To Simplify or Not to Simplify, a Longitudinal Case Study of Smart Speakers Appropriation Over Time

Master Thesis for Interaction Design 2019

Aalborg University
DENMARK

Title:
To Simplify or Not to Simplify, a Longitudinal Case Study of Smart Speakers Appropriation Over Time

Project Period:
1st of February to 3rd of June

Project Group:
hci1019f19

Group Members:
Ayan Nuur
Nanna Møller Calundan
Nichlas Elleby Jensen

Semester:
10th semester

Supervisor:
Anders Rysholt Bruun

Number of Pages:
51

Abstact:
Studies have shown that not all the provided functions in the smart speaker are being used equally. For this reason, we wanted to find out why some functions are used less frequently as others. Two variations of smart speakers were constructed; a simplified smart speaker which had limited functionality and a complex smart speaker with full functionality. A longitudinal case study was conducted with the purpose of gaining an in-depth understanding of the long-term use of the smart speaker and to examine the difference in how users appropriate the speaker based on its complexity. The framework, temporality of experience by Karapanos et al., was used as an analytical lens to see how users appropriate the smart speaker in the three phases based on the complexity of the smart speaker. It was found that the complexity did not have an impact on the appropriation of the smart speaker in the orientation phase, since all the households had been through this phase. It was furthermore found that the households with the complex smart speaker are more likely to develop routines and reach the incorporation phase. The findings suggest that the complexity had an impact on the identification phase. Although none of the households had developed an emotional attachment to the smart speakers, it was found that the households with the complex speakers were more likely to develop this attachment. The simplified smart speaker did not reach the last phase due to the available functions being insufficient for the households.

Summary

A few conducted studies in Human-Computer Interaction (HCI) have been devoted to gaining an understanding of the use of smart speakers over time and how they are integrated into people's home[2, 34, 35]. This longitudinal aspect is important when examining the development of use over time, since it allows to see the change that can occur.

It is apparent in commercials for smart speakers that they are depicted to be able to change people's daily life by assisting people with their routines[43, 47]. It was, however, found that these smart speakers have functionalities that are not used as frequently as others [2, 25, 34, 35]. These findings made us wonder, why these functions are not used as much. What would happen if a smart speaker were simplified in its functionality, would it be possible to compare it to a smart speaker like the ones in the commercials? This led to the following research question:

How are smart speakers appropriated in accordance to the framework, temporality of experience's three phases over time, depending on complexity?

We used the framework, temporality of experience, proposed by Karapanos et al. to get an understanding of the use of the smart speakers over time. The framework will additionally be used as reference to identify the current stage of our participant in their process of receiving the smart speaker and as an indication for when they enter the different phases.

To answer our research question, we needed smart speakers with simplified and complex functionality. Therefore, we constructed our own set of smart speakers for the study, instead of purchasing off-the-shelf products, to make it possible to simplify some of the speakers to only have a few functions. Even though it was not necessary for us to construct the smart speaker with complex functionality, in terms of complexity because it is the same as an off-the-self speaker, we decided to do it nonetheless to eliminate the variables of differences in appearance and sound quality and thereby make it easier to compare.

After constructing the smart speakers, we needed participants for our study where we recruited four households, one from a Facebook group, two with paper flyers and the last one was friends of ours. We recruited the households based on experience and if they were interested in trying a smart speaker. The four households all participated throughout the whole study. We designed this study to be a four-week longitudinal case study, to give the households time to appropriate the smart speaker. We chose to conduct a case study because we wanted to get an in-depth understanding of how our participating households appropriate the smart speaker over time. To observe the use, we used log data of the commands that the participants issued as well as interviews with each household, that were used to get a deeper understanding of their experiences with the speaker.

When we delivered the smart speaker to our participants, we aslo conducted the first interview including an informal conversional technology tour of their home, where we got to know our participants as well as their home. The gathered data was analyzed after every interview with the participants, since the mid- way interview and concluding interview were based on the data that was collected so far. Affinity diagrams were used to get an overview of the gathered data, in which we used grounded theory and let the categories and concepts emerge from the data. Triangulation of data was used to strengthen our findings, in which we used the log data from each household and used the participants' own statements from the interviews to get a holistic understanding of their use of the smart speaker.

We found that all four household had been through an orientation phase, where each household explored their respective smart speaker, to determine which functions were useful for them. It was furthermore found that there was no notable difference in the length of the orientation phase depending on the complexity of the smart speaker. However, when looking at the average of commands used per day during this phase, the simplified smart speaker had an average on 9,15 commands per day whereas the complex smart speaker had an average of 37,5 commands per day.

The findings indicate that the households with the complex smart speaker are more likely to develop routines and reach the incorporation phase. We saw that household 1 had reached the incorporation phase which was evident in both the log data and the statements from the interviews. Household 3 on the other hand was difficult to determine whether they had reached this phase or not. The households with the simplified smart speaker had not reached this phase, meaning they did not have a regular use of the available functions. It was established during the interviews that household 2 and 4 could not see a usefulness of the available functions, where they stated that this was due to not having the relevant functions for their needs.

We did not assess any of the households to have developed an emotional attachment to the smart speaker. Although the households with the complex smart speaker, household 1 and 3, showed some tendencies of potentially developing this attachment. This was prevalent in household 1 as they had developed routines with the smart speaker where they used it to structure their day. This made it slightly difficult to return to their daily life before the smart speaker. Whereas household 3 showed the smart speaker to their guests, which correlates with Karapanos et al. findings, since they state that the emotional attachment entails developing personal as well as social relations with a device [14]. The remaining households that had received the simplified smart speaker had not developed an emotional attachment to the device. This indicates that the households with the complex smart speaker are more likely to develop this emotional attachment, whereas the simplified speaker is insufficient to develop an attachment.

Our findings suggest that the complexity of the smart speaker has an impact on the appropriation of the speaker in the three phases in temporality of experience. The findings furthermore suggest that the households who received the complex smart speaker were more likely to experience all three phases of the framework. While the available functions in the simplified smart speaker stopped the remaining households from experiencing all three phases.

Table of contents

Introduction

Currently there is a wide range of ubiquitous products available to consumers. For instance, the smartphone is a great example, since it has been appreciated by consumers ever since it became available on the market. A trend in Human-Computer Interaction (HCI), is to understand these products, as they become an addition in people's homes and become a part of the user's everyday life, while fulfilling their needs[14].

The latest addition to these ubiquitous products is the smart speaker, where voice interaction is the primary interface in these screenless products. The speakers are pitched on the fact that they can become an addition in the home, since they support a broad set of functions, that helps the user with a variety of different scenarios: to play music or games, help in the kitchen by utilizing timers and recipes, or accessing information about news or weather[33]. For instance, the commercial for the Google Home shows that it can be a part of your daily life and that it can help your daily routines[43]. These speakers are currently becoming a popular addition in people's homes[29]. Though in HCI, only a few studies have been devoted for understanding how these speakers are used over time and integrated into people's homes and daily routines[2, 34]. This suggest that there is a need for more studies devoted for understanding the smart speakers over time as it would indicate how these speakers are incorporated into people's household and reflect how meaningful the speaker becomes in one's life. The framework, temporality of experience, proposed by Karapanos et al., identifies three phases that users are able to experience when receiving a product, namely orientation, incorporation and identification. These phases cover the experience users go through when receiving a product, where they go from exploring its capabilities to being emotional attached to the product, where they have identified how the product can be incorporated into their daily routines and interactions[14].

Studies have shown that although the smart speakers have overly complex functionality, users are unfamiliar with the speakers capability and it is only a few of the functionality which are used frequently[2, 25, 34, 35]. Not knowing the capability have led to frustration among the users, since it is difficult to gain an overview of the available functions, limiting the interactions with the speakers[25,35]. Additionally, the less used functions may be more relevant if the speaker did not flood the user with functionality. Therefore, we found it interesting to uncover whether decreasing the complexity of the speaker, in the form of available functions, would help users experience less frustration and if the less used functions would be more frequently used.

In this report, we compare a simplified smart speaker, a speaker with fewer functions than an off-the-shelf version, to a complex smart speaker, with the same functionality as an off-the-shelf. By using the framework, temporality of experience, as an analytical lens, we want to understand how these speakers compare to each other in the three phases. To get an in depth understanding of how speakers are appropriated over time, depending on the complexity, we conducted a longitudinal case study over four weeks, where we used between-subject design, with four different household. Two of the households received the simplified speaker and the rest received the complex speaker.

This report makes two contributions in the form of empirical findings, to the field of Human-Computer Interaction (HCI). First, it provides an overview of how the complexity of a smart speaker will affect the usage and how users appropriate it. Next it provides an examination of how this complexity affects the three phases proposed by Karapanos et al. in the framework, temporality of experience. To guide this study, the following research question reflects our interest and curiosity:

How are smart speakers appropriated in accordance to the framework, temporality of experience's three phases over time, depending on complexity?

Report Structure

1 BACKGROUND
This chapter presents relevant papers from a conducted literature review on the few devoted studies in understanding the smart speaker and/or virtual assistant in Human-Computer Interaction (HCI). Additionally, it introduces the framework, temporality of experience proposed by Karapanos et al., and the methodology for conducting longitudinal studies, case studies and grounded theory.

2 PROTOTYPING
This chapter provides an overview of our process in constructing four smart speakers with two different complexities in terms of functionality. It includes the inspiration phase where we used mood-boards to gain an insight into different speakers' appearance in addition to the construction phase, where we modified several components in order to construct the speakers.

3 LONGITUDINAL CASE STUDY
In order to get an in-depth understanding of the long-term use of the smart speaker and the change which may occur in the use, we conducted a longitudinal case study. The purpose of the study was furthermore to examine the differences in how users appropriate the smart speaker depending on its complexity. In the study we delve into four households with the aim of gaining an understanding of each household and their specific uses of the smart speaker. This chapter describes the study design, a description of the procedure of conducting our study in addition to the data analysis and lastly it descripts the participating households.

4 FINDINGS
This chapter contains the findings of our study. The purpose of this chapter is to display the findings of the conducted longitudinal case study in addition to shedding light on the differences between the households' appropriation of the simplified and complex smart speaker. The chapter is structured based on the three phases in the temporality of experience.

5 DISCUSSION
This chapter contains a discussion/reflection of the different aspects involved in conducting our longitudinal case study. It emphasizes on key points such as the impact of the complexity in the smart speaker, whether knowing all the available functions in the smart speaker is a necessity in addition to advantages and disadvantages of conducting longitudinal case studies. It furthermore includes a discussion on whether developing an emotional attachment to our smart speakers is possible and lastly the participants' motivation for participating in our study.

BACKGROUND

This chapter presents relevant papers from a conducted literature review on the few devoted studies in understanding the smart speaker and/or virtual assistant in Human-Computer Interaction (HCI). Additionally, it introduces the framework, temporality of experience proposed by Karapanos et al., and the methodology for conducting longitudinal studies, case studies and grounded theory.

1.1 Literature Review of Smart Speakers and Virtual Assistants

We conducted a literature review of 37 papers regarding virtual assistants or smart speakers. These papers were grouped into three categories: understanding usages, user experience and improving. An overview of these categories, with some of the papers, are presented below.

1.1.1 Understanding Usages

The papers that described how users interact with smart speakers in their own home are grouped together in this category. Some papers give an overview of which commands are used the most when users interact with a smart speaker[2, 7, 24, 34, 3]. We want to highlight the work by Bentley et al. and Sciuto et al., they made an overview of all the commands, not only the frequently used but also the least used commands.

Category	% of Total
Music	40%
Information	17%
Automation	9%
Smalltalk	8%
Alarm	6%
Weather	6%
Video	4%
Time	4%
Lists	2%
Other	4%

Table 1.1: Used functions in the study by Bentley et al.

What can be seen in Table 1.1 is an excerpt from the study by Bentley et al. that shows how the commands were used by their participants. *Music, information and automation* were used the most on the other hand *other, lists, and time* were used the least[2].

Similarly Table 1.2 shows the distribution of commands in the study by Sciuto et al. We can see that music, smart home automation and weather is commands that their users

Category	% of Total
Music	25%
Other	20,3%
Smart Home	14,7%
Text not Available	9,7%
Weather	4,6%
Question	3,7%
Timer	3,6%
Wakeword	3,5%
Lists	2,8%
Volume Change	2,7%
Personality	2,6%
Time	2,2%
Alarm	1,5%
News	1,1%
Joke	1,0%
Connectivity	0,6%
Purchase	0,3%

Table 1.2 : Used functions in the study by Sciuto et al.

the most. Where commands regarding purchase of items, connectivity and jokes, are the commands that are used the least[34]. We can see that there is a general overlap in these findings by Bentley et al. and Sciuto et al. If we look at the most used commands, in the other papers that we found, there is the same overlap in the most used commands[7, 24, 35].

When the user needs to place their smart speaker in their home, it depends on how many smart speakers they already have. If it is the first smart speaker in the home, it tends to be placed in a room where all the members of the household have access to it, the speaker is therefore often placed in the kitchen or living room[29, 34, 35].

1.1.2 User Experience

If a user of a virtual assistant needs to have a great user experience and be satisfied, the assistant needs to fulfill the user expectation in a satisfying way[18, 25, 28]. This does however not necessarily mean that if the user experiences some usability problems, that the assistant will be seen as useless[31]. Another way that the virtual assistant can create great user experience, is if it manages to surprise the user with new and usable functions[25].

What generally contributes to bad user experiences are the lack of what can be defined as natural conversation. Porcheron et al. describe natural conversation as a human to human conversation. When we talk to each other we do not only pay attention to the phrasing, but also how it is said and what it implies[33]. Studies found that it can lead to frustrations if a virtual assistant do not understand the commands issued by the user, because of accents or word editing. Therefore, the user will then start to talk slowly and hyperarticulate to make sure they will be understood[7]. Participants in another study expected that the assistant would understand the context and remember the previous commands as a human would do[36]. All in all, virtual assistants and thereby smart speakers do not support natural human to human like conversations, but can instead be seen as single-turn, command- or task-based [4, 7, 35].

