaker. R298 of Echo Dot stated “invasion of privacy” as a con, but the reviewer had also “weighed the pros and cons; at this time this works for me”. R32 of Echo Dot warned potential customers how “smart speaker may also be compromising your privacy … fortunately, these mics may be manually shut off easily by a switch”. Going beyond the switch solution, one reviewer stated how they unplug the device a lot, due to concerns about a device which is ”constantly listening to everything” (R277 of Echo Dot). In R421 of Echo Dot, the reviewer has tackled the “always hearing” microphones by disabling them completely – the reviewer unmuted the devices only when needed. Of course, this takes away some of the ubiquity aspects of the assistants, as the mute setting must be always tweaked before using the assistant.

Comparison between technology giants was constant, with Apple having the advantage in beliefs. R55 of HomePod believed that Amazon uses the user’s data to target advertise, or that Google sells the data to advertisers. R91, R195 and R198 of HomePod all shared the same worries – the choice was HomePod due to concerns about competitors’ intents. One interesting remark which also somehow fits into identity paradox was about data breaches: “I know that an actual person is most likely not listening on the other end, but with all the data breaches going on, can you really be sure?” (R233 of Echo Dot).

Sometimes the concern was not about the service provider itself, but the authentication of the voice itself. R31 of Echo Dot had concerns about using Echo Dot with security systems, as “you can create a voice profile, that does not prevent anyone else from accessing your house”. On the other hand, Alexa was sometimes used to boost security: “it can listen for glass break or loud noises … and alert me as part of our home security system” (R67 of Echo Dot). Sometimes also the devices might randomly turn on with assistant stating how “she didn’t catch that when we weren’t talking either at all to her” (R228 of Echo Dot).

4.1.5 Friend

Friend paradox included several issues from prior research – for example, the emotional connection caused by the assistants (Moussawi, 2018) and whether the conversational communication with the assistant is natural (Portet et al., 2013). This paradox had 32 reviews, and the users many times used rather emotional language when describing the device and assistants. Also, companionship and even friendship were mentioned a few times, providing a peek into more humane side of technology.

The friend paradox showed many times with rather emotional adjectives – for example, R30 of Echo Dot stated the device being “an adorable Echo”. R31 of Echo Dot also showed emotion towards the device: “many times Alexa has … told me a daily joke to keep my smile going”. First mentioned in the value chapter, the review which mentions physically disabled father also tells how “his smile was the brightest this past Christmas”, as the device had really big value to helping be more independent (R45 of Echo Dot). Also, the review where the reviewer was worrying about 85-year-old mother falling in the dark was categorized in friend paradox: “she tells me she doesn’t feel as lonely now that she can tell Alexa to play music while she is working in the kitchen” (R367 of Echo Dot).

The companionship can be one of the core features of these assistants – in R96 of Echo Dot, the reviewer was planning to upgrade the 2nd generation Echo Dots to 3rd generation ones, and after this donate the old ones to “local convalescent home so the infirmed can use them for entertainment and communicate with friends and family and

not feel so lonely". R116 of Echo Dot stated how the reviewer has been successful with family trivia arguments thanks to Alexa, which is continued by how "I'm really not sure how I could get along without her constant help and friendship".

Sometimes the identity paradox went very close to the friend one – feminine pronouns softened the reviews to be more friendly. One HomePod review mentioned how "Siri actually listens to us and can understand us", and the continued to refer Siri with pronoun she (R143 of HomePod). With Echo Dot, the entertainment features were emphasized, making the assistant seem as a companionship and more emotional: "she sings us songs, reads us poetry, tells us jokes" (R290 of Echo Dot). Overall, the friend paradox was formed mostly from keeping company, emotional adjectives and pronoun usage – when the device is extended to be beyond a utility, friend paradox has its place.

4.1.6 Credibility

Rather surprisingly, credibility had only nine hits among the reviews. The supportive questions came from Lau et al. (2018) and Luger and Sellen (2016), where the trust on assistants and situations where the user prefers another medium over voice were evaluated. Credibility's low presence most likely is explained with the nature of reviews – the users do not see the assistant per se as untrustworthy. On the other hand, the privacy aspects in larger picture were much more common topics to be discussed, for example from device microphones' and companies' point of view.

R28 of Echo Dot stated how there were problems with simple web searches and explaining the user the color light indicators of the device, making the user most likely to turn into another way to get answers. Another issue was the accuracy and levels of detail between assistants – R49 of Echo Dot felt that Google offers much more detailed information about weather than Alexa, thus making Google's alternative much more reliable.

Trust on the device was built slowly but nonetheless firmly. R79 of Echo Dot mentioned how Alexa "sometimes does odd things … I suppose it is still "learning"". Another reviewer also felt this way: initially, there was hesitation but after renting a house where Echo Dots were used, the reviewer acquired own devices (R81 of Echo Dot). The initial trust issues were tackled after starting to use the device – but the privacy issues in bigger picture were more present in the reviews.

