OK, GOOGLE: DESIGNING INFORMATION ARCHITECTURE FOR SMART SPEAKERS

Bailey Smith

AALBORG UNIVERSITY

Master's Thesis for Information Architecture

Submitted 19 October 2017

Character count: 186,105

Page count: 77.5 pages

Contents

Table of figures

1. Summary of project

The question addressed in this project, “What constraints do users experience with respect to finding and managing information, and what might an IA designer do to improve user experience?” is approached from a qualitative case study of smart speaker users. Over the past five years, two smart speakers have been released, becoming part of a group of devices with intelligent personal assistants that operate via voice interaction, a development that only recently become commercially available and viable. The technology of intelligent personal assistants has finally become available for home use in the form of the smart speakers Google Home and Amazon Echo. Consumer interest in these smart speakers has increased since the Echo’s announcement in 2014 and release to Amazon Prime customers in 2015. As other companies begin to add their smart speakers to the market, the number of available devices and contributors to the field increases, and the conversation about what role information architecture and information architects can play becomes more and more relevant. However, research on these intelligent personal assistants has only recently started to be published, and research on the smart speakers is almost nonexistent. Therefore, this project serves to help further the research into and discussion of designing information architecture for smart speakers and intelligent personal assistants in the academic world.

In this project, information architecture is addressed more as a broad concept of structural support of user navigation of an information space rather than a finely grained set of rules and regulations for guiding a user through a space. This is because information architecture is highly dependent on context, content, and user, and the relevant components of any information architecture will be defined on a case-by case basis. This issue is even more relevant in terms of the conversational nature of the intelligent personal assistant (conversation itself relying on context and the collaboration between machine and human to be successful) and the fact that these devices are in the user’s home (further

pushing the relevance of the consideration of context and the user in design of information architecture). This idea of the relevance of the context, content, and user in information architecture design is corroborated by computer scientist Joseph Weizenbaum, psycholinguist Herbert H. Clark, computer scientist and natural language programmer Terry Winograd, and professor of Anthropology of Science and Technology Lucy Suchman. Considering the nature of these smart speakers and the role they serve in the users' homes and lives, the inclusion of users is even more important in this context.

In order to investigate the context in which the smart speakers and the users exist and to propose guidelines for designing information architecture for smart speakers, the scope of the existing literature on these smart speakers (and similar technology) is presented in order to establish the current context of this project. Following this contextual exploration, I conducted 19 user interviews either over Skype or in the users' homes to ensure that, in recommending information architecture design considerations, I was speaking from an understanding of the context, the users, and the content of the user's experience, and to ensure that the voice of the users continues to be heard in the building of guidelines for designing information architecture for smart speakers.

2. Introduction

Over the last decade, devices which allow user-machine interaction via voice have picked up momentum and consumer and market interest, mostly due to their recent commercial availability and viability. While the existence of voice control and continuous speech recognition are not a recent development, the technology has only recently become commercially viable and available. The introduction of the intelligent personal assistant (IPA) called Siri into iOS, Apple's mobile operating system (OS) for the iPhone, is the most notable instance of the introduction of a successful voice-user interface (VUI) into a consumer product—the introduction of VUIs into mobile devices is what arguably marked the breakthrough of VUIs into the consumer market.

The technology behind voice control (including voice recognition and command recognition) has improved in many ways, and the interactions have become more fluent. This development has allowed voice control to become more accepted in everyday interactions with technology and thus is being implemented in more and more consumer products, such as watches, cars, phones, televisions, and home automation systems. This last area, the domain of home automation systems, is where the smart speakers called Amazon Echo and Google Home, both voice controlled wireless speakers, find themselves. The Amazon Echo came first in 2014, and was the first product available for the home (as opposed to on a mobile device) on the consumer market to use voice interaction with an IPA as the primary interaction modality. The Google Home, serving largely the same purpose, was released in 2016, also featuring voice interaction with an IPA as the primary interaction modality. Both products are also supported by an app, but users are expected to interact primarily via voice with the products.

2.1. Motivation

As more systems and products take over traditional forms of interaction (that is, face-to-face, person-to-person interactions), there is an opportunity for these smart speakers to also adopt some part of the functionality currently controlled by mobile apps and websites. However, there is a key difference between websites/mobile apps and the way these devices respond to interactions with the user. One key element of these devices, the Google Home and the Amazon Echo, is that they attempt to be conversational. Whether they are truly able to have a conversation is not the question addressed by this project. For this project, it is accepted that these devices can and do provide conversational interaction for users at some level. For the Echo, Amazon Web Services have created Amazon Lex, "a service for building conversational interfaces into any application using voice and text," and they state that "[w]ith Amazon Lex, the same deep learning technologies that power Amazon Alexa are now available to any developer, enabling you to quickly and easily build sophisticated, natural language, conversational bots

('chatbots')" (Amazon Lex, 2017)—clearly, Amazon's goal with the Echo is to have apps (what Amazon calls "skills"), through the Alexa assistant, capable of conversation. The Google Home operates through the Google Assistant, and development of apps for the Google Assistant can be designed "for conversations for a variety of surfaces, such as a voice-centric conversation for Google Home or a visual conversation on an Android phone" (Extending the Google Assistant, 2017). Google Developers claims that, "Unlike with traditional mobile or desktop apps, users interact with apps for the Google Assistant through a conversation or natural-sounding back and forth exchanges, and not traditional, computer-centric paradigms" (Extending the Google Assistant, 2017). Both products claim to offer the possibility of delivering services to the user through a conversational mode.

Furthermore, much of the research on VUIs in the household has been in the field of accessibility for disabled or elderly. Now that these conversational VUIs are entering people's households and everyday use, and there appears to be interest in the consumer market, there is an interesting research question to be asked about how information architects can help improve the user experience with smart speakers. The idea of designing an information architecture around a conversational, VUI-based device, provides a relatively new challenge, as most of information architecture has been designed around navigating a graphical user interface (GUI). I hope to address the fact that an information architecture designer still has a place in designing for a technology with no visible interface.

2.2. Research scope

This project is written with a focus on the role of information architecture designers of a certain group of technologies released in the last five years which allow user interaction via voice with an intelligent personal assistant within a smart speaker. The technologies in focus in this project are the Google Home, supported by the Google Assistant IPA, and the Amazon Echo, supported by the Alexa IPA, with some examples from Siri, an AI available on Apple mobile device operating systems in the iPhone and iPad.

The goal of the project is to take a critical look at designing information architecture for smart speakers through the lens of users' reported experience with them. This will be done by exploring basics of information architecture provided by Rosenfeld, Morville, and Arango (2015), taking an overview of the history of research in this field, and interviewing users of the smart speakers. From this, I develop an outline based on the user responses for considerations to be taken in designing information architecture for smart speakers.

2.3. Research Question

What do users of smart speakers experience with respect to finding and managing information, and what might an information architecture designer to do improve user experience?

2.4. Thesis structure

The rest of this thesis proceeds as follows: Chapter 3 provides a brief introduction to the smart speakers being discussed in this project for the reader's reference. Chapter 4 includes a brief history of human-computer interaction, provides this project's definition of information architecture, and explores relevant concepts to consider in designing information architecture. Chapter 5 is the literature review. Chapter 6 describes the methods employed to collect the user interviews and information about the smart speakers from the users. Chapter 7 presents the interviews in narrative form, telling the story of each participant's experience with their smart speaker, as well as photos of the smart speakers in the users' homes to provide concrete examples of the context of these devices. Chapter 8 explores the themes from the data to understand the users' experience, and Chapter 9 takes these themes and helps provide insight into where an information architect designer might help the most. Chapter 10 takes the themes and uses them to provide suggestions to designers of information architecture for smart speakers and intelligent personal assistants, and it also presents possibilities for further research. Finally, chapter 11 presents the conclusions of the thesis.

2.5. Contribution

This thesis contributes to the information architecture community in the form of a set of proposed design considerations for information architecture designers of smart speakers, grounded in research from the fields of information architecture and interaction design.

Additionally, this work will hopefully contribute to a growing field of research being done with intelligent personal assistants (also referred to as conversational agents by some authors) and smart speakers. There has been little, if any, work done on understanding and assessing the role of information architecture as these devices begin to grow in availability and presence in users' homes, and this project will hopefully take its place as part of information architecture's contribution to a larger context of designing for this new and growing interaction modality.

3. Introducing the Google Home and the Amazon Echo

3.1. Google Home

Figure 1 Google Home images, from Target.com: front (on left) and back (on right)

Google Home was released in November 2016 for users in the United States of America. It continuously listens (but does not record) for the wake-up word, "Ok, Google." Once this is heard by the device, the user may continue with their query. Its physical features include a set of colored LEDs which help inform the user of the status of the device (e.g., if its loading information), capacitive touch controls which can start and stop music and adjust speaker volume, and a mute button on the back (seen on right in Figure 1) which disables the microphones. The Home includes Google Assistant as the intelligent personal assistant, which is also available on Android phones. The Google Home speakers can be linked to provide synchronized music in every room where it sits, and it also can serve as a hub for controlling smart home devices (including the Chromecast, products from Nest, and lightbulbs that can be controlled over Wi-Fi).

3.2. Amazon Echo

Figure 2 Amazon Echo (on left), and Echo Dot 2nd generation (on right), from Amazon.com

The Amazon Echo (left in Figure 2) was originally released in 2015 to Amazon Prime users. It responds to the wake word "Alexa," which allows the user to continue once the device has recognized the activation word, and refers to the intelligent personal assistant called "Alexa." Its physical features include a mute button on the top to deactivate the microphone; a button called the "Action button" which can turn off a timer or alarm, or wake the device, or enable Wi-Fi setup; and a light ring around the top which informs the user about the device's status. Underneath the light ring, there is the volume ring, which can be turned to manually change the volume of the device. Finally, there is a power LED on the back by the power port to indicate if the device is plugged in.