Smart speakers offer many functions to the user, but the manufactures only include a small get started brochure[41, 42, 44], but there are more functionalities than that. Participants in the studies by Luger and Sellen as well as Smith found it difficult to figure out the capabilities of smart speakers. Luger and Sellen suggest it is due to a lack of feedback from the smart speaker, that the participants are not cable of knowing all the functions[25]. Smith's participants were frustrated when they did not know the capabilities of the smart speakers and saw the newsletter from Amazon with suggestions to commands as a burden[35].

1.1.3 Improving

It is evident in the literature that the virtual assistants and smart speakers can be improved upon. In the literature there is also papers that suggest how to improve the smart speaker. According to Iba, when improving a smart speaker, it is firstly important to consider which market segment we need to target to make a successful product[13]. Some literature has suggested improvements based on a specific segment, like the work done by Kim et al. where they studied teenagers and what their needs are[16].

To improve the smart speaker, Kim et al. tried to add a new modality to the speaker, that made it possible to change the temperature of the speaker, this was done to support the understanding of weather[17]. Other ways of improving the virtual assistant powered smart speaker, is by letting the assistant continuously improve itself e.g. with the help from machine learning[5, 8, 27].

The researchers from three studies want to improve the virtual assistant to support natural conversation by making it more context aware, by letting it understand emotions or remember previous conversations[6, 12, 37]. To make it more human like, some papers also suggest incorporating anthropomorphism in the assistant[19, 23, 30, 38, 39]. Lastly in the work by Milhorat et al. they state it is important that virtual assistant fulfills the expectations that the user has to the product[26].

1.2 Temporality of Experience

Karapanos et al. conducted a five-week ethnographic study in which they examine how user's expectations and experiences with the iPhone develop and change over time. They developed a framework, temporality of experience, where they identified three phases when receiving a product, namely orientation, incorporation and identification, see Figure 1.1. The first phase, orientation, refers to the novelty phase of receiving a product, in which a user explores the product. In the second phase, incorporation, the users have decided which functions are useful to them and their daily lives, i.e. they have incorporated the product into their daily routines. In the third and final phase, identification, the user has formed a personal and emotional attachment to the product. In their study they additionally identify three forces, familiarity, functional dependency and emotional attachment, which they state shifts the users' experiences through the three aforementioned phases.
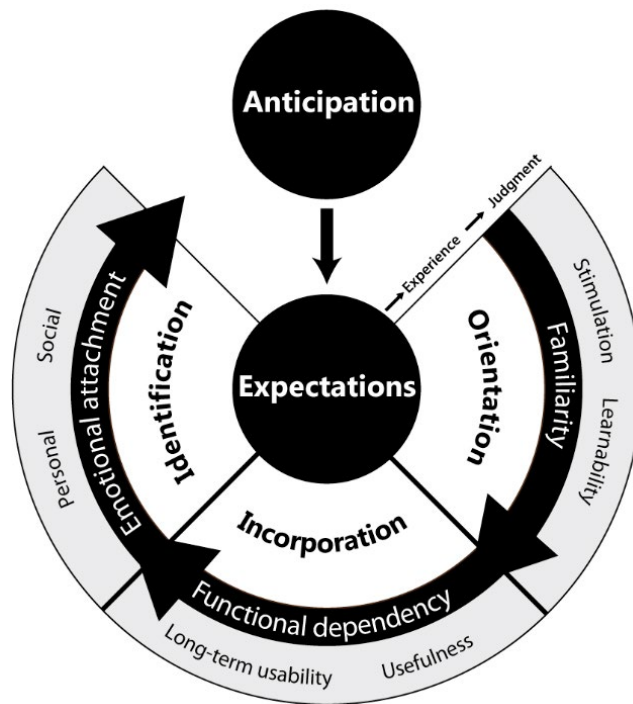
Figure 1.1: The framework of Temporality of Experience proposed by Karapanos et al

We use the temporality of experience framework as an analytical lens to see how users appropriate the smart speaker in the three phases based on the complexity of the smart speaker. Three expectations have been formed, for each phase, based on how we expect the smart speaker to be appropriated in the three phases, based on the complexity.

We expect the first phase, orientation, for the simplified speaker to be shorter, than the complex speaker. Since it only contains a few functions, the users should be able to have tried all the functions in the speaker thus decreasing the time for exploring the speaker's capabilities. For the second phase our expectation is that the simplified speaker would have a harder time being incorporated into the household, since there are less functionalities to choose from. For the final phase, we expect the complex speaker to be more likely reach this stage, since it has more functionalities that allows for users to incorporate it into its household.

1.3 Taxonomy for Longitudinal Studies

Some of the longitudinal research that has been conducted within the field of HCI have allowed for a better understanding of how ubiquitous devices are being implemented into the household of consumers and how these are used over time. These studies have focused on understanding a given device over an extended period of time and studied the changes that may occur during this period[2, 14, 34, 35].

Gerken and Reitere encourage the HCI community to conduct more longitudinal studies since the amount of studies are limited even though researchers have acknowledged the need for more longitudinal research. They have proposed a taxonomy, see Figure 1.2, that is structured around a research objective, that helps guide the rest of the study, where they have identified three objectives, within HCI studies[9].

The first objective, interest in change, revolves around trying to improve the understanding of the HCI. This may be through learning how long it takes people to learn using an application or device. The second objective, change as a confounding factor, revolves around what can be observed when a change has happened. Here it can for example be interesting to compare devices in order to understand which device is better or worse.

To the extent of our knowledge, studies that have referred to Karapanos et al. have not made use of the proposed framework but rather focused on the aspect of conducting a longitudinal study. One study uses it for determining the duration of a longitudinal study[22]. Another study discusses which methods are best suited for capturing emotional experiences over time, in which they decided on using the day reconstruction method, as they deemed this to be fitting for their case and since it worked well for Karapanos et al[20]. A study reviewed how empirical research on user experience is conducted, by looking at existing empirical studies. They criticized Karapanos et al. study for not focusing on an individual level when looking at differences in use, in addition to not giving a clear indication of how long it took for the participants to progress through the three phases[1].

Lastly, ecologic usability, is about reaching a more valid assessment of a device for instance by observing users in their environment over a longer period of time.

From the research objectives, they present seven aspects that needs to be considered when doing a longitudinal study. Here we will briefly summarize some of these.

Research design establishes the foundation of how the data for the research objective should be gathered and which method should be used. Here researchers mostly use retrospective- or prospective panel designs, where retrospective design revolves around asking participants to reconstruct several points in their past through their memories. While prospective panel design resolves around observing the same participants multiple times.
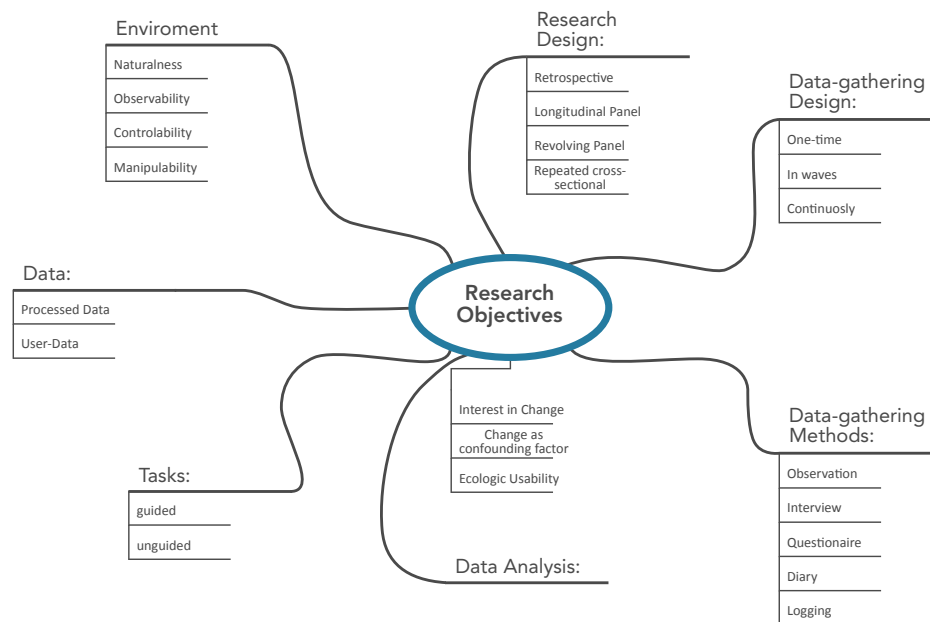
Data-gathering design contains how often data should be gathered. The data can either be gathered once, in waves or continuously over the study period.

Data-gathering methods is closely related to both the research design and the data-gathering design, since it entails which different methods researchers can use to obtain data for their research objective. For instance, observation can be used to observe users in their personal environment where interviews can be a supplement, when there are observations that requires the user's statement in order to clarify a situation. Thus, methods that correspond with the research objective and the research design should be chosen[9].

We have used the taxonomy to structure our longitudinal study, since it provides a good overview of the elements included in conducting a longitudinal study.

1.4 Case studies

Case studies in HCI are described by Lazar et al. as an in-depth analysis of a real context with a small sample size and can be used to generate hypotheses, theories or novel insides to an area with not well-established foundations of literature or research[21]. There are three points which makes case studies the preferred methods compared to others. Firstly, "How" or "Why" are the beginning of the research question for the study. Secondly, there are either little or no control over behavioral events and lastly a present-day phenomenon is the main objective of the study.

Conducting a case study consist of two parts and in order to generate a research objective or question that can guide the study, a thorough literature review should be conducted in the beginning. The first part refers to the investigation of a present-day phenomenon in the real-world context[40]. For instance, Bhattacharya and colleagues conducted a case study on a new group activity function, raiding, in the augmented reality mobile application Pokémon Go. They were interested in how the players organized these group activities when the application did not, at this point, support chat between players[3]. The second part refers to the design of the case study and the data collection. Though it may seem that the design of the case studiy is limited to a single case based on the name, it is not the case. The studies can additionally consist of multiple cases and there are

Figure 1.2: The taxonomy of Longitudinal Studies proposed by Gerken and Reiterer
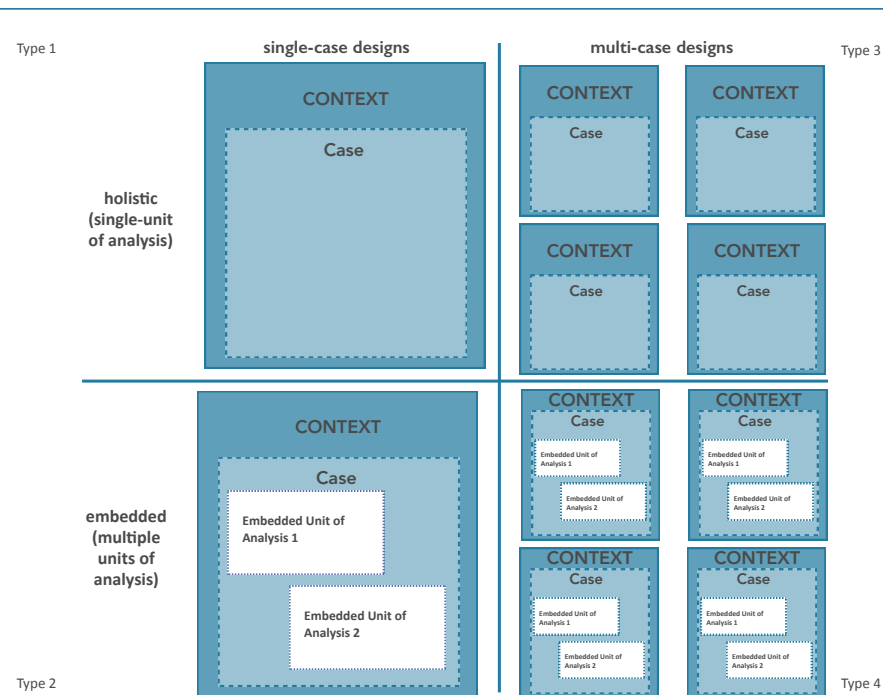
Figure 1.3: Basic case study designs proposed by Yin

both advantages and disadvantages with using either designs. For instance, Gustafsson states the advantages of the single case design is that they are less expensive and time consuming, which makes it easier for the researcher to have a deeper understanding of the subject[1]. These designs can either be holistic or embedded, where holistic refers to having one level of unit analysis and embedded means that there can be multiple sub-units of analysis, see Figure 1.3[40].

For the data collection the researcher often needs multiple data sources in order to data triangulate and reach a more valid assessment of the research question or objective. Additionally, there is an emphasis on the data being qualitative[21, 40].

According to Yin, addressing five concerns about case studies will result in conducting a proper case study. The first concern is that the study needs to be done in a rigorously and structured way, and the researcher needs to keep this throughout the whole study. The second concern is that it is important that the researcher is not doing a teaching case, which is very similar to a case study, but the difference is that in a teaching case, is used for teaching and can have materials that is altered to show a point. It is not allowed to alter material for a case study. The third concern is when questioned if it is possible to generalize based on a single case study or not. It is possible to generalize theories based on a case study, but it cannot be applied else were, like the whole population of a country. The fourth concern is that case studies can require a lot of effort and maybe also too much compared to the final result. This does not necessarily need to be the case, if the researcher considers this when designing the study. The last concern is that it can be unclear how case studies can be used to compare, when researchers compare it to other methods. But case studies are brilliant to answer why or how a researcher found a given result[40].

1.5 Grounded Theory

Grounded theory is a research method that is beneficial to use when conduction case studies, interview or using ethnography, because grounded theory can be used when there does not already exists or only limited literature to base a study on. In grounded theory, the researcher does not base their research on hypothesis gained from others previous work, but instead designs a study with the purpose of gathering data. This data is crucial to the study because it is based on this data, that the researcher will form theories. The researcher will make analysis of the data, often gathered and analyzed over multiple sessions, with an open mind and without any prejudiced or biased assumptions. After the analysis the researcher can then form theories. Grounded theory contains four stages; with an open mind the data is coded, the researcher will make concepts of the similar codes, some of the concepts will then form a category and lastly the based on the concepts and categories the researcher forms theories[21].