4.2 Other common themes and use cases

In addition to the six paradoxes, common themes among reviews were also categorized. These included for example ease of setup and using the device, usage rates, feature discoverability, motivation for using the devices and the future (Kinsella & Mutchler, 2018; Lau et al., 2018; Lopatovska et al., 2018; Luger & Sellen, 2016; Pradhan et al., 2018). In value paradox, several use cases for the assistants were discussed – in this section, the most common use cases are examined more deeply and the numbers behind them. 153 reviews mentioned at least one other common theme – the use cases were spanning across the categories, thus the number of them was formed by calculating all the use cases in reviews despite what category the review represented.

The use cases were the constant theme in the reviews – most of the reviews mentioned at least one use case which provided value for the reviewer. Device specific differences were present: although listening to music was clearly the number one use case with both devices, the number of music use cases was higher with HomePod. Many times the

reviews mentioned that HomePod has a clear advantage and reasoning for purchase, if the main use case is to listen music with good sound quality. In other cases, other solutions were recommended, for example if the customer is looking for smart assistant which handles broadly other functions. The use cases and their appearances are listed in descending order in the table 6 below.

Table 6. The distribution between use cases.

	Amazon Echo Dot (445 accepted reviews)	**Apple HomePod (366 accepted reviews)**	**Total (811 accepted reviews)**
Listening to music	212, ~47,6% of the device's accepted reviews	266, ~72,7% of the device's accepted reviews	478, ~58,9% of the total of accepted reviews
Controlling other devices	152, ~34,2% of the device's accepted reviews	79, ~21,6% of the device's accepted reviews	231, ~28,5% of the total of accepted reviews
Asking general questions	113, ~25,4% of the device's accepted reviews	31, ~8,5% of the device's accepted reviews	144, ~17,8% of the total of accepted reviews
Alarms, notifications, reminders, timers and similar utilities	105, ~23,6% of the device's accepted reviews	32, ~8,7% of the device's accepted reviews	137, ~16,9% of the total of accepted reviews
Weather inquiries	73, ~16,4% of the device's accepted reviews	20, ~5,5% of the device's accepted reviews	93, ~11,5% of the total of accepted reviews
Entertainment (for example games and jokes)	49, ~11% of the device's accepted reviews	7, ~1,9% of the device's accepted reviews	56, ~6,9% of the total of accepted reviews
Calls and messages	18, ~4% of the device's accepted reviews	14, ~3,8% of the device's accepted reviews	32, ~3,9% of the total of accepted reviews

The second most popular use case, controlling other devices, also showed some differences between the devices. Amazon's solution was considered more open, whereas Apple's HomeKit solution was considered to be more restrictive and closed. This continued with the same pattern of recommendation as with music – if the user's main needs are outside of music, other solutions than HomePod were recommended. Asking questions and utilities such as alarms and timers were considered useful, especially in the kitchen. This is most likely due to the ubiquity, as cooking and other activities in kitchen require hands and eyes for the main tasks – it is easier to tell the assistant to set a timer without any physical effort while the focus is on the more prioritized task.

Many times the reviews also mentioned weather inquiries, being especially useful with morning routines: the users can dress and prepare themselves accordingly at the same time with the ease of just asking the day's weather. Entertainment was very kid-centric – the reviews mentioned several times how the reviewers' kids loved to play games and ask the assistant trivia questions and jokes. A bit surprising, calls and messages were in clear minority and the users seemed not to care about this feature – but as stated before, at least with HomePod there were some hearing issues for calls.

Ease of setup and use were also common themes in reviews – generally, the reviewers felt that the devices were quite straightforward to setup and use. For example, HomePod's setup goes in a following way, as stated by R42 of HomePod: "you have an iPhone or iPad, you put them close, confirm you want to set up HomePod, and that's it!" After this, HomePod configures everything through the device which was used to initialize the setup. One particular problem rose from this – is the HomePod's setup as straightforward if the user does not own an iPhone or an iPad? R321 of HomePod states how the reviewer "had to purchase a newer iPad for setup", but no further details of alternative setup processes are disclosed.

When examining the reasoning with HomePod, the ecosystem loyalty was seen in the reviews. For example, in R34 of HomePod, the reviewer mentioned how "my home is filled with Apple products so naturally I gravitated towards the HomePod". The Apple ecosystem and its reasoning for acquiring a HomePod is mentioned in several reviews – it was more likely to be recommended if the user has already invested in Apple's ecosystem of products and services.

From the Echo Dot's perspective, perception on ease of setup varied. R205 of Echo Dot stated that the reviewer "experienced some difficulty in setting up dot initially", continued by R217 of Echo Dot mentioning how "if you aren't a tech person, you might struggle with set up on this device. You have to download an Alexa app on your phone and then setup the device". On the other hand, R226 of Echo Dot stated that "this device is perfect for the not so technical person" – forming a contradiction against the R217 of Echo Dot's statement. Another review mentioned how "Dot is an easy set-up and can be accomplished in a matter of minutes" (R253 of Echo Dot).

Usage rates were a little mystery in the reviews: supportive questions did mention these issues, but very rarely did the reviews tell about them. However, common topic in the reviews was the awareness of possibilities and the changes in use over time. Many times the reviewers did not know initially the possibilities, or if they did, only a few of them. Over time, the reviewers discovered more features and were thrilled about them, making them more invested in the ecosystem. For example, someone might buy a virtual assistant device for alarm and question use, but as time goes on, the user might extend the usage into handling other devices as new discoveries of features happen. On the other hand, some reviewers felt the device as a novelty – they bought it and used it for a while, but then forgot about it.