The Amazon Echo Dot is similar to the regular Echo in that it has a mute button, an action button, a light ring, and a power LED. There might be differences between the first generation (which has a volume ring like the original Echo) and the second generation (which has volume buttons on the top instead of a ring), but overall the two devices are the same.

The Echo and the Home as presented here are the most common devices mentioned in this study, and the only variant described in this project is the Echo Show, which is rectangular and has a screen. This afforded the user with the ability to see suggested uses, song lyrics, videos, video calls, lists, and much more.

Since this project was written, and since interviews were collected, Amazon and Google both announced new devices, and other companies announced competing devices, in Autumn 2017. These announcements only serve to highlight the relevance of this study and others like it at this time.

4. Human-Computer Interaction and Information Architecture

4.1. History of human-computer interaction

Human-computer interaction has developed alongside the computers themselves. Each time a new development has come to computers, there has also been a change in how humans and computers interact. The first computers used batch interface, which involved punch cards or paper describing a dataset and program for the computer to carry out, providing a series of written language commands to the machine, which took a very long time to process and respond to the commands (Butow, 2007). Next, command-line interfaces came along in the 1970s (Butow, 2007), when teleprinters entered the scene. These devices were computers whose modality of interaction was that of a typewriter and telegraph. When the usability of the teleprinter, with its quick feedback and ease of use compared to the batch interface, was supplemented with video display terminals, which allowed for easier visual representation of a program and data set, as well as a text field which could be rapidly and reversibly modified, the command-line user interface became very popular. Some standardization of this text-based interface occurred in the mid-1980s with IBM creating the Systems Application Architecture (SAA) standard (IBM, n.d.), which included regulations that a pulldown menu system should be at the top of the screen, that the status bar should be at the bottom, and that shortcut keys should function the same across interfaces. Shortly after this, the learning curve for computer use leveled out significantly. The time it took to learn and become adept at using a computer was reduced even further by the sudden popularity of graphical user interfaces, which, while in development since the 1960s, finally came to widespread use in 1984 (Butow, 2007) with Apple Macintosh's extremely successful Superbowl advertisement (and the comparatively low cost of the computer they were selling) and grew dramatically in the following years as other companies promoted competition to develop a more user-friendly system by which their customers could interact with the computers they were developing.

Human-computer interaction interfaces are not limited to graphics-based systems, however (Butow, 2007). There are also touch-based systems and voice-based systems (among others), the latter of which takes the spotlight in this project. Human-computer interaction as a field primarily focuses on the interaction with interfaces between people and computers. Below, I clarify why information architecture, while tangentially related, is fundamentally different.

4.2. Information Architecture

Rosenfeld, Morville, & Arango (2015) present four definitions for information architecture. These are as follows:

1) The structural design of shared information environments
2) The synthesis of organization, labelling, search, and navigation systems within digital, physical, and cross-channel ecosystems
3) The art and science of shaping information products and experiences to support usability, findability, and understanding
4) An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape. (p. 24)

This project takes its starting point from the third definition: information architecture as a process of shaping information products and experience to support usability, findability, and understanding. Information products here means any product that can support information seeking. Usability, according

to the Interaction Design Foundation, "is part of the broader term 'user experience' and refers to the ease of access and/or use of a product or website" and the official definition is "[…] the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (Usability, 2017). Findability means the user needs to be able to find information "through some combination of browsing, searching, and asking" (Rosenfeld, Morville, & Arango, 2015, p. 25). Finally, understanding means that the user needs to be able to understand how to find the information, what the available information includes, and how to go further with their search once they find that first bit of information.

Information architecture's focus, as stated here, is the shaping of both the product and the experience to support usability, findability, and understanding. It takes the focus of human-computer interaction and deepens it to explore the information structures that support an information environment, as opposed to looking at the interactions and interfaces in that ecosystem.

4.2.1. Information environments

In Rosenfeld et al., the authors acknowledge "that the information ecosystem landscape is richer and more complex today [than it was when earlier editions of the book were released. Many people's experience of interacting with information increasingly occurs via smartphone apps and other channels that do not involve a traditional web browser" (2015, p. xii). The authors describe current and developing information ecosystems which now include everyday items like doorknobs and thermostats which include sensors and system components which allow them to communicate with each other and with an extended system of devices such as apps on mobile devices, and the authors claim that, despite not requiring the same structures that traditional websites did and do, "[these everyday items with sensors and system components] are still key components in information ecosystems and thus subject to many of the same design principles presented in previous editions of the book" (p. xii).

They suggest considering their designs "in the abstract—as *information environments* instead of websites" in order to "see that the design principles that inform these semantic structures have broad applicability beyond design for the Web," (p. xii) and that "as long as we are dealing with the design of information environments for use by human beings, we will have a need for tools and techniques that allow us to structure that information to make it easier to find and understand" (p. xiii). In the introduction to their book, they express a hope that "even as technologies and techniques come and go, you will be able to continue drawing from the well of information architecture for many years to come," to help bring "consistency, coherence, and understandability to digital products and services" (p. xiii).

Information architecture, as presented by Rosenfeld, Morville, and Arango (2015), "allows us to think about problems through two important perspectives: that information products and services are perceived by people as *places made of information*, and that these information environments can be *organized for optimum findability and understandability*" (p. 1). In the context of this project, the information products of interest are the conversational voice-user interfaces, specifically the Amazon Echo/Alexa IPA, and the Google Home/Google Assistant IPA. These information products both provide a service to their users which is nearly identical in purpose, if not necessarily in performance. So, these places made of information (the smart speakers and the plugs, lightbulbs, thermostats, etc., that can be integrated with them to create a smart integrated home) can be "organized for optimum findability and understandability," but what does that mean? Information architects should consider the following: places made of information with distinct experiences and purposes need to be designed with coherence and consistency across channels in mind. Andrea Resmini and Luca Rosati describe consistency as the following:

> Consistency is the capability of a pervasive information architecture to serve the contexts it is designed for (internal consistency), and to preserve this logic across different media environments, and uses (external consistency) […]. Consistency needs to be designed with the

> context it is addressing clear in mind, and in respect to the several media and environments that the service or process will span. (Resmini & Rosati, 2011, p. 90)

Consistency/coherency is important especially for pervasive information architecture, and is therefore extremely relevant in the case of solving problems with these smart speakers and their information environments. If the information architect understands what problem their architecture is solving, it is easier to design coherently and consistently and avoid adding confusing or excess features which muddy an otherwise good information space.

4.2.2. Context, Content, and Users

Rosenfeld et al (2015, p. 25) describe key concepts of information architecture that I will briefly define below to provide a more concrete understanding of the discussion central to this paper.

- **Information:** information exists in the space between pure data (numbers, names, facts and figures) and knowledge (what people know)—this can be something such as a document, an image, or a website, as well as the metadata which describes and represents these content objects.
- **Structuring, organizing, and labeling:** structuring involves determining the size and relation of different information chunks; organizing involves "grouping these components into meaningful and distinctive categories, creating the right contexts for users to understand the environment they are in and what they're looking at;" and labeling involves finding the right descriptive language and navigation structures for these categories.
- **Finding and managing:** users must be able to find what they need in the information environment, or the system fails; in order to maintain findability of information, content must be managed efficiently and there must be clear policies and procedures in place in the business for managing the information.
- **Art and science:** information architecture relies on the disciplines and methodologies backed by the rigor of the scientific method to understand users' needs and information-seeking behaviors, but the study of usage is affected by the complexity and ambiguity of the human factor of the users, designers, and others involved in information—therefore, "information architects must rely on experience, intuition and creativity. We must be willing to take risks and trust our intuition. This is the "art" of information architecture."

Information must be structured, organized, and labelled to support it being found by the users (and therefore be useful)—this process is supported by effective management of these information structures, organization, and labels, and the whole system is supported by the combination of careful research into the systems and users and by the artist's intuition and a willingness to take some risk and be creative. This abstract outline of the concepts behind an information architecture highlights the complex dependencies and relations that exist in an information environment.

These information systems are not standalone; they do not exist in a sterile, unchanging environment. Instead, they are affected by the intended audience, the users of the information contained therein; they are affected by the content they are intended to contain and make available; and they are affected by the context they are created for. Rosenfeld et al provide a useful Venn diagram to help illustrate these interdependent relations, seen in Figure 3.

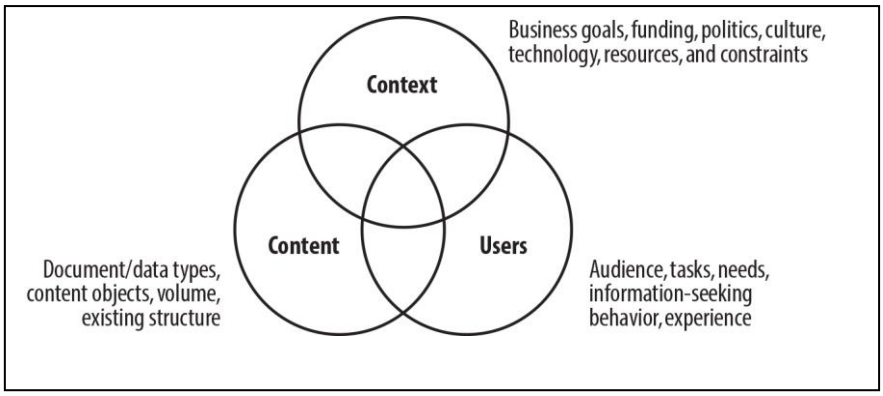


Figure 3 The three interdependent components of an information architecture.