1.6 Summary of Chapter 1

This purpose of this chapter was to introduce the background knowledge which guided our study. The presented papers gave an insight in some of the studies that have been conducted on smart speakers and virtual assistants. Some of the papers focused on how

the smart speaker is used, where others focused on the user experience with these devices, lastly papers presented alternative ways in which the smart speaker could be improved. These papers will be used as reference to validate the findings in our study. The framework, temporality of experience has been used as an analytical lens to conduct our study, where as the taxonomy for longitudinal studies was used to structure our study. The aspects presented in case studies have been taken into consideration when designing our study. The research method grounded theory was furthermore used since we conducted a study in which there is limited research.

PROTOTYPING

This chapter provides an overview of our process in constructing four smart speakers with two different complexities in terms of functionality. It includes the inspiration phase where we used mood boards to gain an insight into different speakers' appearance in addition to the construction phase, where we modified several components in order to construct the speakers.

2.1 The Smart Speaker

For the purpose of our study, we have constructed four smart speakers. Two of these speakers are referred to as simplified, meaning that they have limited functionality compared to what an off-the-shelf smart speaker offers. The rest are referred to as complex, where they have the same functionality as an off-the-shelf smart speaker. The reason for constructing our own speakers is that the off-the-shelf speakers does not allow for modification in the form of limiting the available functionality. Additionally, we decided to construct both variations of the speaker though we did not have to construct the complex speaker since it is the same as an off-the-shelf speaker, we decided to do it nonetheless to eliminate the variables of differences in appearance and sound quality. Since these speakers needed to be dispatched to our participants and incorporated into their home, we wanted to mimic the appearance of an off-the-shelf smart speaker.

2.1.1 Inspiration and Early Process

Before we constructed our speaker, we wanted to get an insight into the appearance of different speakers such as, Bluetooth, WI-FI and smart speakers, and use that as inspiration. This was done in the interest of simulating an off-the-shelf speaker.

For this occasion, mood boards were created with pictures from a web search that included the different terms for speakers. This resulted in various types of speakers with different shapes and sizes, see Figure 2.1. From the mood boards we saw that the most common speakers are cylinder shaped. Based on the mood boards we started modeling potential shapes for the smart speaker out of paper and cardboard, see Figure 2.2. The dimensions of the speaker were dictated by the components needed for creating a smart speaker.

Figure 2.1: One of the mood board used for inspiration

Figure 2.2: Paper prototypes for determining the potential shapes

2.1.2 Investigation of Hardware Components

In order to figure out which components are needed and available for creating a smart speaker, a web search was conducted. Here we gained an insight in how others have created their own smart speakers and which components they have used. We found that the most common solutions were to use a Raspberry Pi together with a microphone and a speaker. This combination allowed us to implement Google Assistant, and through a microphone array and USB powered speakers, we were able to receive and process input, where the microphone serves as the input device, registering voice commands. Here the Google Assistant then processed these commands and generated an output, through the speakers.

The Raspberry Pi is the main component in the speaker it processes the input and output, through the Google Assistant allowing for voice interaction. The Pi we used in our study is the Raspberry Pi 3 model B +, see Figure 2.3C. This particular model was chosen since the extra computing power it provides allows for wake word detection and playing music. This means that the user is able to activate the virtual assistant by issuing a wake word, such as "Hey Google".

For our input device, we used the ReSpeaker 4 Mic array, that consist of an array with four microphones and 12 LED that together form a circular pattern, see Figure 2.3A. This means that we are able to give the user visual feedback in terms of light. The array consisting of four microphones allows for better detection in terms of distance and volume. Additionally, the ReSpeaker is an extension hat for the Raspberry Pi, meaning that it is compatible and can be connected to the Pi through the General Purpose Input Output (GPIO) pins, see Figure 2.3B.
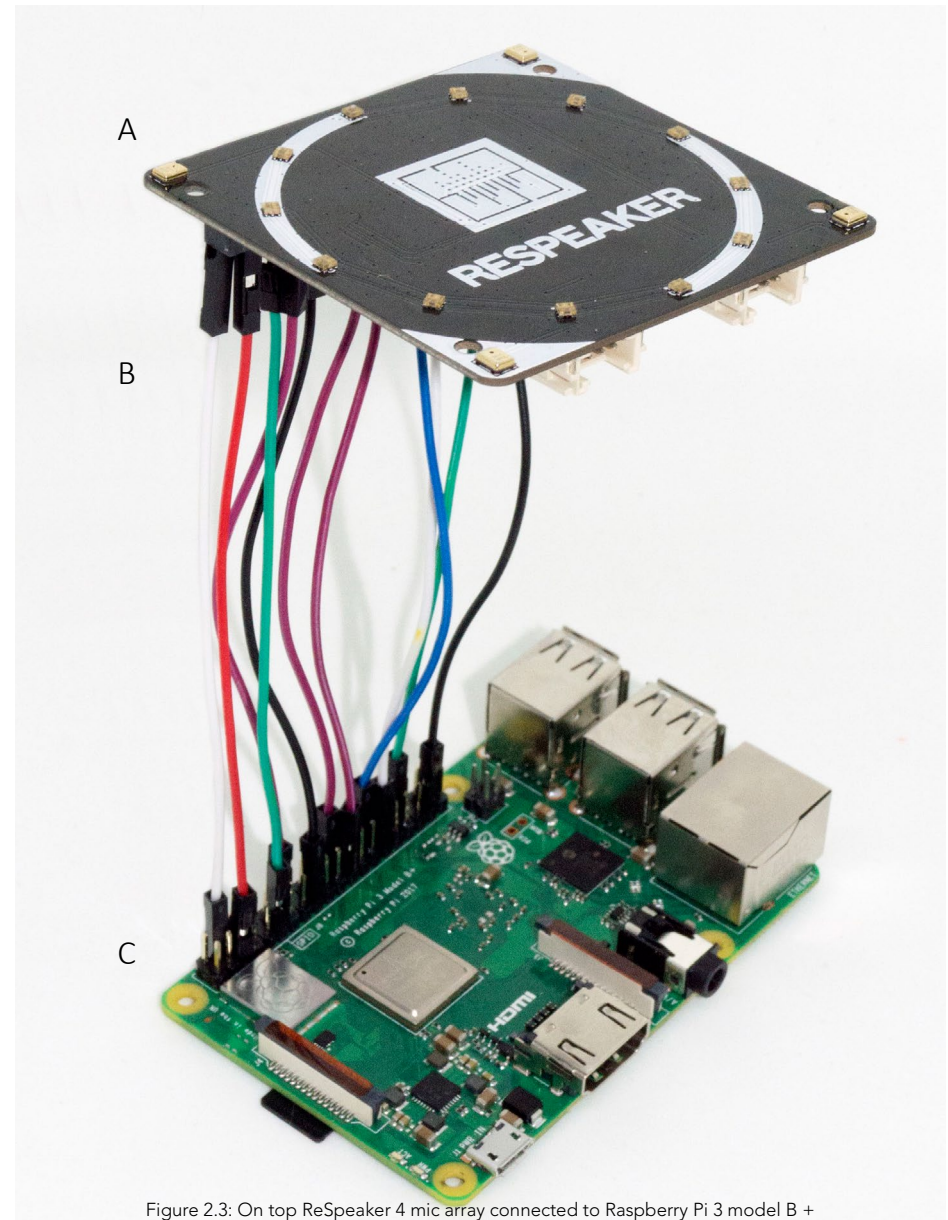
Figure 2.3: On top ReSpeaker 4 mic array connected to Raspberry Pi 3 model B +

Figure 2.4: The tested speakers, rearmost the computer speakers, middle Bluetooth speaker, font USB-speaker

For our output device, we experimented with three different speakers, since the audio quality needed to be decent enough that it could play music.

The first speaker that we tested was a USB powered speaker from Amazon, see Figure 2.4A. Here we found that when the speaker reached close to maximum volume, the sound would be distorted with a crackling sound, which we determined would not be great for listening to music or when generally using the smart speaker.

The second speaker we tested out, was a portable Bluetooth speaker from Biltema, that was powered by a battery which could be recharged through a USB port, see Figur 2.4B. Though the sound quality was better, an issue occurred when the speaker was connected to the Raspberry through the AUX- and USB port, where a high pitch sound made the speakers practically unusable in terms of sound output.

The last speakers we tested were two stereo computer speakers from Biltema which outperformed the first two speakers, see Figure 2.4C. They were connected to the Raspberry Pi, through the USB and the AUX port. In terms of audio quality, they were louder, had no issues with distorted or crackling sounds. Therefore, the decision to use these speakers as the output was made.
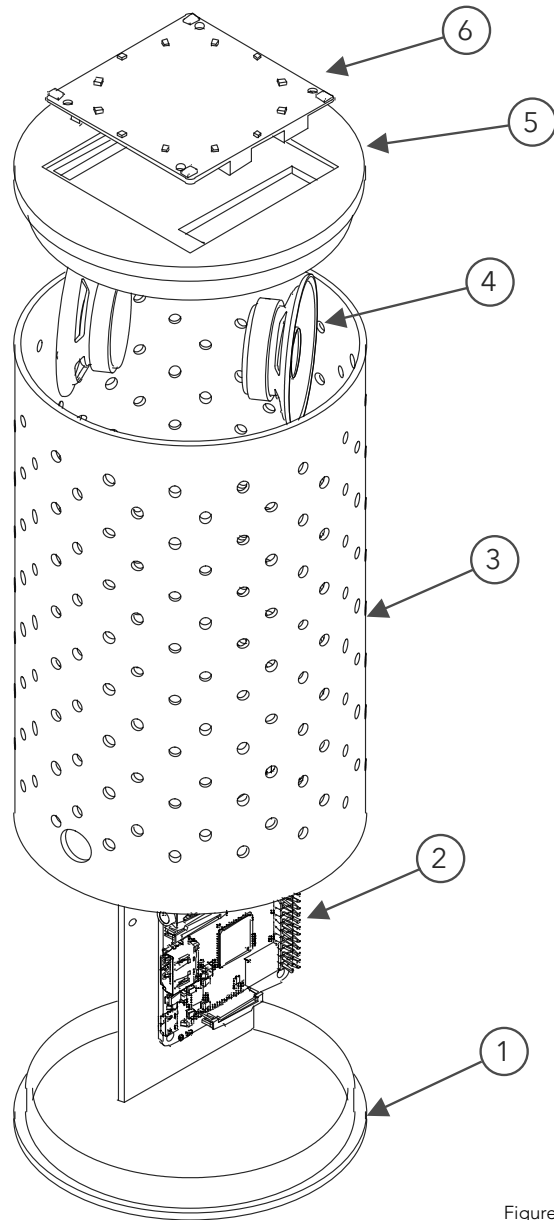
Bill of Material

#	Materials	Quantity
1	3D printed bottom part	1
2	Raspberry Pi 3 Model B +	1
3	PVC sewer pipe Ø110 mm	1
4	3,5 Ohm speakers	2
5	3D printed top part	1
6	ReSpeaker 4 mic array	1
	Bolts	4
	Nuts	4
	Wires	~
	Fabric	~

Additional Components
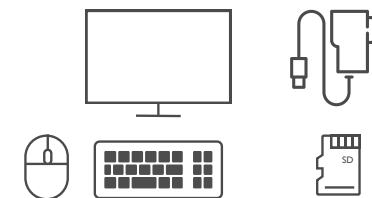
Figure 2.5: Exploded view of our smart speaker including bill of material

2.2 Physical Component Preparation and Assembly

In order for us to assemble a smart speaker, we needed to prepare additional components other than the hardware components. For a full view of the used components, and how they are assembled, see Figure 2.5, for an exploded view and a bill of material.

For the main piece of construction, we used a sturdy PVC sewer pipe Ø110 mm, since it has the cylinder shape, wide enough to fit the microphone array horizontally on top of the pipe and the height could easily be modified with a hand saw, see Figure 2.4B. In order to determine the height, we made paper prototypes with the same diameter as the PVC pipe and experimented with three different heights 14, 18 and 22 cm. We saw the height of 18 cm as the most fitted because of better balance between height and diameter was most harmonic and made it easy to see the top of the smart speaker as well. To improve the sound quality, we drilled holes in the pipe to let the sound escape the pipe and padded the inside with foam washcloth to soften the surface that the soundwaves hit. A template with a grid pattern was created and by wrapping the PVC pipe with the template, we used it as a guide to drill equally spaced holes.

We 3D modeled and printed a bottom and a top part for the sewer pipe. The bottom part was printed with a vertical plate, which allowed the Raspberry Pi to be mounted on, see Figure 2.6C. The top part was printed so that it allowed the microphone array to be embedded into 3D print, see Figure 2.6A. The array was placed here since we wanted the users to be able to receive the visual feedback that the array can provide through the 12 LED's. The speakers we had chosen to work with were encased in a plastic housing. We removed the housing in order to fit the speakers into the sewer pipe and to additionally optimize the space.

We sewed and covered our speaker with a mottled fabric, in order to achieve an appearance that would simulate an off-the-shelf speaker. We choose the heather grey fabric, since that is the most common type of fabric, we saw in our mood boards. Additionally, we ensured that it was structured and sturdy enough to holds its shape and that the LEDs were able to be seen through the fabric when powered.

A

B

C

Figure 2.6: The assembly of our smart speaker

2.3 Implementation of Google Assistant

We accomplished the implementation of the Google Assistant by following three steps. These steps are simplified explanations and require more in-depth explanation, which can be found in the online community surrounding the Raspberry Pi[32]. The first step involves installing the operating system, Raspbian, unto the Raspberry Pi. The second step requires installing the Google Assistant Software Development Kit (SDK) unto the Raspberry Pi and the last step involves programming the Google Assistant to suit your need.