Motivation-wise, there was a lot of variance, as the devices do have a lot of features for different needs and customers. Some reviewers wanted convenience, some wanted a smart home with easily controllable devices over voice, some just wanted a speaker with great sound quality. As stated before, recommendations of HomePod were many times based on the music features. This exhaustive statement summarizes the common recommendation type by one reviewer:

> The big distinction to keep in mind here is that if you're looking for a smart assistant/speaker whose primary focus is voice commands, the Echo is the one to have. If you're looking for something to actually play music through with great sound where the voice commands are the interface, the Homepod is night and day better. (R124 of HomePod)

The motivation for acquiring the device was supported by the numbers of use cases in table 6. As Echo Dot was generally considered to be more general all-in-one device, it had more even distribution between use cases. Although music was the number one use case with both devices, the Echo Dot's other use cases formed around half of the use cases, whereas with HomePod the number of other use cases was only around 30

percent. As the HomePod was being lauded for music features, the features had much higher use case frequency of around 73 percent.

Overall, the constant themes among the devices were the ease of setup, different motivations for use and constant new discoveries of features. For example, R197 of Echo Dot stated how the reviewer is "still learning exactly what the Echo Dot is capable of, but so far I like it a lot". Value was perceived in various places with different users, and this was seen in the reviews, as the reviewers argued against or for their devices and compared them to alternative devices.

4.3 Improvement points and the future

Third category handled improvement points and wishes for the future. The reviewers had rather modest hopes – the supportive question by Lau et al. (2018) about dream voice-based user interface with no technical limitations thus did not go as far as the question can seem, as the product reviews showed rather realistic hopes for near future. Other supportive questions tackled the issues of user's feelings and experiences with the devices and what would be the factors to rise the usage rates of the devices (Kinsella & Mutchler, 2018; Moussawi, 2018). Improvements were categorized in 180 reviews, with a skew towards HomePod's development.

As HomePod was constantly considered to be behind the competitors, the improvement points for this particular device were mentioned several times. R101 of HomePod stated how there are some smart features by Siri, and more are coming with software updates: "some might want to wait for software updates to add more functionality". R103 of HomePod was also aware of the lacking smartness compared to others, but was sure that Apple will "eventually catch up or surpass the others". R104 of HomePod offered a concrete example between the differences in same task with Alexa and Siri: "if I ask Siri if it is going to rain, all I get is a "yes" or a "no". If I ask Alexa the same thing, I get a percentage chance of rain and expected amounts." R22 of HomePod continued by comparing the smart device controlling between the two assistants: "Siri can't do anything in my house with the smart devices I have set up, whereas Alexa is all over it."

Overall, the experience with Siri on HomePod was considered rather lacking: R131 and R135 of HomePod stated how "Siri is useless" and "dumb compared to Alexa". The latter however stated that there is potential – when Siri and HomeKit are improved, the reviewer will switch from Alexa. The reviewer in R141 of HomePod was a long time Apple user who also saw improvement points for Siri, but hoped that the updates will make Siri better – in fact, the reviewer stated that incremental upgrades have already happened. One reviewer felt that Siri was lacking, but privacy concerns about other companies made the reviewer stay with HomePod (R175 of HomePod).

As Echo Dot's reviews were about the 3rd generation of the device, some reviews compared the current version against older ones. There have been improvements – for example, R148 of Echo Dot stated how "Alexa's voice controls have gotten better since the first generation". R153 of Echo Dot agreed with the improvements over previous generations, but still hoped for further learning and recognition ability improvements. R267 of Echo Dot mentioned how the most recent generation has become quicker with faster response times. R46 of Echo Dot however told opposite experiences, although the step backwards is about the Alexa itself, not the device: "the Alexa services and the app's progress seems to have come to a complete halt and there are many quirks here and there regarding understanding, hearing, learning, usability and functioning in general".

One reviewer did test out both platforms and felt that Alexa offered “greater functionality”, and the reviewer eventually returned the HomePod. The HomePod had advantage with music features, but “it’s too confusing having two AI’s for different purposes” (R249 of HomePod). The reviews many times agreed with the target segments of devices: R290 and R292 both stated how HomePod excels in areas where Echo devices lack, the music features.

Overall, the hopes for future were small but important. More accurate filtering, results and functionalities are needed, but the reviewers many times had faith that the future is getting better with the devices. For example, R300 of HomePod mentioned about Siri how “she’s getting smarter with time”. The devices can have excellent sound quality or other great technicalities, but they are not the only issue when considering the bigger picture of voice-based user interfaces. At the core, the main hopes are for the assistants – the technicalities itself do not provide the best experience with the devices.

5. Discussion

The conducted research and prior research are connected together in this chapter. Prior research heavily relied on different methods on research and data gathering, for example from qualitative interviews to similar method as with this research's document-based product review analysis. The research questions are also examined here – there are three distinct ones with rather different points of view, ranging from today's modern voice-based user interfaces, their use cases, and to see what the future of these interfaces might hold. First and foremost, presenting the research question about the today's modern versions of voice-based user interfaces, reflecting them against the six paradoxes, and forming implications what these devices have or not to offer right now:

RQ1. What are today's voice-based user interfaces in modern web and mobile environments, reflected against the six paradoxes?