Where context, content, and users intersect is the information space, and their intersection provides clues for how to build an information architecture. So, what does that look like? Rosenfeld et al provide some answers. First, it is important to determine whether the information architecture is/should be built as a top-down information architecture or a bottom-up information architecture. Top-down information architecture is defined by the fact that the designers of the information space have decided, based on what they think the user needs from the space (users), what the space should include (content), and how users should or want to navigate through the space (context). Bottom-up information architecture is "suggested by and inherent in the system's content" (Rosenfeld, Morville, & Arango, 2015, p. 86). Rosenfeld et al suggest that bottom-up information architecture design is ideal for anticipating users' desire to "jump to other relevant content on your site without learning how to use its top-down structure" (2015, p. 86). From this definition, a bottom-up design of information architecture could be the best approach for the design of information architecture for smart speakers, as a conversation will be more actively built by the user than most parts of information architecture in, for example, mobile apps. This will be an idea I return to later.

5. Literature review

As has been established, these smart speakers are relatively new type of technology to be introduced to the user's home information environment. As such, the amount of research on these specific devices (smart speakers) is less than on, say, chat bots or smart phones (which present a similar conversational assistant without necessarily having the same position in or effect on the user's home information ecosystem). In this section, I present my journey through the existing world of research on these smart speakers, as well as some of the tangential concepts that came to the fore as I dove into the pool of papers.

5.1. Initial steps

To determine the scope of the available literature, Scopus is a valuable tool. Understanding these devices and their users requires an understanding of what research has already been done on these devices. The first step was to search "Google Home" and "Amazon Echo" on Scopus.

Searching "Amazon Echo" on Scopus in October 2017, with no other quantifiers, gives 16 document results. Of these, one is from 2018, eleven are from 2017, three are from 2016, and one is from 2015. The subject areas covered by these documents are as follows: computer science accounts for 12 documents (75%); engineering accounts for 7 (43.8%); health professions, 3 (18.8%); social sciences, 3 (18.8%); medicine, 2 (12.5%); and materials science, 1 (6.3%). The question being addressed in this paper concerns constraints users face in finding and managing information, and how IA designers can use this information to improve the user experience. Therefore, with such a small pool of results, it was worth the

time it took to read through the abstracts and the keywords to extract the most relevant documents to this project.

After removing documents which fell outside the field of relevance to this project (for example, those that deal with developing the software or hardware), the following documents remained: Purington, et al., ("Alexa is my new BFF", 2017), which explores the links between personification of the Amazon Echo in reviews on Amazon.com and reported user satisfaction; Porcheron, et al., (Talking with Conversational Agents in Collaborative Action, 2017), which explores the collaborative aspect and social implications of using these smart speakers, but is a description of a workshop rather than a true research paper. These two documents each provide useful insight into how to refer to the devices, but most importantly two terms come forward from these papers that warrant further investigation: "Intelligent Personal Assistant," and "conversational agent." Before following that path, the same process should be followed for "Google Home."

Applying the same process for the "Google Home" Scopus search requires that the search is restricted to just results from 2017 because the device itself was released in 2016, meaning that no documents about using the device would be produced before then. After 2016, eight documents were published (seven in 2017 and one in 2018). The subject areas break down as follows: computer science accounts for 6 documents (75%); engineering, 3 (37.5%); social sciences, 2 (25%); and medicine, psychology, and health professions, 1 (12.5%) each. Again, with so few documents to sort through, it is worth the time it takes to read through the abstracts and pull out some specific papers.

After removing the documents which are inappropriate for this project, the following documents remain: Druga, Breazeal, Williams, and Resnick ("Hey Google is it ok if I eat you?" Initial explorations in child-agent interaction, 2017), a paper about child-agent interactions (they call the devices "agents" in their paper) and future design considerations around voice, interactive engagement, and facilitating understanding; and Martin (How echo, Google home, and other voice assistants can change the game for content creators, 2017) an article about the possibilities for content creators with these new platforms and building for conversation. While these papers will bring an interesting filter to see the results of this study through, neither of these papers yielded specific words for these devices the way the search for "Amazon Echo" did.

In the next section, I will expand a little on the difference between the provided terms for these devices (Intelligent Personal Assistant, conversational agent), and why I decided to use the term "smart speakers" instead of either of those terms. Despite not using either of the two terms brought forward by the initial search, looking more closely at their meanings and the scope of research around the terms helps to focus the scope of this project by defining it against what it is not.

5.2. Conversational Agent vs. Intelligent Personal Assistant

Conversational agents are classified by Lester, Branting, and Mott (2004) as follows: "Conversational agents integrate computational linguistics techniques with the communication channel of the Web to interpret and respond to statements made by users in ordinary natural language" (p. 1). The chapter they contributed to the *Practical Handbook of Internet Computing* was written in 2004, and at that point, the primary application of these conversational agents was as customer service, help desk, website navigation, guided selling, and technical support—task-oriented or information-seeking enterprise-centered applications. Since that time, several mobile phone producers have developed conversational agents capable of voice interaction (e.g. Apple's Siri, Google/Android's Google Assistant, and Samsung's Bixby), Microsoft has released Cortana to operate with Windows desktops, laptops and phones, the Xbox, and Android, and now Google and Amazon have their smart speaker products on the market. A brief survey of results for "conversational agents" on Scopus follows.

Between a search on Scopus done in June of 2017 and a search done in October of 2017, the number of results for “conversational agent” increased from 1,576 documents (June) to 1,625 documents (October). Below, a chart of the document results with the keyword “conversational agent” by year can be seen.

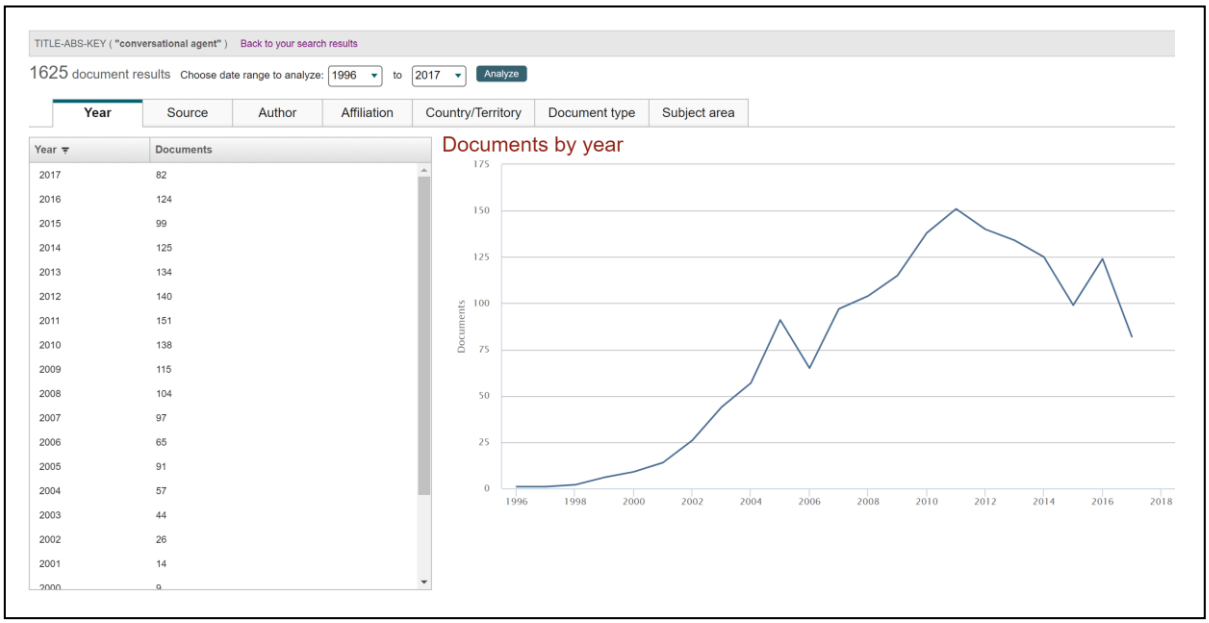


Figure 4 Scopus results for search TITLE-ABS-KEY ("conversational agent") by year.

Clearly, this topic is still of interest to research, but it seems to be either hovering in popularity as devices are released with this conversational agent feature or declining in popularity as these technologies continue to change and the references to them change as well.

Adjusting the filters of the data further and acknowledging the nuance in term use, I also looked at the numbers for the query “conversational agents” on Scopus only in 2017, with the term “embodied conversational agents” removed. Embodied conversational agents are agents that interact with the environment through a physical body (including a graphical body on a screen), and this term is therefore not applicable to these smart speakers and their functionality. This adjustment brings the results to fewer than 70 documents. Reading the abstracts and titles on the first page of results helps narrow the search further: restricting the search to the terms “intelligent agents,” “intelligent assistants,” and “intelligent personal assistants” results in just five documents.

The first article, “Incorporating android conversational agents in m-learning apps” (Griol, Molina, & Callejas, 2017) deals with incorporating the conversational framework of the intelligent assistant in Android phones with the possibilities offered by these devices for mobile education. Their result combined a rich and engaging mobile learning experience with a multimodal (voice or tactile) input and output that allowed for functionally diverse students to access the app. One of the benefits they cited in using the Android phones to implement their educative app was that they could integrate the Google Speech application programming interfaces to emphasize interaction management and build sophisticated, robust, and manageable applications. While this article does not specifically talk about an intelligent personal agent, they use the Google Assistant as the baseline for their conversational educative agent, and that is the same intelligent personal assistant used in the Google Home, explored in this project. Therefore, it can be understood that the natural language processing of the Assistant could be used to help supply the user with a rich and engaging interactive experience.