2.3.1 Simplified Functionality

We limited the Google Assistants functionality by incorporating a whitelist. This list is a filter that contains words or sentences that are allowed through to the assistant. This means that when the speaker receives a command and recognizes a word that is on the list, the assistant processes and completes the command. When the speaker receives a command that does not contain a word that is on the list, it simply responds with an error message. We created and recorded our own custom error messages, by using the translation function in the Google Assistant. Though the SDK allows for custom responses, it uses another voice when outputting these, which we deemed would be confusing for the participants. Therefore, we recorded audio files where we asked the assistant to translate different sentence, so that when the user for instance asked the Google Assistant what it could do, it responded: "I can only do a few things, I can set a timer or an alarm, tell you about the weather or make you laugh with a joke". Additionally, when the participant had received two error messages, the next error message they would receive would contain some of the functions that the speaker was able to process.

Based on Bentley et al., and Scuito et al., findings of which particular functions that are less used, we chose that the simplified speaker should only be able to, tell a joke, set a timer or an alarm and check the weather, see sub-section 1.1.1. Here the words and sentences we whitelisted are based on the keywords that are used to access these different functions. Words such as, weather, timer, alarm, where added to the list though when adding the different keyword, we need to consider that certain keywords could unlock some of the functionality that we did not want the particiapnt to have access to. For Instance, a word like "bored" allowed the user to access both the ability to hear a joke, and the ability to play a game.

2.3.2 Complex Functionality

The Google Assistant SDK is limited in its functionality. For instance, music is not able to be played, which is one of the most frequently used functions of a smart speaker[2, 34]. Fortunately, there exist workarounds where we can implement these missing functions back into the assistant. For implementing the music function, we were inspired by GassistPI, a repository by Shivasiddharth where he used YouTube as the media to play music on the Pi[10]. By taking the input from the participant, we can then identify which song that is requested through a YouTube search. Afterwards, the song is downloaded unto the Raspberry Pi and played through a media player. This media player can be controlled through the assistant and when the music is done playing, the song is then discarded of.

2.4 Summary of Chapter 2

Here an overview of how we created two variants of speaker, one simplified and one complex, have been given. Three different speakers were tested due to the audio quality and there were different challenges in creating the different complexity. Here the online community surrounding the Raspberry Pi helped in solving these challenges.

LONGITUDINAL CASE STUDY

In order to get an in-depth understanding of the long-term use of the smart speaker and the change which may occur in the use, we conducted a longitudinal case study. The purpose of the study was furthermore to examine the differences in how users appropriate the smart speaker depending on its complexity. In the study we delve into four households with the aim of gaining an understanding of each household and their specific uses of the smart speaker. This chapter describes the study design, a description of the procedure of conducting our study in addition to the data analysis and lastly it descripes the participating households.

3.1 Study Design

This section entails the different aspects of conducting a longitudinal study, with an emphasis on how our study was designed. This is described with the use of the framework of temporality of experience in addition to the taxonomy for longitudinal studies, both were used to structure our study.

3.1.1 Applying the Framework of Temporality of Experience

We have used the framework, temporality of experience developed by Karapanos et al. as an analytical lens to identify the different phases of our participants in their process of adopting the smart speaker. Additionally, it serves as an indicator for when to contact our participants to interview them on their progress and their experiences with the speaker. Our aim was to interview the participants once in each phase. The framework was used to identify when the participants enter the different phases. This meant that we looked through the collected log data to determine whether the participants had entered a specific phase of the framework. This assumption was then confirmed or denied by getting them to verbalize their own experiences with the different phases in the following interviews.

3.1.2 Applying the Taxonomy for Longitudinal Studies

We conducted a four-week longitudinal case study to get an in-depth understanding of how the smart speaker is appropriated by the participants depending on its complexity. The study is structured on Gerken and Reiterer's taxonomy for longitudinal research, see Figure 3.1, where we are interested in change and change as confounding factor of the two variations of the smart speakers, meaning that we are interested in the change itself and what can be seen after the change has occurred. The research design that has been applied in this study is a longitudinal case study, which Gerken and Reiterer states is a type of prospective panel. This research design allowed us to focus on a small sample size and monitor the changes that occur with them, this way gaining an in-depth understanding of the specific participants' use of the smart speaker over time. To monitor this change, we sent out the smart speakers to the respective homes of the participants. Data has been collected by letting the smart speakers continuously gather log-data, we furthermore conducted two to three semi-structured interviews in waves, guided by the three phases described in the framework by Karapanos et. al. To make
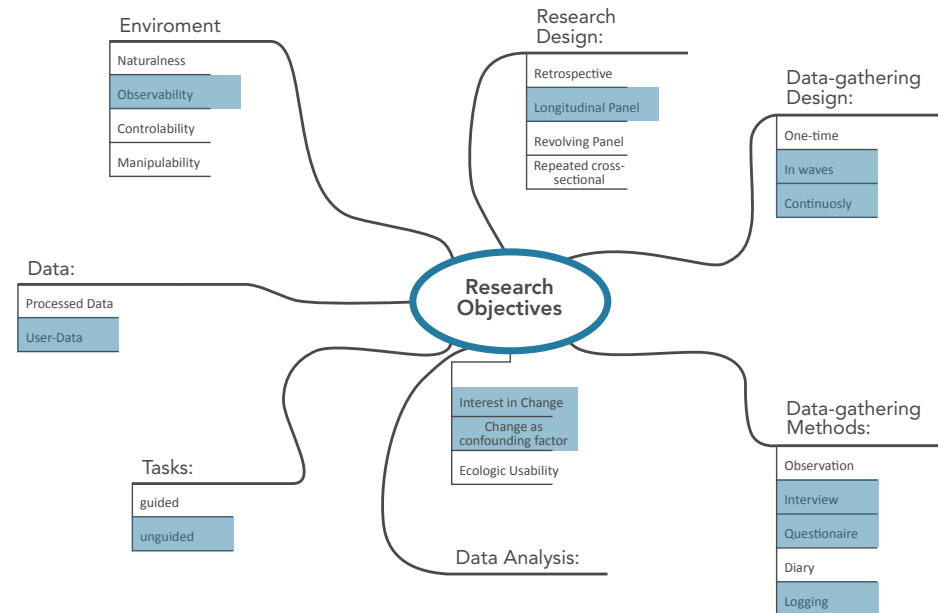

Figure 3.1: Our data gathering process

sure that our data was as natural and unguided as possible, we invited our participating households to use the smart speaker as if they bought it themselves by only giving them a brief introduction and a starting guide, like an off-the-shelf smart speaker.

Additionally, between subject-design has been used to conduct the study, since this was suitable given our time restriction as well as this study design enabling us to deploy the smart speakers for a longer time with each participant. Our hope was furthermore that deploying the speaker for four weeks will ensure that the orientation phase of receiving a new product is passed, which Karapanos et al. describes as a period of proximally one week[14]. However, we decided to conclude the study early for two of the participating households, namely household 2 and 4, the ones that received the simplified smart speaker, due to them only generating data that showed a steady use for less than a week of having the smart speaker.

3.2 Process

This section entails the procedure of conducting our longitudinal case study, with the purpose of giving an insight into our processes. It includes a description of the requitment of our participating households, the three interviews that were conducted in addition to the procedure of our data analysis.

3.2.1 Recruitment

We recruited participants for our study by making flyers and placing these around campus, in addition to posting in a Facebook group where the purpose is to recruit participants. Both the flyer and Facebook post included information about our study, just enough to arise their interest, as well as the expected time frame of the study. The limited information was given to reduce the risk of bias, this furthermore allowed us to start recruiting participants before all the minor details of the study were determined. The physical tear off had the statement "Study of smart speaker" and our contact information, enabling people to send us an e-mail if they wanted to participate. The criteria for participating in the study included not having any prior experience with a smart speaker in addition to being able to talk and understand English, since the language of the speaker is English. This resulted in a recruitment of four participants and their households.

3.2.2 Our Study

First interview: Orientation

The process of conducting the longitudinal case study was initiated with the development of two variations of a smart speaker, a simplified smart speaker and a complex smart speaker see chapter 2. To initiate the study and get the user's addresses we sent them an e-mail containing a video we had created, which presented our smart speaker and showed how to use it. We created two videos; one depicting the simplified smart speaker with all the functions available to the users[46] and another depicting the complex speaker and some of the functions which were available to them[45]. The first mentioned video was sent to the households who received the simplified smart speaker and the latter was sent to the households who received the complex smart speaker.

The purpose of the video was to create realistic user expectations by showing which functions that are available to the households so that they did not expect that it was capable of doing everything, and furthermore to show them that they were not receiving a Google Home, Apple HomePod or Amazon Echo, but rather one we had constructed. Before creating the videos, we looked at commercials for Echo[47], Home[43] and HomePod[48], to get some inspiration. We took the elements from Echo and Home, where they showcase their smart speaker in a household where the members ask it questions. The video additionally attempts to recreate the user experience of purchasing such a device, where the costumer often will do some research, by looking at reviews. We furthermore created accounts for the four households in advance, which meant that they did not have to spend time on doing this.

After having scheduled a date for the setup, we went to the participants' homes to setup the speaker. We decided to set it up in their respective homes, since it had to be connected to their WI-FI, which was done with the use of a Raspberry Pod. Deciding which household would receive which variation of the smart speaker was chosen at random. Next, the initial interview was conducted, with the purpose of gaining an understanding of the participants' expectations of the smart speaker as well as to understand what they hoped to gain from the study i.e. their motivation of participating in the study. In addition to the interview we furthermore facilitated technology tours in the different households, since the initial interview was conducted in the respective household's homes we saw it as an opportunity to get an insight in which types of technologies the different households had and to get a better understanding of the participants.

Second interview: Incorporation

Before conducting the mid-way interview, we analyzed the data that had been generated with the use of an affinity diagram. We let our four households know that we wanted to collect their data a couple days in advance and downloaded the data, this way giving them the opportunity to remove sensitive data if necessary. Based on the initializing interview and the affinity diagram we made individual interview guides for each of our households and invited our participants to a mid-way interview. The purpose of this interview was to gain an understanding of the household's use of the smart speaker and to make the participants verbalize their experiences. The questions posed in this

interview were therefore based on each household's data, which meant that each interview was specified to the individual household and their use.

To incorporate the participant's log data in the interviews, we visualized the subcategories from the affinity diagram in a pie chart and visualized the distribution of the commands in a bar chart and used these as artefacts to facilitate the conversation. The pie chart was used to get an indication of who had used which commands in the households with two members, by asking the participants to indicate which member had used the different functions. We made use of a tangible artefact of the framework, temporality of experience, during the mid-way and concluding interviews to facilitate reflection about the participants' experiences with the smart speaker. In this interview we introduced our participants to the first two phases of the framework and asked them to reflect on their experiences with the smart speaker in addition to verbalizing if they felt that they had been in the presented phases.

For two of our households, the two with the simplified smart speakers, this mid-way interview became the closing interview, since we did not receive more log data from these households and therefore asked them to bring the speakers. Since this interview was the closing interview for household 2 and household 4, we introduced them to all three phases of the framework and thanked them for their participation at the end of the interview. We scheduled when to collect the remaining households' speakers as well as when to have the closing interview. Our aim was to have a week between collecting the smart speaker and conducting the initial interview, to let our participants reflect on not having the speaker anymore and give them a change to miss or forget the smart speaker. Again, we asked our participants to look through the log data and edit commands if necessary.

Third interview: Identification
The purpose of the concluding interview was to determine whether the participants identified themselves as being in the third phase of the framework. We therefore introduced the participants to the last phase of the framework and made them reflect on if they made it to this phase or what would make them reach this phase.

The questions posed in this interview were yet again based on the individual household's data and focused on making the participants talk about their experiences. The conversation was furthermore facilitated with a pie chart and a bar chart visualizing their data. We asked the participants to predict how the bar chart developed since last time. This was done to get an insight into how the participants imagined they had used the smart speaker. We furthermore showed them what their actual bar chart looked like, as well as the pie chart from the first gathering of log data and the last gathering. The focus was to get them to verbalize their usage, which could be used to support the collected log data. By the end of the interview we thanked them for their participation and concluded the study.

3.2.3. Data Analysis
Based on Gerken and Reiterer's taxonomy to gather data we used both continuously logging of the participants use of the smart speaker as well as three in waves interviews, initializing-, mid-way- and concluding interview. Because of our participants privacy, we decided to let them look through their log data and edit out commands they did not feel comfortable sharing, we told our participants a specific date and time for when we downloaded the log data. We did this twice, around half-way in the study and at the end of the study when we got back the smart speakers. Our data collection was structured like this, see Figure 3.2.

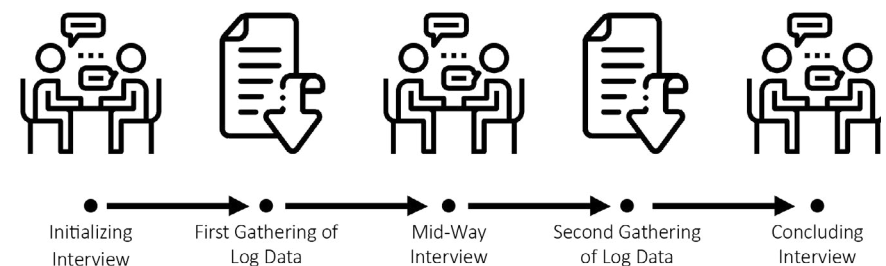| Initializing Interview | First Gathering of Log Data | Mid-Way Interview | Second Gathering of Log Data | Concluding Interview |

Figure 3.2: Structure of our longitudinal case study

Figure 3.3: Our affinity diagram process

The collected log data was analyzed by firstly, cleaning up the data by removing irrelevant data including headings, product type and location, from the spreadsheet so we were only left with, what the smart speakers had registered our participants saying, what the virtual assistant had replied as well as time and date. We did this to make the data more manageable and easier to get an overview. To analyze the data, we decided to use grounded theory and let categories emerge from the data, without having preconceived categories.