Two devices and interfaces were examined – Apple's Siri, which is implemented in HomePod smart speaker, and Amazon's Alexa, implemented in Echo Dot smart speaker. Initial starting level of the assistant varies between the two players in the markets – Apple has been around with Siri since 2011 with smartphone implementation, but only recently expanded to smart speaker markets. Amazon, on the other hand, has been in the markets since 2014, with initial high emphasis to smart speakers. (Kinsella & Mutchler, 2018)

Prior research mentions a few times the overrepresentation problem of Siri – thus, smart speakers were a great research attraction as Amazon has clear market share dominance with them, by around 65 percent (Guzman, 2019; Kinsella & Mutchler, 2018; Luger & Sellen, 2016). This sets an interesting setup, where the evolution of voice-based user interfaces can be examined through different device types and whether the earlier player in the markets has an advantage. Apple continues to dominate in smartphone markets with around 44 percent market share, while having only around five percent market share in smart speakers.

The six paradoxes offered a wide perspective to the interfaces – all of them were covered in the conducted research on product reviews. Some paradoxes were more present than the others – value paradox was the clear number one among the paradoxes, with 455 reviews handling issues about it. In prior research, value's characteristics in voice-based user interfaces is seen in the ubiquity, for example in tasks which require hands or eyes to be used, such as driving, household chores, or cooking (Kinsella & Mutchler, 2018).

Jeong and Shin (2015) argue also with the ubiquity, adding that voice is an appealing medium to interact as it is perceived as natural, everyday way to communicate. In the product reviews, the value was also seen in the ubiquity. Being able to do basic tasks, such as setting a timer while having the main concentration on cooking, was seen as useful and convenient – also, the overall convenience's value was present, for example when handling lights without having to get up to switch them off from a physical switch.

Accessibility was one area where the devices shine when considering value: the interfaces can provide more independence and easier everyday life with elderly and

people with disabilities (Portet et al., 2013). Pradhan et al. (2018) find that there are several advantages, such as the ability to do tasks without visual cues and without any physical effort. Similar results were seen in the conducted research – the interfaces have made life easier for disabled persons, as many very basic everyday tasks could now be done with just using voice. The accessibility was seen in a wide range in the reviews, from restrictive physical disabilities where the person had difficulties to move in the house to children who were worried about their parents falling in the dark when reaching for the light switch. Convenience and ubiquity of basic tasks are the advantages of these interfaces, and they provided a lot of value according to the reviews, no matter what the health state of the person is.

Filter paradox in voice-based user interfaces is about restricting the working domains of interfaces, in order to reduce voice's natural ambiguity – many times computers have problems to handle queries formed in natural language as they can be interpreted in several ways (Mankoff et al., 2000). For example, Apple has restricted the domains where Siri can work – this means that Siri cannot do every single task imaginable as the ambiguity handling would be impossible ("Domains and Intents," n.d.).

In the conducted research, the reviewers did sometimes feel that the filter features were lacking – sometimes the assistant did not catch what was said, or if it did, it did not understand the query correctly. Language problems were also present, as some reviewers reported problems with their accent or speech impediments, such as lisp. Overall, the problems of voice's natural characteristics were sometimes present: the queries had to be many times rather explicit, in order to achieve the desired result. This might hinder the adoption of these interfaces, as naturality of the voice can be considered as an advantage for using these interfaces – when the devices have problem with it, the usage rates can drop.

As Best Buy is US-based electronics retailer, the reviews were also very US-centric. This makes the results very English language based – as stated before, some reviewers mentioned accent differences, but the results might be different when the reviewers represent completely different language. The devices do support a lot of languages, so one future research issue can be the differences between languages. For example, comparing English Siri to Finnish Siri can provide rather different results.

Identity and friend paradoxes handle the issues about the perception of voice and the potential of emotional value by the assistant. Guzman's (2019) research goes rather deep with the existence attributes of the assistants, by examining the assistant's form and location, whereas Luger and Sellen (2016) and Pradhan et al. (2018) observe anthropomorphistic cues in the voice which can make the voice more "humane" and thus, offer emotional connection to the user. In the conducted research, humanization of the assistant was present: for example, when talking about Siri and Alexa, the reviewers used pronouns she and her, or even referred the assistant as a girl.

The borderlines of technology were also present, as some reviewers described the assistants more like a personal assistant or an extra person in the room than a piece of technology. The reviewers seemed to transfer the anthropomorphistic cues to their language in various forms – for example, in addition to pronoun referencing, the reviewers used words such as "companionship" and "friendship". Anthropomorphism is constantly seen in prior research and the conducted research continued the same way, as the users seemed to become even emotional and attached to the devices. These cues are making technology, which is many times seen as cold and alienating, more accessible to even larger audience.

Credibility paradox tackles the issues of assistant's trustworthiness – Nass and Lee (2000) places a personality to a machine, to match it with the user's one. Although the user was reminded that the voice is not human, the users still were convinced with it if it reflected the same personality traits of the users (Nass & Lee, 2000). In conducted research, credibility unfortunately had a very low presentation. Mainly credibility was seen in the initial setup of the device – there were hesitations, but after starting to use the device, the potential was seen. Credibility also seemed to blend very easily into other paradoxes, such as privacy or identity. For example, the assistants or their voices themselves were not concerning the reviewers, but the larger stakeholders such as the companies with their data handling policies were more concerning.