The second article dealt with using embodied intelligent conversational agents and turning them into autonomous agents and can be interacted with (Bonsch, Vierjahn, Shapiro, & Kuhlen, 2017). The focus is

on modeling virtual agents, rather than on the way these agents can help the user with searching for information. Therefore, I will not go into any more details on this paper.

The third article was a description of an upcoming workshop to "bring together academics and industry practitioners to explore collaborative challenges in speech interaction" (Porcheron M. , et al., 2017). Their focus is on exploring the challenges of speech-enabled, multi-user devices such as the Google Home and Amazon Echo in collaborating with devices in social interactions. While there is no result from this workshop published at the time of this project, any results from this workshop would inform the area of research to which this project will contribute.

The fourth article dealt with conversational agents, which the authors also call intelligent personal assistants, in a conversation among a group of friends in a café (Porcheron, Fischer, & Sharples, 2017). Their focus was the agent as an assistant, so for argument's sake, the focus in this paper is the intelligent personal assistant, and they used the examples of Siri (Apple's assistant on the iPhone), Cortana (Microsoft's assistant), and Google Now (the precursor to Google Assistant). Their hope was to "provide an understanding that contributes to the design of IPAs" (p. 209), and thus falls into the same category of research as this project. This article focuses more on multi-party conversation than on individual expectations and experiences, as this project does, and on the interactions with the assistant as it exists in a mobile phone, as opposed to interactions with intelligent personal assistants as they exist in a speaker that is stationary in the home (again, as this project does). However, their findings might lend some light to the results from this thesis. They found that interactional and technological problems are accommodated in the following ways:

- Repeating and refining: in case of failure of the assistant, users will repeat or refine their query within a few seconds. They state that the design solution to this is the intelligent personal assistant (IPA) providing the user with more meaningful feedback such as helping the user in finding the right words or refining the query. They cite inspiration for future design considerations in auto-completion of searches such as what currently happens on Google searches "in order to support query formulation and refinement without the need for members to *recall* or reason about terms which would be more likely to result in a successful query" (p. 215). They also state that allowing repair of a query (saying, for example, "oh no, I meant…") would help alleviate misunderstandings and result in more correct responses from the IPA. Last in this section, they add that "contextual relevance could be gathered from prior failed queries, as a utility to both improve accuracy in understand interlocutor's intent during successive queries, albeit at the potential expense of privacy" (p. 216).
- IPAs as humanlike conversational partners: users were more likely to treat the assistant as humanlike if they were in private (citing Luger and Sellen's (2016) results from interviewing private users of IPAs on phones), but that they still treat the IPA as a machine and not human. The design suggestion in this result was that a better understanding of the nature of talk with IPAs could help address "nuanced interactional troubles" (p. 216).
- IPA use in multi-party conversation: speaking to an IPA makes interaction with the device (in this case, a mobile phone which is normally a single-user device) available to all members in the setting, and the democratic nature of this interaction with a voice-based interface could support collaboration. This result is not as helpful for this thesis, but still interesting to note.

This paper revealed some interesting design suggestions for the future of IPAs, and more importantly, design suggestions that would be relevant for IPAs in smart speakers as well as IPAs in other contexts, thereby making it relevant for this thesis.

The fifth article explored what conversational search would look like with an intelligent assistant (Vtyurina, Savenkov, Agichtein, & Clarke, 2017). The authors compared users' experience with an information search on a text-based chat interface with three conversational agents ("a commercial

system, a human expert, and a perceived experimental automatic system, backed by a human 'wizard' behind the curtain" (2017, p. 2187)). One of the key similarities between Vtyurina et al. and this project is that their results help provide understanding of what users' hopes are and what the limitations are of existing conversational agents to guide future work to improve conversational agents/intelligent assistants for search scenarios. It is not unlikely that this project will also result in providing an understanding of user hopes for and limitations of existing intelligent assistants. A direction for future research they cite is "investigating the possibilities for improving existing conversational agents" (Vtyurina, Savenkov, Agichtein, & Clarke, 2017, p. 2192), and that is the direction I hope this thesis takes. The design suggestions that come from their data are as follows:

- Context: allowing the user to give short questions and comments as opposed to long sentences
- Provide sources of answers: trusting the sources improves credibility of the assistant
- Use feedback: allowing the user to provide feedback could help improve the system and the results
- Opinion aggregation: hearing a summary of other people's experiences was a wish of participants in their study
- Direct answers vs. expanded information: users were split into 2 camps, one of which wanted direct answers to their query (complained that the answers were too long) and the other who wanted a broader context (wanted more information than just an answer to the immediate question).

Their conclusion was that people were happy to use automatic systems for complex information seeking tasks "as long as their expectations about accuracy were met" (2017, p. 2192). I hope this holds true for the participants in this study.

One of the most relevant works was not found in the Scopus literature search but was instead cited in Vtyurina et al (2017) and Porcheron, Fischer, and Sharples (2017). This article is from Luger and Sellen ('Like having a really bad PA', 2016), and it focused on interviews of 14 participants regarding their daily life experience with an intelligent assistant that they use regularly. Luger and Sellen's work largely inspired the structure of this thesis. They reported on the experiences and expectations of users, and they discussed successes and failures of conversational agents from the user reports. One of the major findings of their work was that users would most frequently use relatively simple tasks (e.g. weather, checking for reminders, directions).

Another work that provided inspiration to this project is Bødker and Christiansen's (2012) work exploring "the new opportunities brought to use by App technologies available on mobile devices" (p. 78). Their approach used data "from interviews with 12 iPhone users triangulated with models of appropriation, theories of micro and macro level appropriation, and the concept of 'expansive learning'" (p. 78) to provide a new understanding of how users make the iPhone their own, focusing on how the phone and apps themselves help or prevent the device becoming a personal gateway for the user to access more apps. They show that a major part of users appropriating the device is its relevance in social contexts (e.g., discussing it on Facebook, at dinner, building identity with others around owning an iPhone), which to some extent rules out the application to the smart speakers, as those are relatively immobile, stuck in one place, and the social network to build an identity around owning one is still too small and widely spread geographically (for the most part) to contribute to the maturation and advancement of use patterns. Future research they predicted was a study of "what mechanisms interaction designers may employ in supporting directly and indirectly both appropriation and expansive learning through design" (Bødker & Christiansen, 2012, p. 84). This project contributes to the same field, in that it applies to the design of how users approach an IPA, though the focus of information architecture design is the information architecture, and the focus of interaction design is the design of elements of the user-machine interaction. Designers in these fields may answer many of the same questions and problems, but

this project focuses on the information architecture design of smart speakers, following Rosenfeld, Morville, and Arango's (2015) framework as laid out above.

5.3. Smart Speaker

Choosing to use the word "smart speaker" limits the scope of the reference to just those intelligent personal assistants or conversational agents which function from a speaker which sits in the environment. Until Autumn 2017, Google and Amazon were the only two with popular, functional, affordable tech on the market. Therefore, there are more users of those two devices to interview about their actual, lived experiences interacting with their device. The goal is not to talk about the conversational agents or intelligent personal assistants which function from mobile phones or computers, or other devices. Therefore, while these devices can be talked about as intelligent personal assistants or conversational agents, it is important to the central question of this project that the function of these smart speakers and their existence in the user's home is the focus.

5.4. Older literature

The fields of artificial intelligence, machine learning, and conversational interaction are by no means new. Therefore, it is useful to briefly supplement newer literature on these IPAs with older literature on their foundation.

Joseph Weizenbaum was a computer scientist and professor emeritus at MIT, and his contributions to the field of artificial intelligence have given him a place among the fathers of artificial intelligence. In the early part of the 1960s, he developed a program called ELIZA that could use natural language to interact with a user. This program operated under the concepts of Rogerian psychotherapy, which functions by means of asking the client (the user) reflexive questions about the user's utterances, making the user feel understood and responded to. Interaction with ELIZA consisted basically of a user and the program, and a typewriter to relay utterances between the user and the program. Weizenbaum discovered that the program performed best "when its human correspondent is initially instructed to 'talk' to it, via the typewriter of course, just as one would to a psychiatrist" (Weizenbaum, 1966, p. 42). He was "startled to see" the emotional connections the human users made with the program and their anthropomorphizing of it (Weizenbaum, 1976) and that the users would go so far as to make (perhaps generous) assumptions about their conversational partner, and they would "contribute much to clothe ELIZA's responses in vestments of plausibility" (Weizenbaum, 1966, p. 42). In the context of this project, it could be fair to assume that users who "talk" with their smart speaker could make an emotional connection with their device (or the IPA) and that their anthropomorphizing of the IPA would help ease frustrations with the certain failures of the IPA.

Herbert H. Clark, a psycholinguist at Stanford University, focuses on cognitive and social processes in language use, interactive processes in conversation, and word meaning and use. He developed and is well-known for his theory of "common ground": individuals engaged in conversation must share knowledge in order to be understood and have a meaningful conversation (Clark, 1985). He and Wilkes-Gibbs (1986) take the theory of common ground further, establishing that common ground must be mutually established by both parties before moving on to the next utterance in a conversation. This establishment of common ground can be iterative, but users want to minimize the effort it takes to collaboratively create common ground. Additionally, Clark and Isaacs (1987) worked with the common ground theory in terms of mismatched expertise between two human conversationalists, and they determined that with two partners, they would "accommodate to each other's expertise […] by following a basic strategy: (a) Begin assuming only as much expertise as you think might be shared by your partner; and (b) use your partner's responses to adjust to his or her actual expertise" (Isaacs & Clark, 1987, p. 35). They found that "people accommodate to each other quickly and automatically in the very process of making themselves understood" (Isaacs & Clark, 1987, p. 36). This could be applied to the smart speakers and IPA in that the IPA must collaboratively build shared knowledge and grounds for understanding, and

this can be done iteratively by both the user and the IPA adjusting their interactions to better suit the other's expertise.