We used affinity diagrams to analyze the log data, since this method supports our grounded theory approach and provides an overview as well as a visual representation of the data. We therefore printed out all the log data and cut out each individual command, see Figure 3.3, we took one command at a time and by comparing the commands to each other, we started forming categories and sub-categories of commands. After all commands were grouped together in sub-categories, we organized each sub-category in chronological order by date and time. We did this to make it easier to get an overview of the distributions of commands over time. The commands were then glued to a piece of paper. This was done for all four households.

When we received the smart speakers from the households, we once again downloaded the log data, cleaned it up, printed it and cut it out. This resulted in an extension of the existing affinity diagram where we created new subcategories if necessary.

We had a final round of analysis, where we used triangulation of data, where all the gathered data was considered, this included all three interviews, questionnaire, tech tour, data log, including affinity diagram, pie- and bar chart. We made a combination of the pie- and bar chart, into a bar chart with stacked columns, to make it easier to see how the subcategories of commands changes over time. We furthermore used their statement from the interviews as validation for our assumptions based on their generated data.

3.3 Households

This section gives a detailed description of the four participating households. The information stated is based on a questionnaire, which the participants filled out in addition to the interviews that were conducted. During the initial interview, we facilitated a technology tour, with the purpose of getting an insight in what type of technologies the participants owned, in addition to developing an understanding of how they appropriate and use these.

3.3.1 Household 1

Household 1 received the complex smart speaker. This household consists of two people, namely the couple Vicky and Kevin, both aged 26, who live in a two-room apartment, see Figure 1 for reference. Our contact person was Vicky, which meant that she was the one who took part of the different interviews. She is a student and has stated that she has an interest in IT, she furthermore assessed her technical skills to be between practiced and expert.

Vicky is familiar with smart speakers and has gained familiarity with the devices through the internet. She does not use the virtual assistant on her phone regularly but has tested it to see what it is capable of. She stated that it does not make sense for her to use the virtual assistant on her smartphone:

> "When I have the phone at hand, I think it is a little silly to
> start talking to it, so I can just as well type it." - Vicky

She expected that using the smart speaker would teach her how to use hands-free

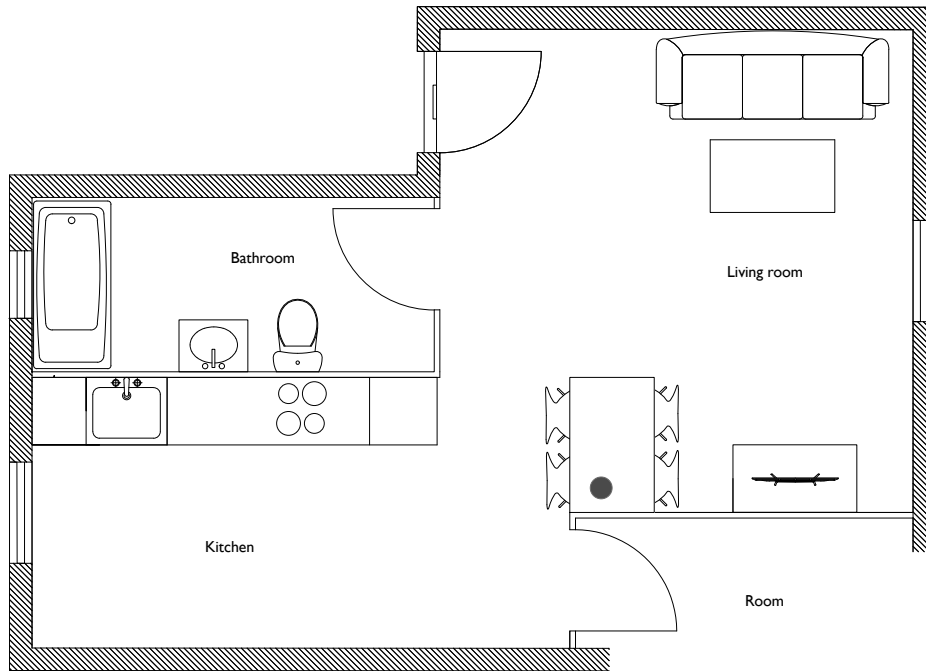Figure 3.4: Floor plan of household 1, where the black circle represents the placement of the speaker

Figure 3.5: Floorplan of household 2, where the black circle represents the placement of the speaker

technology, and she furthermore expected to use the functions music, timer and alarm the most since she thought these would be most convenient for her. Vicky participated in the study to figure out whether or not she should invest in a smart speaker, she stated that our study was a great opportunity for her to test it.

During the technology tour, Vicky talked about some of the technological devices she had in her home, she stated that they each own a smartphone, they furthermore have a desktop computer, a laptop, PlayStation, Wii, wireless speakers, iPod, and a TV. Out of these technologies they use the PlayStation and Wii once or twice a week, these were placed in their living room close to the TV. Whereas they use their smartphones the most, which she stated they use daily. She expressed, however, that she tends to misplace her phone regularly. When asked where she would place the smart speaker, it was important for her that it was placed central in their apartment, this way making

it easier for them to interact with it from all rooms. She therefore decided to place the speaker on their dining table, which is placed central in their apartment.

3.3.2 Household 2
Household 2 received the simplified smart speaker. This household consists of Matt who is a 23-year-old student, who lives in a one-room apartment, see Figure 2 for reference. He has an interest in IT and assessed his technical skills to be between practiced and expert. He is familiar with virtual assistants and assess his use of the assistant to be on a weekly basis, he uses Siri on his smartphone to set timers.

Matt is familiar with smart speakers and stated that he keeps up with technology, he has been keeping up with smart speakers since the first one was released. He has an Amazon Echo stored in a drawer, which he does not use this was unbeknownst to us

prior to conducting the study. He clarified that internet was included in the rent in his old apartment, but this is not the case in his new apartment. Now he uses a personal hotspot from his iPhone to connect to the speaker. He stated that the process of connecting to the smart speaker is complicated and an inconvenience to him. He used to use his Amazon Echo to listen to music and control the lights since he has his Philips Hue lightbulbs. He found it to be more convenient to say play some music rather than finding his phone and choosing a song to listen to.

He had great expectations to our smart speaker and expected it to be able to do everything he asked it to. He furthermore expected to use the speaker to listen to music and control his lights, just as he did with his Amazon Echo. He added that weather also could be relevant to him.

During the technology tour it was evident that Matt appreciates Apple products he owns an iPhone, iPad, MacBook, Apple Watch, he uses these products daily but mostly his iPhone. He additionally owns Philips Hue lightbulbs, which he does not use. Matt stated that he would purchase an Apple HomePod to expand his Apple ecosystem if he could afford it, additionally he states that he prefers listening to music via Spotify, but the Apple HomePod only supports Apple Music. We placed the smart speaker on his desk in the living room when we were there to setup the speaker, but he talked about placing the speaker in a central place in his apartment, which would be the triangular table close to the window.

3.3.3 Household 3

Household 3 received the complex smart speaker. This household consists of the couple Hank and Lucie, who live in a three-room apartment, see Figure 3 for reference. We were in contact with Hank, meaning that he took part in all interviews, however they both took part of the initial interview. Hank is a 25-year-old student and Lucie is 24 years old. Hank stated that he has an interest in IT and assessed his technical skills as being expert. He is familiar with virtual assistants and uses one on a weekly basis. They use the virtual assistant to play either music or films on their TV via Google Chromecast in addition to if they are unable to find the remote, they use the virtual assistant to turn off the TV. They furthermore use the virtual assistant in their car to speak on the phone while driving, but this a new feature that they are getting used to.
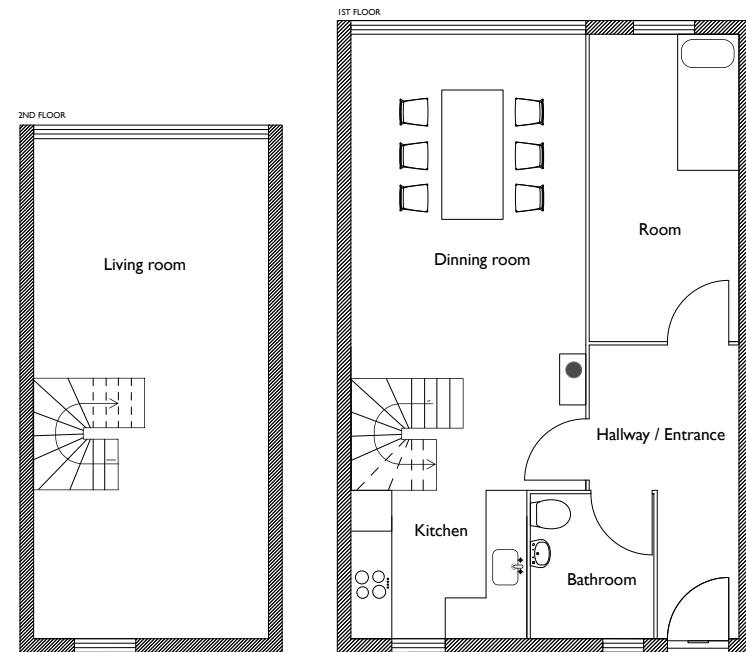

Figure 3.6: Floorplan of household 3, where the black circle represents the placement of the speaker

Both Hank and Lucie are familiar with smart speakers and they have gained this familiarity through commercials, the internet and news. They have spoken about purchasing a Google Home, but were afraid that they would not use it enough. The price was another reason that they had not bought one yet. They saw partaking in our study as a great opportunity to determine whether or not to purchase one, since they wanted to make sure that they used it enough and that they enjoyed using it.

They had limited expectations to the speaker, which was due to their existing knowledge about smart speakers. They expected it was able to play music and support home automation, since they owned Philips Hue lightbulbs. They furthermore saw the convenience of using the smart speaker while cooking to set timers. Overall, they expected it would be more convenient to use in certain situations since it was easier than to first locate their smartphones.

During the technology tour Hank and Lucie stated that they have smartphones, Google Chromecast, Philips Hue lightbulbs, two laptops, two desktop computers and a radio. They use all the listed technology but use their smartphones and laptops the most. They use their smartphones all the time except for when they are asleep, and they use the integrated virtual assistant to interact with lights and the TV. Hank is more prone to speak to the virtual assistant, where Lucie is more hesitant stating that it was unable to understand her at times. Hank added that it is because she tends to mumble. Lucie reflects on the placement of their devices, and decides the placement based on where it is most practical. For this reason, they had decided to place the smart speaker on a table close to the kitchen, since it was a central place in the apartment, which allowed them to talk to it while cooking and while eating in the dining room.

3.3.4 Household 4

Household 4 received the simplified smart speaker. This household consists of the married couple Jacob and Rachel, who live in a two-room apartment, see Figure 4 for reference. We were in contact with Jacob who is 25 years old and has a full-time job. He has an interest in IT and assesses his technical skills as being between practiced and expert. He is familiar with virtual assistants and uses an assistant on a weekly basis. He uses a virtual assistant in his car, which is integrated in his new car, he is therefore still learning how to use it.

Jacob is familiar with smart speakers as well as Apple CarPlay and has gained this familiarity through coworkers and friends. He expects the smart speaker to make it easier to access things that are online, and he can see the convenience in using hands free technology. He furthermore expected to use the smart speaker while cooking, to find and follow recipes. Jacob was not sure whether or not he would purchase a smart speaker, this was mostly due to him not owning any home automation devices in addition to Rachel not being fond of technological devices such as the smart speaker due to the privacy issues that can arise. Jacob did not see enough potential in the smart speaker without having these additional devices.

During the technology tour Jacob stated that he and Rachel own smartphones, a desktop computer, two laptops, TV, Google Chromecast, PlayStation, Xbox and a
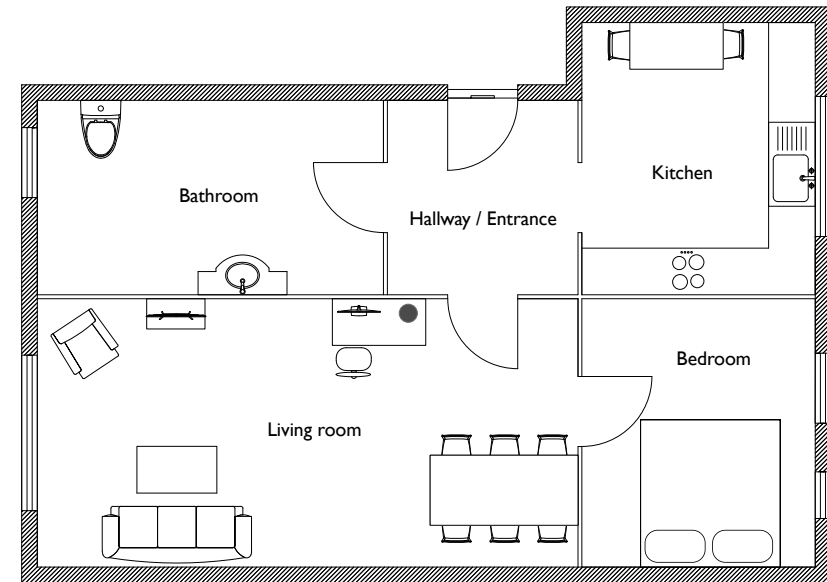

Figure 3.7: Floorplan of household 4, where the black circle represents the placement of the speaker

Bluetooth speaker. They each use their computers and smartphones but tend to use the TV together. Jacob stated that they use their TV and computers the most, the TV for relaxing and the computer for gaming. Jacob and Rachel have considered where to place their technological devices, and Jacob further stated that they will place the smart speaker centrally in their apartment, which was the table in the living room that was placed close to the door. This would make it easier to hear it throughout the apartment.

3.4 Summary of Chapter 3

This chapter entailed the different aspects of conducting our study. The study design included how we applied both the framework, temporality of experience in addition to the taxonomy for longitudinal studies, these were used to structure our design. With the purpose of giving an understanding of how we designed our study. The chapter additionally describes the process of conducting the study, to give an overview of how and when we contacted our participants, and describing how we conducted the interviews. Lastly, we gave a detailed description of our participating households, with the purpose of creating a relation between the households and their findings.