Privacy paradox handles the issues of user's privacy with the voice-based user interfaces. As the devices have access to users' homes, there are several aspects to be taken into account. For example, data handling policies, potential misuses and the responsibilities are important aspects of privacy (Oinas-Kukkonen & Oinas-Kukkonen, 2013). Lau et al. (2018) examine these aspects more profoundly especially with these interfaces – the topic is rather sensitive, as some users feel that their privacy is violated with the devices, whereas some users accept this and use the devices. Providing an access for these service providers in homes is a dealbreaker to some users, as they feel that more data than necessary is collected, for whatever reasons the companies have (Lau et al., 2018).

Some reviews in the conducted research shared the same concerns mentioned in prior research. Constantly listening devices with no warrants where the data is used were alarming to some users – although some users acknowledged these issues, they have still decided to deal with them and use the devices due to advantages they bring. Sometimes the users were worried if the data was collected to use in advertising, or if the privacy is compromised if a data breach happens. Blue et al. (2018) note that most of the modern voice-based user interfaces do not have any authentication features, making them vulnerable in security-related situations. This concern showed a few times in the reviews, making it a possible future improvement point as the devices develop further.

Overall, the paradoxes offered a wide perspective for the interfaces. Two popular device types were detected – smartphones' assistants do have their own mobile-based context, whereas smart speakers are more locked into one place. The two popular interfaces, Siri and Alexa, were examined to see the current modern situation. Although Siri might have an advantage in smartphone section, Alexa seemed to triumph in smart speakers. This is most likely due to different focus points – Apple is late in the smart speaker game as Amazon has achieved a clear leadership in implementing smart speakers into homes. A lot has been achieved, whether the focus is on smartphones or smart speakers: the assistants understand a lot, manage to perform various tasks and sometimes offer value when other mediums simply just cannot do.

Voice's value and use cases vary a lot – the assistants have a plethora of features. This makes the review analysis interesting, as the users' preferences vary a lot. Use cases were calculated from the reviews, and there was a lot of different situations where the assistant showed their capabilities. The second research question is solely about the use cases – offering a look where the highest potential is and implications how to develop the assistants even further:

RQ2. What are the most common use cases with modern voice-based user interfaces?

Lopatovska et al. (2018) examine Amazon Alexa's use cases among the users – mentions of checking weather, fact searching and news are mentioned. Weather inquiries and music features are among the popular ones (Lopatovska et al., 2018). Apple's site for Siri gives examples of music controlling and everyday tasks, such as alarms, traffic and reminder information. When going beyond the device-specific commands, the use cases can reach external devices, such as controlling the lights. Both Apple and Amazon emphasize the rather similar use cases, with no drastic differences. ("All things Alexa," n.d.; "Siri," n.d.)

Kinsella and Mutchler (2018) discuss how the use cases of the assistants are heavily skewed towards activities where hands, eyes or both are occupied. These activities include driving, relaxing at home, doing household chores, cooking and walking (Kinsella & Mutchler, 2018). In the conducted research, the results naturally did not cover every aforementioned category – for example, as smart speakers are home-based, static devices, driving was not one of the examples where the potential of smart speaker assistants was unleashed. However, the ubiquity and the possibility for hands-free communication with the device seemed to be the essential features behind the use cases.

As the examined devices were speakers, the number one use case was, not surprisingly, music-related. This could be categorized to relaxing at home – just say the command, and music starts playing. The differences between devices did show especially in here – Amazon's Echo Dot was considered to be inferior to Apple's HomePod when it comes to music, but HomePod was clearly lacking in other features. Other popular use cases of the reviewers matched rather well to the showcased ones in Apple's and Amazon's websites – the reviewers controlled other devices, used alarms, timers and other similar utilities, asked questions and used the devices as entertainment.

As controlling other devices was the second most popular use cases, it seems like that the users wish to have an ecosystem of smart products to be controlled by voice. Lights, house alarms, and other smart devices with support for voice controlling were popular among reviewers. General questions and the time-related utilities were considered especially useful in the kitchen – many times the device was placed there, as the reviewer could then ask questions and set timers hands-free while cooking or doing other tasks in the kitchen.

Weather inquiries were also popular – the distinct use case for this was among the morning routines. The reviewers mentioned how the assistant was helpful when preparing for the upcoming day, just by asking the weather. This way, the user can focus on getting ready for the day without being fully focused on the assistant itself. Once again, ubiquity was an essential element in this use case. Overall, the motivation for using the device came from various places – as stated before, the devices provide a lot of features, and users many times use only a few of them. This also led to a situation that the users were constantly learning something new – common theme in reviews was that the reviewers discovered new features over time and sometimes adopted them into their routines. This sets an interesting setup for the future's development: sometimes the users feel the devices useless, but it can be because the user has not found out the right, value-providing features yet.

The most popular use cases fit well into Kinsella and Mutchler's (2018) categories – also the companies' websites reflect very well the results. Overall, the value is emphasized in the ubiquity and convenience aspects. Constant remark in the reviews was the comparison between devices and assistants – many times the reviewers felt that another option would be better, and the recommendations were based on the reviewer's own needs and experiences. The differences were easily detected, as Alexa was constantly seen as more all-in-one solution, whereas HomePod was seen almost solely

as a music device. As the devices develop, especially the recent HomePod, it will be interesting to see whether HomePod manages to extend its territory from music to more like a general smart home hub, an area where Amazon clearly excels right now.