Terry Winograd is a professor of computer of science at Stanford University who has been instrumental in work surrounding artificial intelligence and natural language processing. Together with Fernando Flores, a Chilean philosopher, he wrote a critical appraisal of artificial intelligence and cognitivism (1986), which provides discussion on use of computers, especially in determining whether it is appropriate to use them at all for certain tasks. They discuss, especially in chapter 12, how to design "computer-based systems to facilitate human work and interaction" (1986, p. 163). They also discuss conversation, in terms of the structure it provides: "A speaker and hearer do not apply 'conversation pattern rules' any more than they apply 'perception rules' or 'deduction rules.' [...] [A conversation's] structure becomes visible only when there is some kind of breakdown" (p. 68). This lack of structural visibility is reflected in information architecture as well: it is difficult to see a "good" information architecture until there is a breakdown. However, to avoid breakdowns, they state, "Successful system builders learn to consider the user's domain of understanding after seeing the frustrations of people who use their programs" (p. 165). This matches with Isaac & Clark's discussion of matching expertise between conversational partners and applies it to designing a computer. In this project, we can apply that to designing an information architecture to help the IPA and user match their expertise. Furthermore, in a book of interviews with interaction designers, Winograd mentions Mark Weiser's notion of "ubiquitous computing" (Moggridge, 2006), an idea that users do not want to interact with computers for the sake of interacting with computers, but they interact with computers to get something done, and the computer should become the tool. Conversational IPAs and smart speakers in the home help make this possible, but the design should allow inclusion of the user and of flexibility to ward against breakdowns.

Lucy Suchman is a professor in the sociology department at Lancaster University with interests in the field of feminist science and technology studies. In a paper directed at Winograd and Flores' book mentioned above, she directs the reader to concerns she has about who the systems serve, and encourages designers to consider and include the user in the creation process to avoid a myopic worldview informing an entire system (and thereby locking certain users out who do not operate within that worldview) (Suchman, Speech acts and voices: Response to Winograd et al., 1995). Much of her other work (e.g., (Suchman, Plans and Situated Actions: The Problem of Human-Machine Communication, 1987)) encourages designers to consider the user's real actions, rather than static ideas of plans for action, when designing. This idea is carried forward in the decision to study real users in this project.

6. Methods

6.1. Approach

The end goal of this project is to explore the experience of users of smart speakers regarding finding and managing information, and from that data, provide suggestions for what an information architecture designer might do to improve user experience. This is approached through an exploratory qualitative case study of a small number of smart speaker users conducted via semi-structured interviews. Qualitative research methods fit perfectly with this project:

> If a researcher is interested in a topic on which little or no research has been done in the past, quantitative research may be difficult to employ, because there is little prior literature from which to draw leads. A more exploratory stance may be preferable, and therefore qualitative research may serve the researcher's needs better, since it is typically associated with the generation rather than the testing of theory […] and with a relatively unstructured approach to the research process. (Bryman, 2016, p. 36)

I have established that very little, if any, research has been done on the information architecture of these devices. An exploratory approach to qualitative research works well in this project for this reason, and

will allow me to generate suggestions for design of information architecture for smart speakers which will improve user experience. Furthermore, qualitative research allows for a relatively unstructured approach to the research process, allowing me more freedom and flexibility in my approach to data gathering with the users.

Choosing semi-structured interview for data collection ensures that Suchman's (1995) emphasis on the inclusion of user perspective in design remains a focus. A form of qualitative interviewing, this approach allows for open-ended conversations with users that can evolve both within the individual interview and between interviews. Semi-structured interviews are ideal for research with a fairly clear focus because they allow the researcher to address specific issues related to the research question, and "the emphasis must be on how the interviewee frames and understands issues and events—that is, what the interviewee views as important in explaining and understanding events, patterns, and forms of behavior" (Bryman, 2016, p. 468).

One of the distinguishing features of a semi-structured interview is the interview guide, "a list of questions or fairly specific topics to be covered […], but the interviewee has a great deal of leeway in how to reply" (Bryman, 2016, p. 468). The researcher also has a great deal of leeway in the order, wording, and inclusion of questions in each interview, though for the most part, all questions in the interview guide will be asked and a similar wording will be used in each interview. In creating the interview guide, it can be helpful to ask, "Just what about this thing is puzzling me?" (Lofland & Lofland, 1995, p. 78). Questions may be generated by random thoughts in different contexts; or they may arise during discussion with colleagues, friends and relatives; or they may come from the existing literature on the topic.

The cycle of formulating questions for an interview guide in Bryman is relatively linear, proceeding from research area to research question, then to a cycle of revising the interview questions, followed by testing the guide (or doing a pilot study), then revising the interview questions, then finalizing the guide (2016, p. 470). My process was much less linear, as the questions and the list itself evolved as I progressed through user interviews, and while I did read through the interview guide with a few colleagues/fellow students before starting interviews, I primarily used my first interviews as pilots for the questions, and revised in an ongoing process even until the final interview. The interview guide is discussed below in its final form.

Bryman (2016) also includes that the interviewer should record 'factsheet' information, such as name, age, etc., in order to contextualize the responses. I did record a limited amount of information in the interview scheme which I brought to every interview with the confidentiality agreement and interview guide, though it was originally intended to be a cover page for the interview notes just to help me remember who each interviewee was and to serve as an additional page for notes if necessary. Therefore, it lacks some information such as age, gender, occupation, etc., but other information which could have been included in this page are covered on the interview guide itself, such as how long ago the participants first heard of the devices, when they bought their device, how many they own, and where the devices are located.

6.2. Materials

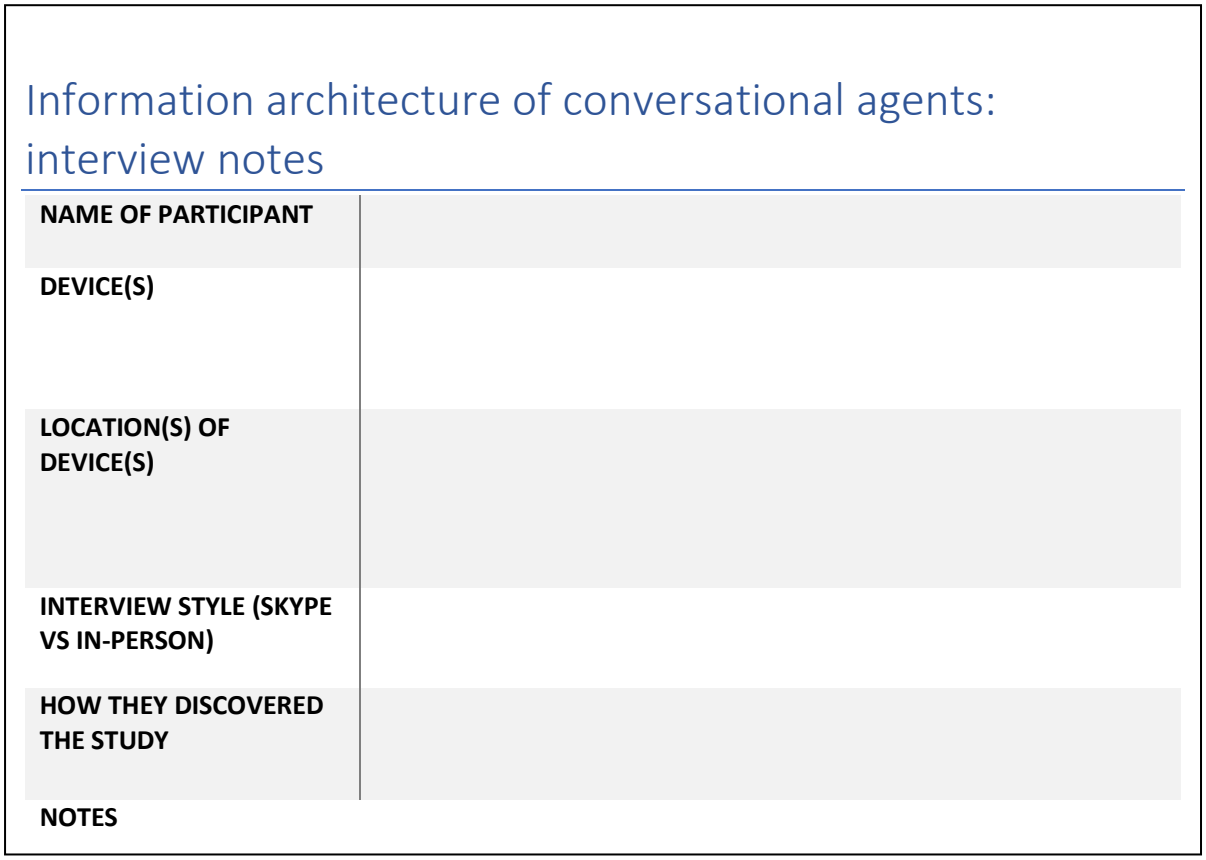
The materials used for the interviews included three documents, my cell phone, and a pen. The first document, the confidentiality agreement, is a standard document, and is therefore only included in the Appendix because it does not need my comments.

6.2.1. Interview notes page

The second document, the interview notes page, is shown in Figure 5. This document was important to the organization of the interview papers, as I occasionally had more than one or two interviews each day, and in the process of transiting between interviews, there might have been little to no time to organize

each interview's papers. Therefore, this document was vital in keeping track of which interview guide belonged to which consent form. However, I reiterate that this notes page was more for my own personal organization than for any tracking of vital user statistics.