FINDINGS

This chapter contains the findings of our study. The purpose of this chapter is to display the findings of the conducted longitudinal case study in addition to shedding light on the differences between the households' appropriation of the simplified and complex smart speaker. The chapter is structured based on the three phases in the temporality of experience.

4.1 Various Usages

Based on the log data that was gathered for the different household, six categories emerged using an affinity diagram. These categories reflect how the households have used their speakers during the time period that they had them. When looking at Figure 4.1, there are four bar charts, one for each household, each chart has the categories, found in the affinity diagram, on the y-axis and the number of issued commands on the x-axis. The bars are divided into different shades, that represent the sub-categories.

Home automation, is where the users of the household, have used the speaker for controlling their different devices in the home. For instance, household 3 used their speaker for controlling the lights in their apartment. When looking at Figure 4.1, we can see that household 3 is the only household that have issued commands regarding home automation. Household 1 did not own any home automation devices therefore they did not use these types of commands. Household 2 and 4 had Philips Hue and Google Chromecast, respectively, but they received a simplified smart speaker and could therefore not use these devices. In the last interview Jacob, from household 4, stated that he wanted to invest in more home automation devices if he were to purchase a smart speaker.

Entertainment are the commands the participants have used in order to be entertained by the speaker. This category includes the subcategories music requests, jokes and Netflix. All households have tried entertaining them self with a joke and music. Household 2 and 4 with the simplified smart speakers have also tried to play a song, even though it is not a part of the available functionality in the simplified speaker. This could indicate that it is counterintuitive for a speaker not to be able to play music.

Information resolves around the user obtaining general information by using the speaker. This category contains the functionality where weather, news and factual search has been requested. Again, all households have contributed to this category. All have asked about the weather and three out of four households have used the smart speaker for fact search. Household 4 have also tried to fact search with the simplified speaker twice. Vicky stated the following:

"Now that it is Google, so could you [Google] tell me this, because that is the point with Google, it would surprise me it didn't work. It was a long shot." – Vicky

This could indicate that "Googling" is a well-established concept, that can be difficult not to explore when the user is saying "hey Google". This category is significantly more used in household 1 than any other of our households in this study and it is the most used category, where the sub-categories weather and news seems to be very popular.

Practical tool, here functions such as the ability to set a timer or change the speaker's settings were placed. All the households have used a practical tool. In the households 1 and 4 it is used very much, especially the timer. This can however be misleading, because for every time a timer is set, it will automatically also have an additional "stop the timer" or even "how much time is left" command, these are also represented in this sub-category.

Limitation is in the context of understanding the user, which have led to some gibberish commands that the speaker was not able to recognize. All households except for household 2 has generated data that was misunderstood.

Unavailable functions is a category that was only seen in the use of the simplified smart speaker. This category entails functionality that was unavailable to use, these unavailable functions included music, navigation, time, factual search and an Easter egg.

Partial Conclusion
Each household used the speaker differently. Many of the same categories and sub-categories can be found in all households. But we can see that the use in household 3 was more home automation orientated. Household 1 have used the smart speaker the most especially the information category. When looking at the households with the simplified smart speaker, we can see that both of them have tried to use commands that were not available and generally used it less, especially household 2.
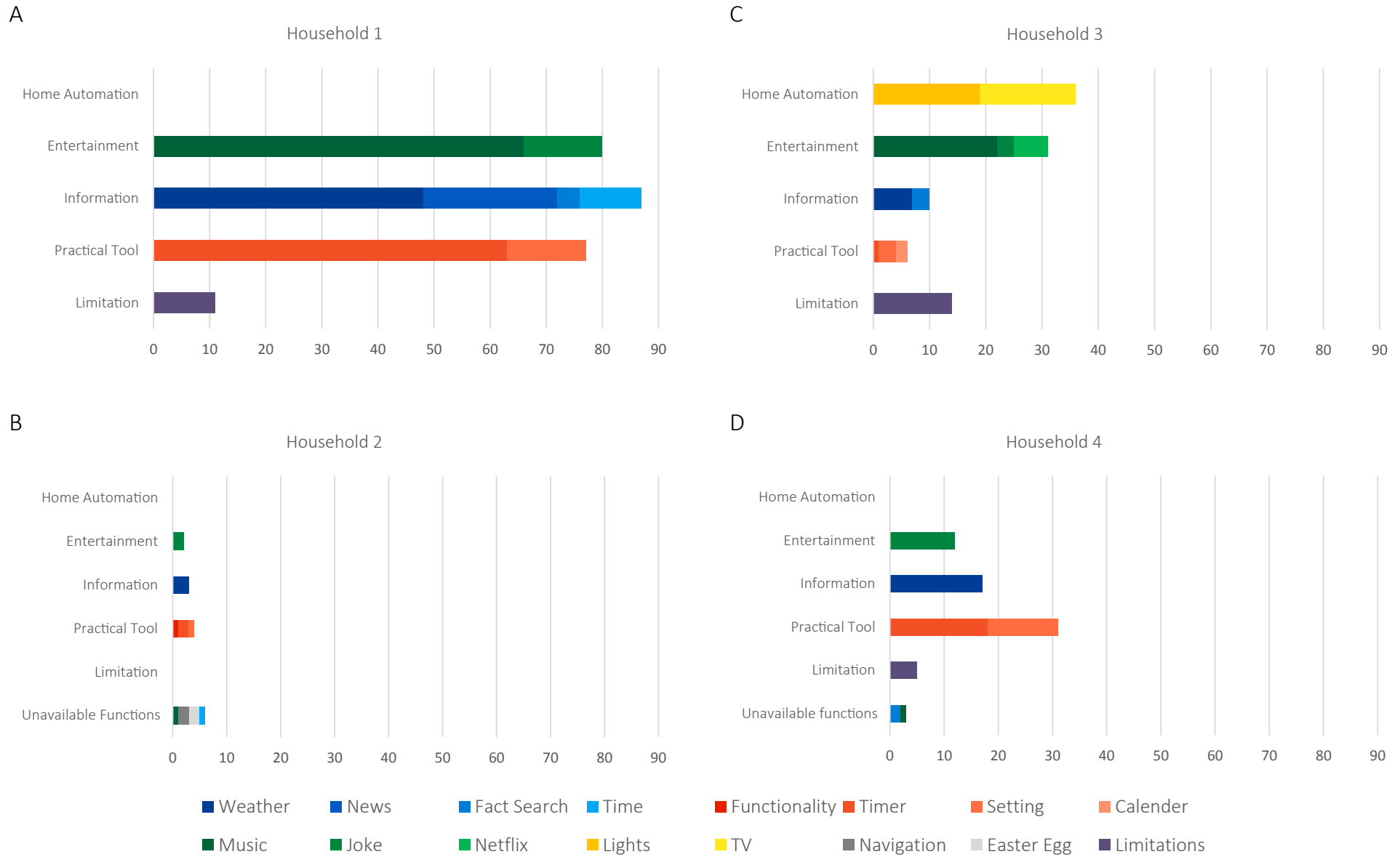
Figure 4.1: A bar chart for each households' representation of results from the affinity diagram
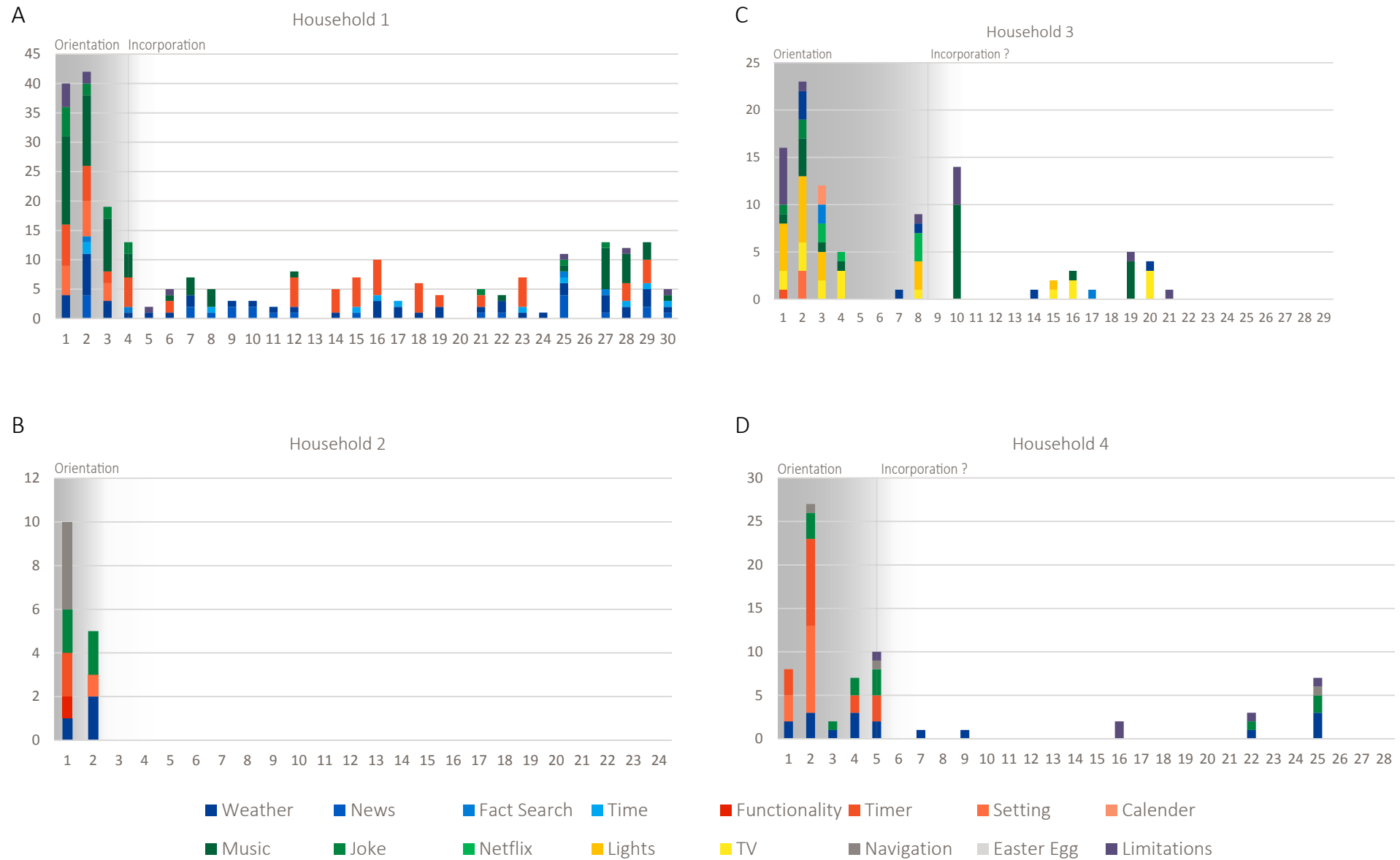
Figure 4.2: Indication of each households' orientation phase and incorporation phase

4.2 Orientation

It is evident in the log data and the interviews that all households had an orientation phase in which the households have experimented with the smart speaker to test its different functions, to then be able to determine which ones are useful for them and their needs. When examine Figure 4.2, there can be seen four bar charts, one for each household, the charts each shows the distribution of commands. The days of the study is on the x-axis, and on the y-axis the numbers of commands. Each bar is divided into sub-categories of the commands issued that day.

"Yes, of course you are more excited in the beginning because then you have a lot of things to test out. That is definitely something I recognize." - Vicky

The orientations phase in household 1 spans over the first, three to four days, which is evident in two ways. Firstly, when looking at Figure 4.2A the number of commands used per day is high and then decreases to a steady level. Secondly, the participants have investigated the various functionalities, which can be seen in the big variation of commands used in these first days. This was furthermore evident in the interview in which Vicky stated that she had been through this phase and that she felt that she had been through all functions. But after a short reflection she changed her mind and now stated that she probably missed some functions; this could indicate that she is not troubled if she has missed some functions. She stated that it is a part of receiving a new device to test it out and discover what it is capable of.

It is difficult to determine whether household 2 has had an orientation phase based on figure 4.2B alone, for this reason we decided to look at the log data. It is evident in the log data that Matt has tested some of the smart speaker's functionality to determine what it is capable of. During the first day he has made use of 10 commands in the time span of 20 minutes. Nine of these commands took him seven minutes to try out, these included weather, functionality, comedy, navigation, music and time. Whereas the remaining command, which was setting and stopping a timer took 13 minutes. On the second day Matt has issued four commands in the time span of two minutes, where he later in the evening issued two additional commands. The time span in which he has issued the commands on these two days is an indicator of the orientation phase.

This is furthermore apparent since Matt tried functions that were not available to him during the two days that he has used the device. Here he tried to get the smart speaker to navigate him to a hairdresser, asked for the time, tried out some pop culture references and tried finding an Easter egg. This can also indicate that his initial expectations to the smart speaker were not met, he also stated that he expected it to be an off-the-shelf speaker and not one we constructed ourselves.

The duration of the orientation phase is also short since Matt received an error message stating which functions are available to him, this way reducing the necessity to explore further, therefor he states that he with certainty had tried all available functions. As mentioned earlier he has an Amazon Echo that he does not use, to gain an understanding of his use of the Amazon Echo we asked him some questions in regards to the framework by Karapanos et al. Matt acknowledged that the orientation phase with the Amazon Echo was longer than with our smart speaker, which is also an indication that our smart speaker's functionality was insufficient for him. When asked if he felt that he found all functions in his Echo, he replied no, but that it did not bother him because he did not find them all useful. He used to use his Amazon Echo before he moved, which was due to the internet being included in his rent.

It is apparent when looking at Figure 4.2C that household 3 has been through an orientation phase with a duration of four to five days, because the bars start high and then decrease to a lower level of commands. We determine day eight on Figure 4.2C to also be a part of the orientation phase. Since Hank and Lucie were not home on day five and six, which was a weekend, we believe they continued their exploration of the smart speaker on day eight, which explains the number of commands and variation of functions used on that day.