Kinsella and Mutchler (2018) examine profoundly the current situation of the voice-based user interfaces, without forgetting the future. The era of modern assistants is divided into two phases – first one was the introduction and evolving of the assistants. During this phase, the assistants spread across different device types, such as smartphones and smart speakers. The first phase was all about getting the users familiar with the modern possibilities of voice, whereas the second, ongoing phase is about persuasion of the voice and extending the capabilities even further. (Kinsella & Mutchler, 2018) Third research question is about the improvement points and the future of voice-based user interfaces – as the assistants have been here for a while and many people now know what they are about, the focus is in the future's improvement points and challenges:

RQ3. What are the success factors of voice-based user interfaces and how these interfaces could be improved in order to achieve wider consumer adoption?

Munster and Thompson (2018) tested the capabilities of the assistants in both smartphones and smart speakers. The understanding of the assistants is accurate, but getting the correct answer can still be difficult. Evolving of the assistants is however constant: in both tests, the evolving technology and the numbers behind the improvements show that the assistants are constantly evolving, allowing the user to perform an increasing number of tasks with just using voice. (Munster & Thompson, 2018; Munster & Thompson, 2018)

In the conducted research, the future improvement wishes were rather modest. The devices are by no means completely ready and perfect – there are a lot of issues to be fixed in order to achieve even more widespread adoption. The gap between Amazon and Apple is rather big, especially in smart speakers. Although Apple has been around with Siri longer than Amazon with Alexa, it is considered to be more lacking. The reviewers did however have hope – many times they believed that future software updates make Siri more intelligent, and maybe someday surpass Alexa.

One clear current success factor is the ubiquity. This could be improved in a lot of ways – the reviewers hoped for more accurate filtering, results and functionalities. For example, if a user has acquired a device for convenience reasons, the convenience factor degrades if the queries constantly fail and the user has to rely on other mediums. If the user constantly encounters invaluable answers from the assistants, the overall experience suffers a lot. This was many times seen in the reviews – for example, Siri can handle a certain set of questions, but wishes for more intelligent assistant were constantly mentioned. The constantly mentioned wishes for better Siri does not mean that Alexa is perfect compared to Siri – Echo devices also gathered a lot of wishes, as there simply are no devices yet which understand the user always and deliver perfectly accurate and right answers. Improvement points from reviews were small, but extremely important.

Overall implications for theory and managerial aspects are well examined in the research. The six paradoxes and their connection to voice-based user interfaces implicate that the ubiquity is the value-providing aspect – meaning that the reviewers simply wanted to ease their life with small things. Devices with microphones were concerning to many reviewers – the reviews showed that beliefs are very stubborn and very company-specific. Rather than the assistant itself, the bigger picture of company-wide privacy policies had a big role in customer's decisions. Thus, if the company's

privacy policies were stained by whatever reasons, the reviewers did not seem to care how well the assistant otherwise does the job. This is one managerial implication for the big players in the markets – value and privacy of the devices are essential, if wider consumer adoption is desired.

Another managerial implication is about the device's value properties. The reviewers constantly did not initially see the value, but also many times the users gave a chance for the device. This many times led to situations where the reviewer was pleasantly surprised after using the device for a while – initially there was no value as it was hidden, but the discovered features made the user deeply engaged to the system, integrating the system to everyday tasks. In the reviews, this was many times seen when the reviewers started to expand the ecosystem, such as with other smart appliances to be controlled via assistant. In order to get more users to use voice-based user interfaces, the value properties of the assistants should be clearer by the companies.

Accessibility is one big research topic for voice-based user interfaces. Like Pradhan et al.'s (2018) research, the conducted research showed that the voice has great potential with people with disabilities. The ubiquity of voice removes many physical restrictions of many everyday tasks, and also the anthropomorphistic voice can work as companionship, or even as a friendship. Many times with the assistants, technology was not seen as a cold machine – the anthropomorphistic properties of the assistants made them easily approachable.

The early problems with voice as an interactive medium were with the natural cues of human communication. These do still exist, but in much smaller scale than initially, almost 30 years ago. The assistants understand a lot – but there also is room for improvement. The issue is not easy, as there will always be different accents, speaking styles, speech impediments and other issues which affect the output of the speech. Machine learning and artificial intelligence possibilities can offer some solutions to the problem – as the assistants get smarter, hopes for always-accurate recognition are there.

6. Conclusion

The conducted research answered well to the research questions – in summary, the voice-based user interfaces have improved a lot since their emergence, but there is still a lot to be done. Natural language issues, company-wide privacy policies and making the value visible instantly to the users are the main improvements points. Since the late 1980s and the early 1990s, voice as an interactive medium has developed a lot: initial vocabulary and complex natural language issues have been greatly reduced. Now, almost 30 years later, the one-sided talking to the devices has evolved into maintaining human-machine conversations, due to great achievements in technologies such as artificial intelligence and voice recognition capabilities.