Having these items on the interview notes page ensured that I always wrote this information down at the beginning of the interview and was not forgotten in the process of the interview. Knowing which device or devices the users owned, and the location of each device, was important for understanding the user's narrative of how they used their devices. Knowing how they discovered the study allowed me to know, to some extent, which method of contacting users had worked most effectively (more on that below). Since the page itself was larger than the table for this information, I left the rest of the page blank for any notes that may need to overflow from the interview guide page.



Information architecture of conversational agents: interview notes

NAME OF PARTICIPANT	
DEVICE(S)	
LOCATION(S) OF DEVICE(S)	
INTERVIEW STYLE (SKYPE VS IN-PERSON)	
HOW THEY DISCOVERED THE STUDY	
NOTES	

Figure 5. Interview notes page—useful for keeping track of which consent form belonged to which interview guide, and as an extra notes page.

6.2.2. Interview guide

The third document, the interview guide itself, is shown in Figure 6. In preparation for the interviews, I created and read through this interview guide with a few family members and friends who are familiar with smart speakers to ensure that, as much as possible, the questions would be comprehensible and relevant to the interviewees. I was curious about the user's experience from the first time they heard about the devices until the time of the interview, and therefore some of the first questions relate to time before the user owned a device. For some users, these initial questions were irrelevant. For example, if they did not purchase the device, the questions about motivation to buy the device, or when they purchased it, did not apply or needed to be altered—in some of the cases where the device was a gift, for example, the question about motivation could be altered to, "What would have motivated you to buy a device?", which, while being a fundamentally different question, still access information about what how the user interprets their device.

Understanding that the primary use of the device might change over time, and that how the user interprets the device in their life would change with use, motivated some of the later questions. Building a narrative around how the users interacted with their devices and how that has changed over time serves the purpose of helping me as the researcher to understand the user's perspective and thereby better understand their relationship with the device and contextualize their responses. This document was printed out for each interview to allow me to take handwritten notes during the interviews, and I often ran out of space on this page. However, since it was difficult to know where the space would be needed, I simply recorded as best I could in the margins and on the back of the page, overflowing onto the interview notes page, and/or onto notebook paper, if necessary.

The list of questions evolved over time. After each interview, I felt that I had changed how I asked the questions, or new questions had come up which seemed to be relevant enough to the research question to warrant being officially added to the interview guide. The latter case applies to the items on the list from "Characterization of interaction with device" and down to "Thoughts on voice based vs screen based interactivity?" It is clear that this interview guide was a combination of specific questions I needed answers to as well as concepts that should be addressed in the later add-ons to the guide, as the formulation of questions disintegrates and the list becomes concepts with question marks. For me, this last area was more of a checklist—many of these items came up naturally during the interviews.

Interview guide for smart speaker users

Which device(s) do you own?

When was the first time you heard of these devices?

What was your impression then?

What motivated you to buy a device? Why did you choose the device you chose?

What did you consider when buying a device?

How long ago did you purchase your device?

Have you been using it since you purchased it?

- If not, why not? What changed to make you start using it again?
- (if not) How long have you been using it?

What were some of the first things you did with your device (after setup)?

Has your understanding of these devices changed since you started using yours?

Has your usage changed since using the device?

What do you do with your device? How do you use it?

How often do you use it? Multiple times daily, daily, weekly, etc.

How did you discover uses for your device?

How did you learn to use your device? (e.g. tutorials, articles, the device helping you, etc.)

Characterization of interaction with device

How does it fit into your home?

How does it feel to use it?

Frustrations?

Skills/add-ons?

Future use?

Suggesting uses:

- -Device-focused interaction
- -Non-device focused interaction

Thoughts on voice based vs screen based interactivity?

Figure 6 Interview guide--evolved over time to formally include questions which arose in earlier interviews.

Having the interview guide was helpful in maintaining my focus during the interview, as well as helping to ground me when conversation with the interviewee wandered off the path of research-relevant information. Often, I would note new questions from the previous interview(s) in pen at the top of this paper if I did not have time between interviews to edit the document itself.

The app I used to record interviews is called Smart Recorder (SmartMob), and is a relatively simple interface for recording voice on a mobile phone. These recordings were uploaded after the interview to my OneDrive account in order to avoid loss of data if my phone or the app malfunctioned.

6.3. Participant recruitment

I conducted interviews while I was visiting my family in the USA over the summer. Neither the Amazon Echo nor the Google Home are activated for use in Denmark, therefore I lacked access to users to interview in Denmark. My family's influence and connections in their community and their network allowed me access to users' homes that I am aware I would be unlikely to be granted otherwise.

Access to participants was achieved through a combination of resources. With the design help of a friend, I was able to create an advertisement for participants, which I released on Facebook, through email, via word-of-mouth and physical door-to-door distribution, and through an app called NextDoor (like Facebook, but for neighborhood-relevant posts such as lost pet announcements or garage sale event notifications). I also received help from my parents and neighbors: my mother granted me access to her Facebook account, where I posted my advertisement to her friends; my father emailed coworkers of his that he knew owned a smart speaker; my neighbor's wife reached out to her friends and colleagues who she knew owned a smart speaker; and my mother's account on NextDoor allowed me access to people in the neighborhoods and suburbs in driving distance from my parents' house who trusted members of the NextDoor app (for example, more than they would trust someone posting on Craigslist).

The advertisement is shown in Figure 7, an advertisement for participants detailing, in colloquial terms, when, where, and why I was conducting interviews, what I wanted to know, and ensuring the users' data would be private. The section at the bottom of the advertisement provides some details about me as the researcher, lending credibility to my call for participants for research purposes by my association with a university. I also felt it was important to include pictures of the devices I was asking about to avoid confusion, and a picture of myself to help users who might invite me into their homes by allowing them to know who to expect at the door.

Figure 7 Call for participants

6.4. Data collection: Contextual Inquiry and qualitative interviews

Contextual inquiry is a style of ethnographic field study used by information architects to learn about user behavior in a specific context. As stated by Rosenfeld, Morville, and Arango, "[Y]ou can only learn so much about the bald eagle or the bottle-nosed dolphin by studying them in a lab" (2015, p. 340). Contextual inquiry, as described by Rosenfeld et al, can be applied somewhat flexibly, depending on what is appropriate to the situation.

Some of the advantages of contextual inquiry include: seeing the workspaces of users to understand "the spectrum of information resources they use on a daily basis (e.g., computer, phone, bulletin board, Post-it notes)" (Rosenfeld, Morville, & Arango, 2015, p. 340); watching people interact with a product during the normal course of their day; and occasionally, watching people do their work in the space. Restrictions also present themselves in terms of using contextual design. First, people might use their device sporadically, and use during an interview might be "forced." Second, being present for either observation or interview in someone's workspace (as contextual inquiry is designed for) is very different from coming into someone's personal space, in other words, into their home, for the same purpose. The interviewer must take precautions to seem as inoffensive and inconspicuous as possible, while understanding that an interview to some extent puts the interviewer and interviewee into a distinct power dynamic as well as into a spotlight in what is normally a somewhat neutral space devoid of that specific type of hierarchy.

The interviews were originally intended to take place in users' homes exclusively. However, at the beginning of the participant recruitment, I worried about having enough users and about having enough usable data. Therefore, I expanded my participant pool to include users who were willing to Skype with me (that is, anyone not living within approximately a 30-minute drive from my parents' home--mostly out-of-state participants). Skype allowed me to see the user, their device(s), and their home, for the most part, though there were limitations. The dynamic of inviting someone into your home versus that of having a corner of your house you are willing to show them over Skype presents different challenges, though I think it was not a major drawback to allow interviews via Skype.

6.4.1. Sidebar on card sorting and why I avoided it

One minor limitation was that I had originally been toying with the idea of using a version of card sorting to provide some insight into users' mental models, "illuminating the ways they often tacitly group, sort, and label tasks and content in their own heads" (Rosenfeld, Morville, & Arango, 2015, p. 344). I had actually tossed that idea before planning the interviews because the application of the test primarily seems to apply well to spaces where navigation of complexly layered information spaces is necessary. The information and interaction available via these devices is more one-dimensional and simpler, unless one decides to go deep on one's own and search for new uses in the apps or on the internet. However, if I had decided to go ahead with card sorting, Skype would have presented a significant barrier for me as a researcher. I would not be in the context with the user to see how they sorted the cards, it would be more difficult with the timing of asking questions about why a user sorted the cards the way they did (because sometimes there are time delays with Skype and other video chat systems), and having used card sorting before, I often find it is easier for me to understand and explain what the participant is doing when I have been in the same room and seen the cards sorted—the process of sorting is just as important as the final result.

6.4.2. Interview process

Two types of interviews were conducted: in-home interviews, where I entered and interviewed the user in their home; and Skype or telephone interviews, where I interviewed the user over Skype or over the phone, if they lived too far away to interview in-person.

Interviews in the home were conducted as follows:

1) Host invited me into their home, perhaps offering a drink (not required, but always appreciated)
2) We talk a little about what I am doing and why I am doing it as I take out my folder containing consent forms, the interview notes paper, and the interview guide, as well as completed interview notes. We pause while they sign the consent form.
3) Once the participant signs, I start the recording.

4) I start by asking which device they have, why they have it there, and when they first heard of it. Depending on the user's response to questions, I either proceed directly through the questions on the guide, or I follow their lead and ask questions inspired by their tangent.
5) Before ending the interview, I have the user take photos of the devices they own either with their own phone's camera or with mine. If photos were taken on their phone, I had them email me the photos.
6) The interview either ends at a natural stopping point (the end of the conversation or the end of the guide), or at a certain time if the user has appointments to keep.
7) I thank the user for their time, providing the option of compensation for their participation (a $5 Starbucks gift card), and leave.