Usually when Hank purchases a new product, he familiarizes himself with the product, this was however not the case with our smart speaker, since he already had required this knowledge beforehand through articles. This is also visible in the bar chart, see figure 4.2C, in the sense that the amount of commands given in the orientation phase are notably lesser than the number of commands given by household 1. We are able to

compare the orientation phase of household 1 and 3 since both of these households were given the complex smart speaker.

Hank has knowledge about the smart speaker, so instead of spending too much time on testing the different functions he rather tested its physical limitations such as how to speak to it and tested if it was able to hear him from certain distances. Despite his knowledge about the smart speaker, Hank stated that it is difficult to assess whether he has tried all the functions. He elaborated this and said he would be surprised if he has missed any vital functionality, but he is certain that there are functions that he has not tried.

When examining the log data, we can see that he used time on connecting their Philips Hue to the smart speaker. In the log we can see that he in the time period of two minutes, askes to turn on a specific light, where the virtual assistant replies that it cannot find the light bulb, Hank then asks which lights that are connected to the account and there are none. A few seconds later Hank is then able to turn on the dining room lamp, this clearly indicates that he successfully connected Philips Hue to the smart speaker.

Hank furthermore acknowledged that he has been through the orientation phase and stated in the interview that it is difficult to assess whether he has tried all the functions.

> "There was not so much that it required orientation.
> It was mostly about which functions were available and
> which were unavailable." - Jacob

It is evident in Figure 4.2D that household 4 has been through an orientation phase, which lasted four to five days. This can be seen since the bars start high and then decrease to a lower number of issued commands. Although day three has a low number of issued commands, we still view this day and the following two days to be a part of the orientation phase. This is due to Jacob and Rachel having prior engagements on day three and therefore being unable to use the smart speaker as much on that particular day. However, on the following days, day four and five, they continue with exploring the functions where it is seen in Figure 4.2D that they used weather, timer and joke on

day four and on day five they used the same functions in addition to the unavailable function factual search and asking the device why it was slow. This addition on the day five indicates that they were not done exploring on the previous days. When exploring functions in this phase, Jacob believe that he had tried all the functions that is available in the simplified speaker and have even tried functions that were not available. Jacob likes the concept of a simplified speaker, but he still wishes it had more functions.

Partial Conclusion
All the households have been through an orientation phase, with varying lengths two to five days. There are no significant differences in the length of the orientations phase nor number of used categories of commands. This finding is not as we expected it to be, as stated in section 1.2, we expected that this phase would be shorter for the households with the simplified smart speaker. When comparing the simplified and complex smart speaker based on the average of commands used per day in this phase, there is a significant difference. Where households with the simplified speaker have an average of 9,15 commands per day and the households with the complex have an average of 37,5 commands per day, it must be mentioned that household 1 raises this average.

4.3 Incorporation
The log data shows that household 1 have reached in the incorporation phase, see Figure 4.2, from day five and onwards we see that the bars reached a steady level, both in the numbers of commands per day and types of commands.

> "It also came more naturally after I had practiced
> it [using the smart speaker], that's also why I missed
> it in the end. I was used to saying, "Hey Google, set timer",
> wait no "Phone where are you?"" - Vicky

This quote from the interview suggests that she is in the second phase, incorporation, as she states that something is missing from her everyday life now. This furthermore shows that there has been established routines in the household. She was dependent on the speaker for some activities and uses the smart speaker to structure her day and as white noise for when she is alone. She also placed the smart speaker in a central spot

in their apartment, so that she and Kevin could access the speaker from most of the appartment, see sub-section 3.3.1. She describes herself and Kevin as being creatures of habit, meaning that they are prone to creating routines and maintaining these.

Vicky has during this study used nine types of commands. We especially want to draw attention to four of them, weather, news, timer and limitations. Weather is used every day except for a few days, and this is due to Vicky had asked about the weather on the previous days. News has been used the days Vicky has been home alone as she uses this function as white noise. The days in which news has not been used is mostly weekends and the Easter holiday where both Vicky and Kevin have been home, meaning that white noise has not been a necessity. Timer has been used frequently since she uses this function when cooking, she added that when the timer has not been used it is due to her not having to cook on that day. The category limitation is seen on the first two days and returns when problems occur such as when they had trouble with playing music, due to an update that had occurred which prevented them from playing music via YouTube, which they had been used to.

Matt has not reached the second phase, incorporation, since he only generated enough data to depict the orientation phase. When asked, during the interview, what was needed for him to reach this phase, he stated that home automation and connectivity to other speakers in addition to having Wi-Fi access could persuade him. In the interview he expressed that he has reached the second phase with his Amazon Echo by consistently using it for lights and music. He was not able to use these functions on our smart speaker, since he had received a simplified smart speaker.

It is difficult to determine whether or not household 3 have reached the second phase, incorporation, based on Figure 4.2C alone. This is due to the lack of issued commands after the orientation phase had ended. However, Figure 4.2C indicates that the incorporation phase spans from day 14 to day 21, since the usage of the smart speaker is fairly regular in these days.

During the second interview Hank expressed that he expected his usage of the smart speaker to be more than it was and was surprised when he saw his actual usage.

Household 3 used the smart speaker for home automation and an explanation for this lack of usage, Hank refers to, could be due to light not being as necessary during the spring months.

The days in which Hank or Lucie have not issued any commands is either due to them not being home or the placement of the speaker. As aforementioned household 3 lived in a two-floor apartment, in which the smart speaker was placed downstairs in the kitchen-dining area and their living room is upstairs, the floorplan made it difficult to use the speaker to interact with their TV and light upstairs. Despite this, Hank acknowledged that he has entered the second phase, incorporation.

Like household 3 it is difficult to determine if household 4 reached the incorporation phase. When examining figure 4.2D, we can see that Jacob had used the smart speaker in the first five days. After this there is occasionally issued a couple of commands, we wanted to investigate this through the last interview with Jacob. There he told us two reasons why there is days where the smart speaker was inactive. The first was the fact that they were not home during Easter, day 18-21.

"Rachel has sometimes unplugged it [the smart speaker] when it was on standby when she didn't want it to hear her. Because of that I have asked it about somethings when it wasn't on, then I perhaps turned it back on, but it loses some of the functionality it has by always being on and ready to be asked about something." - Jacob

But the second and very interesting reason, Rachel is a private person that wants to uphold her privacy therefore she unplugged the smart speaker when not in use. This can explain the inactivity, because the potential that Jacob saw in the handsfree interaction were shattered. Despite this, we can see a steep increase in commands on day 25, this can be bias because we reached out to arrange the last interview the day before.

Partial Conclusion
There is no doubt that household 1 has reached the incorporation phase with the complex smart speaker, it is a little unclear if household 3 reached it or not. We can with

confidence say that neither household 2 nor household 4 with the simplified speakers made it to this phase. This could indicate that a simplified speaker is not enough to reach the incorporation phase, but we need to consider that Household 2, was not the most ideal participant, because Matt already owns an Amazon Echo as well as the fact that Rachel in household 4 turned off the smart speaker when it was not used. It is therefore difficult to determine whether or not our expectation, from section 1.2, about this phase were fulfilled due to of the extreme instance with the simplified speakers.

With the obscurity in household 3, it is difficult to determine whether they reached the incorporation phase, makes it difficult to determine. Hank stated himself that they had reached this phas, but looking at Figure 4.2 and the interviews, we are not sure. After the water damage in the apartment, they did not turn on the smart speaker again even though it was not damaged. They could have moved it upstairs to their living room, which they stated also could be a useful placement.

4.4 Identification

We do not assess any of our participants to have developed an emotional attachment to the smart speaker. Household 1, however, showed potential in developing an emotional attachment but had not yet reached this phase in the duration of our study.

"Honestly I missed it in some strange way. It wasn't because I thought I used it a lot, but I could feel it [was missing]. But also, in terms of its placement, it is as if I was missing something in my home…" - Vicky

Although Vicky said that she missed the speaker, we interpret this as being an indicator of the second phase, incorporation. This is due to Vicky had stated that herself and Kevin are creatures of habit, which makes her statement of missing the speaker seem more like that she misses the routines that they had already established.

On one hand Hank and Lucie, from Household 3, showed the smart speaker to their guests as a type of gimmick, which could indicate an emerging emotional attachment, since this involves developing a social and/or personal relation with a device. But on the other hand, Hank stated that he could not see himself developing an emotional attachment to the smart speaker.

"It is going to be a glorified power outlet, a switch, an on/off thing." - Hank

He went as far as calling it a "glorified power outlet". He did not view the speaker to have any functions that would enable him to develop an emotional attachment to it. Hank stated, during the interview, that he did not notice that the smart speaker was not in his home anymore. He elaborated this and stated that he had turned off the speaker one week before returning it due to water damage in his apartment.

Matt from household 2 has not reached the third phase, identification, and he even stated that he has not reached this phase with his Amazon Echo. He proclaimed, however, that if he had the functions that he deemed important to him, such as home automation and music via Spotify, then he would feel emotionally attached to the device.

It is evident in the log data as well as the conducted interviews that the functionality provided in the smart speaker with limited functionality were insufficient for him, it is evident in figure 4.1b that he tried using functions that were unavailable to him. He stated that the biggest deal breaker was the functionality, since they did not meet his expectations to the smart speaker. He furthermore stated that the fact that he did not have WI-FI access was an inconvenience, since he then had to connect the smart speaker to his personal hotspot.

To develop an emotional attachment Jacob, household 4, stated that he had to invest in additional home automation devices in addition to mutually agreeing with Rachel to purchase one.

Household 2 and 3 both stated that they consider themselves to have an emotional attachment to their smartphones, this is due to their smartphones providing the functions that they deem important and useful for them and their situation. Hank expressed that his smartphone is capable of many things and can be adjusted to his need. Vicky and Jacob from household 1 and 4 respectively, tend to misplace their smartphones, and Vicky stated that she does not want to have her smartphone on her person constantly.

Partial Conclusion

None of the participants showed a clear indication of having developed an emotional attachment to the smart speaker. Household 1 and 3 both showed tendencies of possibly developing an emotional attachment. This was prevalent in household 1 since they had established routines and Vicky used the smart speaker to structure her days, which made it slightly difficult to return to her previous ways of structuring her daily life. While household 3 exhibited this tendency by showing the smart speaker to their guests. Whereas household 2 and 4 did not show tendencies of developing an emotional attachment to the smart speaker, since Matt stated he needed functions that were better suited for him and Jacob stated he needed to invest in home automation devices. Our expectations were that the complex smart speaker is more likely to reach this phase. None of the households did with certainty reach this phase. But household 1 and 3 shows the tendency and got further in the framework compared to household 2 and 4, therefor this exportation is for filled.

This indicates that the participants with a complex smart speaker are more likely to develop an emotional attachment, whereas the simplified smart speaker simply is insufficient to develop an emotional attachment.

DISCUSSION

This chapter contains a discussion/reflection of the different aspects involved in conducting our longitudinal case study. It emphasizes on key points such as the impact of the complexity in the smart speakers, whether knowing all the available functions in the smart speaker is a necessity in addition to advantages and disadvantages of conducting longitudinal case studies. It furthermore includes a discussion on whether developing an emotional attachment to our smart speakers is possible and lastly the participants' motivation for participating in our study.

5.1 Complexity of Smart Speakers

Our research question stated that we wanted to see how users appropriate a smart speaker based on complexity. Where we created smart speakers with two levels of complexity, to investigate this question. Based on the results from our study we do not have enough evidence to make any general assumption about the participants' appropriation. But it seems that the few functions that we chose, based on the literature, are not enough for the participants to get to the emotional attachment phase of the temporality of experience framework. We also see other factors were decisive to why the households with the simplified speakers did not get to that phase. Matt in household 2 was probably not an ideal participant, because he already owned an Amazon Echo, however he did still contribute with great insights. In household 4, Rachel was very concerned about her privacy to the extent that she often unplugged the smart speaker, this had according to Jacob an impact on his use and attachment to the smart speaker.

5.2 Uncertainty of Capabilities

The related work that described the frustrations that the user can feel when they are unsure about the functions of a smart speaker. The uncertainty of whether users have tried all the available functions or not can lead to this frustration[25, 35]. This made us wonder whether making a simplified smart speaker would eliminate the user frustration. The frustrations which can occur when familiarizing oneself with a new product is also shed light on by Karapanos et al., they state that while the users' experiences evolve over time, as they begin to become familiar with the speakers one would expect them to experience less frustration, as they begin to be accustomed to a few functions but also fewer exciting episodes, since they grow accustomed to these functions [Karapanos et al]. This can be seen with our participants. The category limitations are more apparent in the orientation phase, where the participants are learning how to use the smart speaker. When we asked our participants, with the complex smart speaker, if they felt that they had tried all the functions they said yes but they were also aware that they properly had missed minor functions. When examining the categories of functions in Bentley et al. and Sciuto et al. studies, we can see that none of our participants had tried to make a list, reminder or play one of the many games available on the smart speaker. When comparing functions used in the households, there is an overlap in functions, but 3 out of 4 have also tried functions the other has not.

5.3 Longitudinal Case Study

Conducting a longitudinal case study entails observing people and their use of a specific device over time which has its advantages and disadvantages. An advantage is that by using this approach one can obtain a deep understanding of a given phenomenon, whereas a disadvantage could be that the mere fact that people are being observed can result in them changing their behavior meaning that the study itself can impact the end result.

Vicky from household 1 kept apologizing for not using the smart speaker enough, making it seem like she was more concerned with pleasing "our study" and not as concerned with us getting a representation of her real usage. This was however not evident in the collected log data, as it was apparent, through the interviews and the collected log data, that she used it regularly in her daily life.