As Kinsella and Mutchler (2018) state, the phase two of these interfaces is ongoing. This includes expanding the capabilities of voice to cover multiple devices, creating an ecosystem which makes the user even more invested in the medium (Kinsella & Mutchler, 2018). Persuasiveness of the voice is also one characteristic of this phase – one future research topic could indeed include the persuasiveness of the voice as a medium. As the users many times see personality traits and other human-like characteristics in the assistants, exploration on technology's more humane side and the impact on persuasiveness can possibly significantly improve the assistants even further.

Another future research topic for voice-based user interfaces is accessibility. From prior to conducted research, value-providing aspects and use cases when considering people with disabilities were constant. As the phase two of these interfaces is about creating larger entities to embrace voice as a medium, these entities could provide much more fulfilling life and independency for persons who now have daily struggles with basic, often taken for granted tasks, such as switching the lights on or off. However, it is important to remember that automatizing everything is not the right solution – the users might then feel even bored if the tasks which they feel as activating are taken away.

As a limitation for this research, the data gathering method would have been better if accompanied with another method. For example, interviewing users of Echo Dot and HomePod would have given more detailed answers – also some supportive questions for evaluating the reviews were not very suitable, as rarely did the reviews handle some issues from the questions. The reviews showed a lot of variance: users discussed wide range of topics, making the data rather scattered. Also, the product review gathering was done manually, as Best Buy forbade using any automated tools – if all the reviews could have been automatically mined by relevant keywords, the results might have been more accurate. The conducted research examined only smart speakers – the smartphones were left out, although the use cases are different between the device types. By this, the overrepresentation problem with Siri was avoided.

To summarize, the current state and future of voice-based user interfaces is rather bright. There are improvement points to be considered, but the users do not wish for the impossible. The six paradoxes disclosed a lot about the current state, but also about where the road should go on. In the future, voice as an interactive medium surely has its place, and it could expand to areas which the users even are not yet aware of.

References

All things Alexa. (n.d.). Retrieved December 5, 2018, from https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011

Blue, L., Vargas, L., & Traynor, P. (2018). Hello, Is It Me You're Looking For? Differentiating Between Human and Electronic Speakers for Voice Interface Security. *WiSec '18 Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks,* 123-133.

Brennan, S. E. (1991). Conversation with and through Computers. *User Modeling and User-Adapted Interaction, 1*(1), 67-86.

Corbett, E., & Weber, A. (2016). What Can I say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. *MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services,* 72-82.

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human-computer dialogue. *International Journal of Human Computer Studies, 83,* 27-42.

Customer Ratings & Reviews. (n.d.). Retrieved February 26, 2019, from https://www.bestbuy.com/site/reviews/apple-homepod-space-gray/5902410

Customer Ratings & Reviews. (n.d.). Retrieved February 26, 2019, from https://www.bestbuy.com/site/reviews/amazon-echo-dot-3rd-gen-charcoal/6287974

Delogu, C., Paoloni, A., & Pocci, P. (1991). New Directions in the Evaluation of Voice Input/Output Systems. *IEEE Journal on Selected Areas in Communications, 9*(4), 566-573.

Domains and Intents. (n.d.). Retrieved December 12, 2018, from https://developer.apple.com/design/human-interface-guidelines/sirikit/overview/domains-and-intents/

Dubiel, M. (2018). Towards Human-Like Conversational Search Systems. *CHIIR '18 Proceedings of the 2018 Conference on Human Information Interaction & Retrieval,* 348-350.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review, 114*(4), 864-886.

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. *2017 IEEE Intelligent Vehicles Symposium (IV),* 365-372.

Google Assistant. (n.d.). Retrieved December 5, 2018, from https://assistant.google.com/

Guzman, A. L. (2019). Voice in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior, 90,* 343-350.

Jeong, J., & Shin, D.-H. (2015). It's not What It Speaks, but It's How It Speaks: A Study into Smartphone Voice-User Interfaces (VUI). In Kurosu, M. (Ed.), *Human Computer Interaction – Interaction Technologies* (pp. 284-291). Los Angeles, CA: Springer.

Kiesel, J., Bahrami, A., Stein, B., Anand, A., & Hagen, M. (2018). Toward Voice Query Clarification. *SIGIR '18 The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval,* 1257-1260.

Kinsella, B., & Mutchler, A. (2018). Voice Assistant Consumer Adoption Report November 2018. Retrieved December 5, 2018, from https://voicebot.ai/voice-assistant-consumer-adoption-report-2018/

Kiseleva, J., Williams, K., Jiang, J., Awadallah, A. H., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding User Satisfaction with Intelligent Assistants. *CHIIR '16 Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval,* 121-130.

Lanamäki, A. (2017). *Qualitative Research Lecture 1: Overview* [PDF document].

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction – CSCW, 2.*

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., … Martinez, A. (2018). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science,* 1-14.

Luger, E., & Sellen, A. (2016). "Like Having a Really bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 5286 -5297.

Mankoff, J., Hudson, S. E., & Abowd, G. D. (2000). Providing Integrated Toolkit-Level Support for Ambiguity in Recognition-Based Interfaces. *CHI '00 Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* 368 -375.

Mogalakwe, M. (2006). The Use of Documentary Research Methods in Social Research. *African Sociological Review, 10*(1), 221-230.