While conducting interviews, I attempted to maintain an awareness of the participant's time—the stated limit on my call for participants was 30-60 minutes, so I tried to stay below 60 minutes unless the participant requested otherwise.

6.5. Data storage and analysis

Three types of data were stored for each interview: the voice recording, the photos of the devices, and the printed pages. The voice recordings were uploaded as they occurred to my OneDrive cloud storage to avoid any loss of data if something happened to my mobile phone (since the app does not appear to provide its own cloud backup option). The photographs were stored in Google Photos as they are captured, or downloaded from emails. The printed pages (and any other papers with notes on them related to specific interviews) were stored in a ring binder, and each interview was labelled with the participant number and the participant's initials.

Data was processed by means of thematic analysis, an analysis method used "for identifying, analysing, and reporting patterns (themes) within data" (Braun & Clarke, 2006, p. 79). Braun and Clarke's paper is useful for defining thematic analysis as opposed to other forms of analysis used on interview data such as grounded theory and thematic discourse analysis. They emphasize that the flexibility of this style of analysis is one of its strengths, allowing the researcher to identify themes that are relevant to and important to the data in relation to the research question, "and represents some level of patterned response or meaning within the data set," (2006, p. 82) while allowing the researcher to determine whether an item in the data is a theme or not based on their own understanding of the data, rather than using a hard and fast rule for deciding what is a theme or not.

Using thematic analysis, interview notes were read through and, in some instances, were listened to again, to establish an initial understanding of the data. As themes emerged in each interview, or were repeated but specified, I used FreeMind (Müller, et al.) to create a map of themes from the first pass of the information. This map was revised and refined to only results which related to the research question, addressing what users' experience was with respect to finding and managing information and what information architecture designers do to improve user experience.

7. Interviews

The purpose of the interviews was to achieve first-hand stories from users about their experiences with smart speakers in order to provide suggestions for information architecture designers to improve user experience. In this section, the results of the interviews are presented.

7.1. Participant summary

In total, nineteen interviews were recorded. The participants' profiles are included here. First, I present a table with a short introduction to the participants (see Table 1). This is helpful in terms of orientation before the participant interview summaries. The table includes information about the identifier of the

participant (P1-KC is the participant number in order of participation, and the initials of the interviewee), which smart speaker(s) they own, if they are related to other participants, what their gender is, the reason they own a device, and the type of interview they participated in.

Table 1 Participant table

Participant	Smart speaker(s)	Relation to other participants?	Gender	Reason for owning device	Type of interview
P1-KC	Amazon Echo	Husband to P2	M	Trying it out for a family member, wanted to have something unique	In person
P2-LC	Amazon Echo	Wife to P1	F	Trying it out for a family member	In person
P3-PR	Amazon Echo	None	F	Tech enthusiast husband bought it	Skype
P4-PC	Amazon Echo	None	F	Tech enthusiast husband bought it	In person
P5-JC	Amazon Echo	None	M	Tech enthusiast, bought it on Prime offer after announcement	In person
P6-BB	Amazon Echo	None	M	Amazon Prime member, fun gadget	In person
P7-PL	Amazon Echo	None	F	Gift from tech enthusiast husband	Skype
P8-RD	Amazon Echo & Google Home	None	M	Received Echo as a gift, bought Google Home as an improvement on the Echo as a Christmas gift for himself	Skype
P9-PS	Google Home	Wife to P10	F	Received it on 11th of July 2017, "Prime Day"—son suggested it	In person
P10-DS	Google Home	Husband to P9	M	Received it on 11th of July 2017, "Prime Day"—son suggested it, it integrates with tech they have, and could buy one immediately at Best Buy	In person
P11-SO	Amazon Echo	None	F	Tech enthusiast husband bought it	In person
P12-AH	Amazon Echo	Father to P13	M	Daughter's husband works for Amazon, gave it to P12 as a gift	In person
P13-SB	Amazon Echo	Daughter to P12	F	Christmas gift from brother-in-law, who works for Amazon	In person

P14-NL	Google Home	Son to P16	M	Bought it for P16 as Christmas gift, is a tech enthusiast himself	In person
P15-JL	Google Home	Wife to P16	F	Bought it for P16 (tech enthusiast husband) as Christmas gift	In person
P16-TL	Google Home	Husband to P15	M	Received it as a gift from P14 and P15	In person
P17-AN	Amazon Echo	Sister to P9	F	Bought it on Prime offer after announcement	In person
P18-BR	Amazon Echo	None	M	Father's Day gift	Skype
P19-TB	Amazon Echo	None	M	Father's Day gift	In person

7.2. Participant profiles

In this section, I present a short summary of the important points of the interviews, preceded by a quote which highlights one of the elements of the interview which stood out to me as important to answering my research question about the user experience with smart speakers and how to improve their experience through information architecture.

P1

"Obviously it's up to me to do the troubleshooting, it's not going to troubleshoot itself."

P1 was convinced to try out an Echo after their blind friend bought one and liked using it, and he and P2 bought it for Christmas in 2015, to try it out before gifting one to P2's stepfather. P1 and P2 enjoy using the radio, timers, lists, music, news, and weather. He feels as though he is teaching himself how to use the device, as opposed to the device learning from his use. For example, when controlling the volume, he has to understand whether 50% of the current volume is too quiet or still too loud for a specific context—until he learns this for himself, it might begin to feel frustrating to constantly re-specify volume changes.

Learning is also exciting and interesting, despite having owned the device for about a year. Regular use builds into routine, which happens "pretty quickly" according to P1. New uses take time to learn—first to discover the new use (and whether it is relevant or not), then to learn how to implement it, then to remember to implement it enough times that it becomes routine. The investment of time is still worth it as P1 understands that the device is still developing and growing.

The onus of discovery lands largely on the user. P1 stated that he does not frequent the Amazon Alexa app, and acknowledges that his "awareness of some of those things is probably limited." He also acknowledges the usefulness of accessing the app for new ideas: "When you access the app to open the lists, then usually there's some kind of, 'Oh did you know you could do this?' And that kind of opens up other thoughts or ideas."

One of P1's biggest frustrations is troubleshooting with Alexa. For example, the device will intermittently "want to lose the internet connection." After rebooting, the device will usually automatically reconnect to the internet. However, if the issue is not the internet connection, further steps do not always

automatically become obvious. For example, the cleaning worker they hired inadvertently clicked the Mute button on the top of their Echo. Since P1 and P2 did not notice the red ring around the top, nor did they think to notice it as they did not originally know what it meant, P1 ran several reboots before realizing that there might be another reason that Alexa was completely unresponsive.

In terms of satisfaction with use of the device, "I'd say in terms of it doing what we want it to do, it's about 80-85%. It wants to give me some other station occasionally for (X), then it gets in a rut for that, and so then I just have adapted to say, well then I need to give it more cues, and then no problem." P1 seems to be happy using it, but not excessively invested in

P2

"I got a lot of friends, not looking to talk to a robot, that's ok."

P2, being in the same household as P1, faces many of the same issues. P2 is not the tech enthusiast in the household, and she leaves the discovery of new uses and the majority of troubleshooting to P1. P2 claims P1 uses it more, but P2 uses it for timers, and asking about weather abroad where her children live, and shopping lists. P2 qualifies her device use as purely task-based or information seeking as opposed to a conversation, but she states that, while she is not looking for a conversational robot, it is still nice to interact with a pleasant voice. P2 lies using Alexa better than using the microwave for a timer, because she doesn't have to stop what she is doing, go to the microwave, manually enter the time, and then manually end the timer once the time has run out. Most of P2's learning consists of finding out how to make Alexa understand the request/command being given, and this learning often occurs while watching her husband and (when he is visiting) her son interact with the device.

One of P2's frustrations is access to the app: "One thing I don't like, is I use it for a shopping list, and then I go to find it on my phone, I have the app on my phone, but then it asks for a password, and I have so many passwords, there's no way I'm going to…, so then I end up not always being able to access it when I'm at the grocery store." This lack of access to a crucial part of one of the user's key applications for the device is hugely detrimental to the functionality this user experiences. Another frustration faced by P2 is understanding how to speak to Alexa. This is exemplified in her explanation of using the timer feature: Alexa signals that the time is up, and then the user has to say a certain phrase—P2 has learned that she has to get Alexa's attention, then stop and wait for Alexa to respond. P2 explained that she has to train herself to speak a certain way to have success with Alexa.

Overall, P2 was satisfied with the Echo's reliability on giving information such as weather, the time, etc.

P3

"I think for the most part positive. Convenient, you know, helpful. Probably not necessary, but nice to have."

P3 was very enthusiastic about having the family of Echo devices—at the time of the interview, she and her husband had an Echo Show, a Tap, a Dot, and a first-generation Echo. Despite expressing some frustrations, she stated that she was overall happy with the devices and what they could do. Originally, she thought they were just interesting gadgets, and now she has added that they are convenient to have in the home. One of the uses she cited as adding convenience was turning lights on or off as you move around the home, or dimming the lights, or asking for a camera (if you have one) to the outside. P3 stated, "As it's gone on, they've added more and more features, and it has become more interesting." She benefitted from her husband's love of technology in that he used the most time discovering what the

device can do, and then she learned from him, though she also occasionally read the emails sent by Amazon about new features, and looks at the paper that comes with the device which lists a few activities to try.

In terms of choosing the Echo over the Google Home, she explained that her family was "just so deep into [Echo], and [my husband] has got so much connected that it wouldn't make sense for us [to switch to the Google Home]." She and her husband had invested in multiple Echo devices, and had gotten her extended family invested in them as well.