We were slightly concerned that when contacting the participants to schedule an interview or when prompting them to tell them that we were going to download their data for data analysis would have an impact on their use. This was seen once with household 3, where we contacted them on day 24 to schedule the last interview which resulted in an increase of commands issued on day 25. It was, however, not evident with the remaining households that this was the case, meaning that there was not an increase in the number of commands issued based on the days we contacted household 1, 2 and 4. Based on the collected data through both the log and interviews, we believe that we got a representational insight in the different household's usage.

5.4 Emotional Attachment

Karapanos et al. describe in their framework, three phases when receiving a new product[14]. The last phase, emotional attachment is especially of interest. Karapanos et al. assessed whether their participants had reached this phase based on if they had formed a personal relationship with the device, meaning they would personalize the device to fit their personality and needs[14]. We do not assess any of our participants to have developed an emotional attachment to the smart speaker, this assessment is based on the log data in addition to their own statements in the interviews.

Vicky from household 1 had tendencies of potentially forming an emotional attachment to the smart speaker but had not yet reached this phase in the duration of our study. Vicky stated that she missed the smart speaker, which could indicate an emotional attachment, but is the word missing enough to determine whether she has developed an emotional attachment? We interpreted her statement of missing the speaker as an indication of the second phase incorporation, meaning that the smart speaker had become a part of her routines and she viewed it as an assistive technology in her daily life. It was furthermore evident in the fact that upon returning the smart speaker she did not purchase a speaker of her own.

Matt and Jacob stated that although they had not developed an emotional attachment to the smart speaker, they did have this attachment to their smartphones. They argued that it was due to the smartphone having the exact functions that were useful for them, in addition to the fact that they always had their phones on their person. The fact that the participants had not developed an emotional attachment to the smart speaker could be due to them having to return the speaker at the end of the study – or to the fact that it is not a personal product.

Karapanos et al. found in their study that developing emotional attachment was dependent on the type of product that was being used. In their study they made use of the iPhone which is a product that is considered to be personal due to it enabling people to connect with others, allows for personalization to fit the individual user and is furthermore close at hand making it mobile and convenient to use[14]. This is consistent with what we found, since Matt and Hank stated that they had developed an emotional attachment to their smartphones for these exact reasons.

The smart speaker is not a social device compared to the smartphone. This is seen in the placement of the two devices. While the smartphone is usually nearby or close to hand, the smart speaker tends to be placed in a place that allow multiple people to use it. This is consistent with articles found in the literature review where it was prevalent that when people have one smart speaker, they tend to place it in the living room or kitchen making it available to everyone in the household[29, 35]. If people have more than one smart speaker, they tend to distribute them throughout the house, placing them where

they conduct their daily routines making it able to assist them when needed[34]. This is consistent with the findings in our study, as we saw a tendency for the participants to place the smart speaker centrally in their respective apartments, making it available to the full household.

5.5 User Motivation

Karapanos presented the framework, temporality of experience on CHI '09, where this presentation was video documented[15]. During this presentation Karapanos mentions that for their study they recruited participants that already had decided to purchase an Apple iPhone, the object for their study. They did this because they assumed that, their participants were more motivated to use and familiarize with the product because it is their own, compared to if they borrowed a product form the researchers for a limited period of time[15]. In our study we used this, according to Karapanos, problematic approach where our participants were given a smart speaker that they needed to return to us after the study. Despite this we believe that our participants still wer motivated to use the smart speaker as if it were their own, because three out of four participants participated in our study because they wanted to try out the smart speaker to help them decide whether or not to buy a speaker. The last participant already had a smart speaker stored in a draw, here we saw that this participant was the least motivated to use the smart speaker, because he experienced the same challenge with our smart speaker as he did with his own, a problem with not having WI-FI and only using iPhone personal hotspot.

CONCLUSION

This report presented a four-week longitudinal case study with the purpose of getting an in-depth understanding of the long-term use of the smart speaker and the changes that can occur. The study additionally set out to examine the differences in how users appropriate the smart speaker over time, depending on its complexity.

The purpose of the conclusion is to answer the research question:

How are smart speakers appropriated in accordance to the framework, temporality of experience's three phases over time, depending on complexity?

We saw that all households, despite having a simplified or complex smart speaker, has been through an orientation phase, in which they explored the capabilities as well as functionalities of the smart speaker. There was no significant variation in length of this phase depending on the complexity of the smart speaker, although we saw an extensive variation when looking at the average of commands issued per day across the two variations of the speaker. The simplified smart speaker had an average of 9,15 commands per day, whereas the complex smart speaker had an average of 37,5 commands per day. We saw in the incorporation phase that the households with the complex smart speaker were more likely to establish routines. This was prevalent in household 1 since it was clear through both the log data and interviews that they had established a routine with the smart speaker.

The households with the simplified smart speaker had not reached this phase, meaning they had not a regular use of the available functions. It was established during the interviews that household 2 and 4 could not see a usefulness of the available functions, where they stated that this was due to not having the relevant functions for their needs in addition to not having home automation devices.

None of the four households developed an emotional attachment to the smart speaker. We did, however, see a tendency in the households with the complex smart speaker, household 1 and 3, to potentially develop an emotional attachment. Household 1

displayed this tendency by having difficulty returning to their regular way of structuring their days, since Vicky viewed the smart speaker as an assistive technology. While Hank in household 3 showed this tendency when he showed the smart speaker to his guests, indicating the social aspect of emotional attachment that Karapanos et al. refer to.

We did not see this potential with the remaining households, household 2 and 4, with the simplified smart speaker. They both stated that the available functions in this variation of the speaker was insufficient for them, where Jacob added that he needed additional home automation devices for it to make sense for him to buy one.

Because it is not obvious that any of the households, despite the level of complexity, reached the identification phase it could indicate the reason was the smart speaker and not the complexity. Karapanos et al. presented the framework with an Apple iPhone as the product. This type of product is personal whereas the smart speaker is purchased for the household. This can indicate that users do not easily get emotional attached to a shared product.

The findings from our study indicate that the complexity has an impact on the three phases presented by Karapanos et al., and how the participants appropriated the smart speakers. In conclusion the households with the complex smart speaker were more likely to go through all the three phases in the temporality of experience framework. The study furthermore indicates that the available functions in the simplified smart speaker were not sufficient for the participants to inkorporate the speaker into their daily life.

REFERENCES

[1]   Bargas-Avila, J. and Hornbæk, K. (2011). Old wine in new bottles or novel challenges. Proceedings of the 2011 annual conference on Human factors in computing systems- CHI '11.

[2]   Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B. and Lottrjdge, D. (2018). Understanding the Long-Term Use of Smart Speaker Assistants. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), pp.1-24.

[3]   Bhattacharya, A., Windleharth, T., Ishii, R., Acevedo, I., Aragon, C., Kientz, J., Yip, J. and Lee, J. (2019). Group Interactions in Location-Based Gaming. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems- CHI '19.

[4]   Cho, J. (2018). MENTAL MODELS AND HOME VIRTUAL ASSISTANTS (HVAs): A SINGLE TURN-TAKING IS NOT ENOUGH TO HAVE A HUMAN- LIKE CONVERSATION.

[5]   Cohen, P., Cheyer, A., Horvitz, E., El Kaliouby, R. and Whittaker, S. (2016). On the Future of Personal Assistants. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems- CHI EA '16.

[6]   Conati, C. (2013). Virtual Butler: What Can We Learn from Adaptive User Interfaces?. Your Virtual Butler, pp.29-41

[7]   Cowan, B., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D. and Bandeira, N. (2017). "What can I help you with?." Proceedings of the 19th International

[8]   Czibula, G., Guran, A., Czibula, I. and Cojocar, G. (2009). IPA- An intelligent personal assistant agent for task performance support. 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing.

[9]   Gerken, J. and Reiterer, H. (2009) A Taxonomy for Longitudinal Research In HCI. Mensch & Computer 2009: Grenzenlos freil?. Oldenburg Verlag

[10]   GitHub. (2019). shivasiddharth/GassistPi. [online] Available at: https://github.com/shivasiddharth/GassistPi [Accessed Mar. 2019].

[11]   Gustafsson, J. (2017). Single case studies vs. multiple case studies: A comparative study.

[12]   Huber, B., McDuff, D., Brockett, C., Galley, M. and Dolan, B. (2018). Emotional Dialogue Generation using Image-Grounded Language Models. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems- CHI '18.

[13]   Iba, W. (2013). Before We Get There, Where Are We Going?. Your Virtual Butler, pp.54-69.

[14]   Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J. (2009). User experience over time. Proceedings of the 27th international conference on Human factors in computing systems - CHI 09.

[15]   Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J. (2009). User experience over time: an initial framework. [online] Available at: https://dl.acm.org/ft_gateway.cfm?id=1518814&ftid=1186132.

[16]   Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M. and Lee, J. (2018). Can a Machine Tend to Teenagers' Emotional Needs?. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18.

[17]   Kim, S., Row, Y. and Nam, T. (2018). Thermal Interaction with a Voice-based Intelligent Agent. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems- CHI '18.

[18]   Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A., Zitouni, I. and Anastasakos, T. (2016). Understanding User Satisfaction with Intelligent Assistants. Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval- CHIIR '16.

[19]   Koda, T., Maes, P. (1994) Agents with faces: The Effects of Personification of Agents In Proceedings of Human-Computer Interaction'96. London

[20]   Kujala, S. and Miron-Shatz, T. (2013). Emotions, experiences and usability in real-life mobile phone use. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems- CHI '13.

[21]   Lazar, J., Feng, J. and Hochheiser, H. (2017). Research Methods in Human-Computer Interaction. 2nd ed. Morgan Kaufmann.

[22]   Leite, I., Martinho, C. and Paiva, A. (2013). Social Robots for Long-Term Interaction: A Survey. International Journal of Social Robotics, 5(2), pp.291-308

[23] Li, J., Zhou, M., Yang, H. and Mark, G. (2017). Confiding in and Listening to Virtual Agents. Proceedings of the 22nd International Conference on Intelligent User Interfaces- IUI '17.

[24] Lopatovska, I. and Williams, H. (2018). Personification of the Amazon Alexa. Proceedings of the 2018 Conference on Human Information Interaction&Retrieval- CHIIR '18.

[25] Luger, E. and Sellen, A. (2016). "Like Having a Really Bad PA." Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems- CHI '16.

[26] Milhorat, P., Schlögl, S., Chollet, G., Boudy, J., Esposito, A. and Pelos, G. (2014). BUILDING THE NEXT GENERATION OF PERSONAL DIGITAL ASSISTANTS. International Conference on Advanced Technologies- ATSIP'2014

[27] Mitchell, T., Caruana, R., Freitag, D., McDermott, J. and Zabowski, D. (1994). Experience With a Learning Personal Assistan.

[28] Moussawi, S. (2018). User Experiences with Personal Intelligent Agents. Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research- SIGMIS-CPR'18.

[29] National Public Media. (2018). Latest Report | National Public Media. [online] Available at: https://www.nationalpublicmedia.com/smart-audio-report/latest-report/.

[30] Nunamaker, J., Derrick, D., Elkins, A., Burgoon, J. and Patton, M. (2011). Embodied Conversational Agent-Based Kiosk for Automated Interviewing. Journal of Management Information Systems, 28(1), pp.17-48.

[31] Pyae, A. and Joelsson, T. (2018). Investigating the usability and user experiences of voice user interface. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct- MobileHCI '18.

[32] Raspberry Pi. (n.d.). Raspberry Pi Community - Projects, Blogs, and more. [online] Available at: https://www.raspberrypi.org/community/ [Accessed Mar. 2019].

[33] Porcheron, M., Fischer, J., Reeves, S. and Sharples, S. (2018). Voice Interfaces in Everyday Life. Proceed- ings of the 2018 CHI Conference on Human Factors in Computing Systems- CHI '18.

[34.] Sciuto, A., Saini, A., Forlizzi, J. and Hong, J. (2018). "Hey Alexa, What's Up?" Proceedings of the 2018 on Designing Interactive Systems Conference 2018- DIS '18.

[35] Smith, B. (2017). OK, GOOGLE: DESIGNING INFORMATION ARCHITEC TURE FOR SMART SPEAKERS.

[36] Velkovska, J. and Zouinar, M. (2018). The illusion of natural conversation: interacting with smart assistants in home settings.

[37] Vtyurina, A., Savenkov, D., Agichtein, E. and Clarke, C. (2017). Exploring Conversational Search With Humans, Assistants, and Wizards. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems- CHI EA '17.

[38] Walker, J., Sproull, L., Subranmani, R. Using a Human Face in an Interface. In Proceedings of CHI'94.

[39] Wexelblat, A. (1998). Don't Make That Face: a report on anthropomorphizing an interface. American Association for Artificial Intelligence.

[40] Yin, R. (2013). Case Study Research. 5th ed. [Erscheinungsort nicht ermittelbar]: SAGE Publications - Books.

[41] YouTube. (2017). All-New Amazon Echo (2017, 2nd Generation) Unboxing and First Impressions. [online] Available at: https://www.youtube.com/watch?v=Qg4H9YDFdh8.

[42] YouTube. (2018). Apple HomePod: Unboxing & Review. [online] Available at: https://www.youtube.com/watch?v=XrLh0vAuIKs.

[43] YouTube. (2017). Google Home Official Ad. [online] Available at: https://www.youtube.com/watch?v=OsXedJq1aWE&t=45s.

[44] YouTube. (2016). Google Home: Unboxing & Review. [online] Available at: https://www.youtube.com/watch?v=m2lyB_gzR84&t=116s

[45] YouTube. (2019). Here is your new smart speaker - Complex. [online] Available at: https://youtu.be/3IiWTehi5A0.

[46] YouTube. (2019). Here is your new smart speaker - Simplified. [online] Available at: https://youtu.be/rqk9PB_pezk

[47] YouTube. (2017). Introducing Amazon Echo. [online] Available at: https://www.youtube.com/watch?v=zmhcPKKt7gw&t=65s.

[48] YouTube. (2018). Welcome Home (Apple HomePod Ad). [online] Available at: https://www.youtube.com/watch?v=BLAnTQfFXOI.