Moussawi, S. (2018). User Experiences with Personal Intelligent Agents: A Sensory, Physical, Functional and Cognitive Affordances View. *SIGMIS-CPR'18 Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research,* 86-92.

Mukhopadhyay, D., Shirvanian, M., & Saxena, N. (2015). All Your Voices are Belong to Us: Stealing Voices to Fool Humans and Machines. In Pernul, G., Ryan, P. Y. A., & Weippl, E. (Eds.), *Computer Security – ESORICS 2015* (pp. 599 -621). Switzerland: Springer.

Munster, G., & Thompson, W. (2018). Annual Digital Assistant IQ Test – Siri, Google Assistant, Alexa, Cortana. Retrieved December 5, 2018, from https://loupventures.com/annual-digital-assistant-iq-test-siri-google-assistant-alexa-cortana/

Munster, G., & Thompson, W. (2018). Annual Smart Speaker IQ Test. Retrieved December 5, 2018, from https://loupventures.com/annual-smart-speaker-iq-test/

Myers, M. D. (2013). *Qualitative Research in Business & Management.* Smy, K. (Ed.). SAGE Publications Ltd.

Nass, C., & Lee, K. M. (2000). Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction. *CHI '00 Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* 329-336.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues, 56*(1), 81-103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors. *CHI '94 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 72-78.

Oinas-Kukkonen, H., & Oinas-Kukkonen, H. (2013). *Humanizing the Web – Change and Social Innovation.* Palgrave Macmillan UK.

Portet, F., Vacher, M., Golanski, C., Roux, C., & Meillon, B. (2013). Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Personal and Ubiquitous Computing, 17(*1), 127-144.

Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.*

Rogoff, R. (2001). Voice Activated GUI – The Next User Interface. *Proceedings IEEE International Professional Communication Conference,* 117-120.

Shechtman, N., & Horowitz, L. M. (2003). Media Inequality in Conversation: How People Behave Differently When Interacting with Computers and People. *CHI '03 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 281-288.

Siri. (n.d.). Retrieved December 5, 2018, from https://www.apple.com/siri/

Terms and Conditions. (2019). Retrieved February 26, 2019, from https://www.bestbuy.com/site/help-topics/terms-and-conditions/pcmcat204400050067.c?id=pcmcat204400050067

Use Siri on all your Apple devices. (2018). Retrieved December 5, 2018, from https://support.apple.com/en-us/HT204389

Voice Assistant Timeline. (2018). Retrieved December 19, 2018, from https://voicebot.ai/voice-assistant-history-timeline/

Appendix A. Evaluation base for gathered product review data

Construction	Questions from prior research
1. Six paradoxes and their connection to voice-based user interfaces	
1.1 Privacy	Does the user have any privacy concerns with the interfaces, for example with how service providers handle the user's data? (Lau et al., 2018) Does the user think privacy is/can be compromised with the interfaces - if yes, how? (Portet et al., 2013)
1.2 Identity	The assistant's voice, who does the user think it is? (Guzman, 2019) Where is the voice of the assistant located? (Guzman, 2019)
1.3 Credibility	Are there situations where the user prefers another medium over voice? (Lau et al., 2018) Does the user trust these assistants? (Luger & Sellen, 2016)
1.4 Friend	Has the user ever felt emotional connection caused by the assistant(s)? (Moussawi, 2018) Does the user think that the conversational interaction is natural with the assistant(s)? (Portet et al., 2013)
1.5 Filter	Has the user been able to complete all intended tasks? (Kiseleva et al., 2016) Have the interfaces understood all the user's commands? (Lopatovska et al., 2018)
1.6 Value	What are the most important qualities of the interfaces? (Kinsella & Mutchler, 2018) How satisfied has the user been with the interfaces? (Lopatovska et al., 2018) Does the user see the interfaces useful? (Pradhan et al., 2018)
2. Other common themes and use cases of voice-based user interfaces	
2.1 Other common themes surrounding voice-based user interfaces: using them, reasoning for use, usage over time	How often does the user use the interfaces? (Kinsella & Mutchler, 2018; Lopatovska et al., 2018; Luger & Sellen, 2016) How does the user foresee the future use? (Lau et al., 2018) How long has the user been using the interfaces? (Lau et al., 2018) Has the use of these interfaces changed over time, from beginning of the use to now - if yes, how? (Lau et al., 2018; Luger & Sellen, 2016; Pradhan et al., 2018) What is the motivation to use the interfaces? (Luger & Sellen, 2016) Was the user aware of the possibilities of the interface before acquiring one? (Pradhan et al., 2018)

2.2 Use cases	What are the user's most frequent use cases of the interfaces? (Kinsella & Mutchler, 2018; Luger & Sellen, 2016; Pradhan et al., 2018) What has been the most memorable interaction with the interfaces? (Lopatovska et al., 2018)
3. Improvement points and the future of voice-based user interfaces	
3.1 Improvement points in order to achieve wider audience	What would cause the user to use the interfaces more (i.e. what is needed to rise the use)? (Kinsella & Mutchler, 2018) How does the user feel about the interfaces right now? (Moussawi, 2018)
3.2 Dream voice-based user interface	With no limits in technicalities whatsoever, what would the user's dream voice-based user interface be like? (Lau et al., 2018)