She enjoyed using the list feature because anyone in the house could access them and add to them. The list feature was yet another convenience, as she viewed the immediate access to a grocery list as more convenient than trying to find your phone and then opening an app and adding the item to a list. Another of the conveniences offered by the Echo was that it offered an easy way to shop online by reordering from Amazon.

One of the frustrations of using the Echo was that Alexa does not always respond—either Alexa would not recognize the command, or it responded as though it has carried out the command without doing so (for example, being asked to turn off the lights, then responding as though it has done so, without changing the lights).

P3 learned to use the device from her husband, but she also recognized that she was learning to use it by trial and error with the device's response or lack thereof. She said that if the command did not work the first time, she would try again with the same words, then again with different words.

She was overall positive about her experience with the Echo, though she admitted it was not necessary. P3 was the only user in these interviews who owned an Echo Show, which was useful in that it showed suggestions for commands on the screen. P3 stated that she looked at the screen more than she expected to do when they bought it.

When asked about suggestions, she said that suggestions would be creepy from a non-device-focused interaction. However, she responded positively to suggestions coming from device-focused interactions.

P4

"We use it probably a little bit more and more every time we find something new to use."

P4's husband (P5) bought their Echo in early 2015 as soon as it was available for preorder. She described P5 as a gadget enthusiast, and at the time, the Echo was the only smart speaker available. They bought it because it would be something fun and different, and Amazon offered a discount for early adopters[1]. Their uses included jokes, business hours, conversions for measurements, weather, temperature local and abroad, news and music. They also used the list feature for grocery lists. As Amazon announced developments, they used the device more and more. They also had a remote for the Echo, making interaction with the device easier by pressing a button to activate the device rather than having to use the activation word. In terms of integration with the home, they had lightbulbs which could be controlled by the Echo, and she saw the possibility of integrating their home further as long as they could solve the issue of the lights requiring very specific vocabulary in order to turn them on or off.

[1] In this paper, and in these interviews, the terms "adopter" and "early adopter" are not in reference to Mahler and Rogers **Invalid source specified.**; these terms were used by the interviewees, and I reported them as such.

She mostly learned to use the device from her husband's research and experimentation. From that, she knew that you could name alarms (e.g. "kitchen alarm") and learned about many other uses.

P4 reported no frustrations with the device initially, beyond activating the lights requiring specific vocabulary. Instead, she was optimistic about her future with the device. She reported that it could be interesting to have suggestions based on location for movies, concerts, etc., or to have the calendar and phone synced with the Echo. However, if the device was to provide suggestions, she said it would be creepy if the suggestions came from content not directed at Alexa.

At the end of our interview, P4 asked Alexa what the deal of the day was. Alexa allowed P4 to order it, and at the end of the order, Alexa provided information about a new feature—Alexa can notify users when the order is out for delivery by making the ring around the top of the device yellow, then the user can ask for their notifications.

P5

> "Just about every day you get an email from Amazon. I'll turn some of them on and try them. But those are little one-offs, mostly."

P5 was an enthusiastic early adopter of the Echo. He bought the Echo originally because as an early adopter, he could get "a Star Trek computer device [he] can talk to" which would allow him to indulge his gadget interest. When he bought it, he thought it would just be a toy, and he was not sure it would do much beyond being entertaining. He bought it because, in his words, "I had to have it. I had to have it right away."

He had an Echo in the kitchen/dining/living room space, and a Dot in his and P4's bedroom.

P5 enjoyed the music aspect of the smart speaker the most. He would ask Alexa for artists he remembered from when he was a child and attempt to "stump" Alexa. Using an unlimited Prime music subscription, he heard songs he had never heard before using Amazon's music service.

Additionally, he enjoys the voice interaction, which allows him to change the music, the volume, the lights, etc. without having to leave the couch. He liked that the device could give him reminders, as opposed to trying to set them on his watch or phone. He also liked that he could act immediately on an idea, without trying to find his phone. Ease of use was one of the highest items on his "pro" list for owning the Echo, and making daily life easier also drove his use patterns. He had installed Echo-controlled lightbulbs and was looking to eventually install Echo-integrated thermostat, garage door controls, and door locks. He wanted future developments to integrate with existing grocery delivery services

In terms of learning to use the device, P5 read the Echo emails, the Amazon site, and the Alexa app, but he claimed that the actions listed in these sources were not as useful as more frequent uses such as music, news, weather, and the integrated lights. He also said he would read online blogs and search on Google for uses.

Responding to the question about the device providing suggestions based on Alexa-directed interactions, he responded with overwhelming positivity that he would like that feature. While he thought that the majority of people would not like that feature, he suspected that, over time, people would like suggestions from Alexa-directed interactions as well as non-Alexa-directed interactions.

In the future, he wanted Alexa to be more intuitive. For example, when he made a doctor's appointment, he would like Alexa to know that he liked the 8am appointment slot. He wanted the calendar integration to push notifications, such as when the location of a meeting had been changed. He liked the newly-

released feature of the Echo showing a yellow ring around the top when notifications were available, but he wanted Alexa to be able to know when they were home and provide notifications when they had come into the kitchen after coming back from work. In the future, he wanted it to have more characteristics of a real person, such as greeting them when they walked in the door, or having more of a "personality," and making interactions less of a one-way, user-driven street.

P6

> "So maybe the real test for any of these new technologies is how tolerant the human operator can be to not give up on it and continue to put new challenges out there to see what its capabilities are."

P6 had an Echo in the kitchen/dining/living room area, a Dot in the home office, and a Dot in the master bedroom. He said that this arrangement was convenient for the way the traffic in the home moved. The Echo in the kitchen had the most use, with timers, measurement conversions, news, and music being the most frequent applications for that space. The Echo in the home office was primarily used by P6's wife asking miscellaneous questions. P6 bought the Echo originally because he was a Prime member and received an early adopter discount, which allowed him to buy the device mostly out of curiosity about the excitement surrounding the voice interface.

When asked about whether he would have chosen differently if he had had the option, P6 stated that he would have chosen the Google Home because he uses an Android phone, and he had found that the apps and services provided through Google were the best on the market in his experience, especially if he wished to integrate his email, calendar, and other services with a smart speaker.

As time passed, P6 stated that his original curiosity about the device wore off, and he and his wife had settled into using the Echoes more for routine use, such as the timer, alarm, questions, news, music, etc. He said he did not use the Echo for anything very "strenuous," where the cost of time and energy to discover and implement a new skill or feature would outweigh the benefits. In terms of discovering new uses, he would read the email sometimes, he listened to tech review podcasts, and he read articles online. Additionally, sometimes he would simply try something he thought it should be able to do. When he first began exploring the app, he found that the skills were difficult to browse, and there was no clear idea of what a "skill" was when they were first announced. Despite the fact that they are now categorized and searchable, the skills often are not as well-developed as they need to be to have success.

P6 cited frustration with not knowing the device's abilities and limitations, and that, while the device is constantly changing and developing, the user might not have the patience to keep up with those developments. Instead, the user is expected to use a certain set of phrases to activate a skill or ability, and the syntax and vocabulary of these commands is more limiting than he expected.

In the future, P6 would like these devices to be more intuitive, using access to information about the user, information about its location, and other references and resources to help the user navigate their world a little more easily. He recognized that the ongoing limitations of the device would be the language barrier and whether or not the user has integrated their home (e.g. lights, thermostat, plugs) and other devices (e.g. mobile phone and apps on the phone) with the smart speaker.

He recognized that the user is the limiting factor in integration, but that Alexa is the limiting factor when it comes to lifting the language barrier. If Alexa was not able to be more flexible with versions of commands it responded to correctly, the user would likely become frustrated and limit or stop experimenting. He said that the Google search engine did a better job of parsing the question and responding correctly than Alexa did, and it frustrated him to have to use the activation word. However,

he did mention that the user needs to train themselves to be more tolerant, patient, and flexible when learning to use the device.

P7

> "I'm not intimidated by it. I thought it was going to be a real technical thing, it's really not. […] I haven't attached it to anything, to my phone or anything like that. I just use it for some simple, everyday things. Pretty much entertaining things."

P7 received the Amazon Echo as a gift from her husband. Her Echo is in the kitchen, which is in the middle of an open floorplan house (allowing access from multiple "rooms" which do not have separating walls), they have a Dot in the sunroom (a room with access to the outdoor pool), and a Dot in the Master bedroom.

When she first heard of it, she had no idea what it was or what it could do. She described herself as "not a tech person at all." P7's husband was the gadget enthusiast, and she saw three benefits to having the device: 1) she likes to listen to music, 2) it can keep lists, get recipes, etc., and 3) it is a fun toy from her husband. As a non-tech-savvy person, P7 keeps the brochure which came with the Echo underneath the Echo for inspiration of what to do with it. At some point, P7 added a skill which provides her with a focus or inspiration word for the day. She said she was not intimidated by using the device as she was originally, but she was unsure of how to do certain things, or her ability to add new skills (e.g., integrating with the phone, ordering from Amazon), and she primarily uses it for entertainment.

Her primary uses for the Echo included music, trivia questions, questions about distance (they had recently moved, so distance between their home and the grocery store, for example), jokes, 20 Questions, weather locally and abroad, and lists. She used the Echo multiple times a day, but she reported that she did not do much experimenting or active searching for new uses, and she did not know about the Alexa app until it came up in our interview.

In terms of future use, she said she would like more Dots around the house, and she would like to make more lists once she discovered the Alexa app thanks to the interview (e.g. movies she wanted to see, gift ideas).

Responding to the question about Alexa suggesting uses, P7 liked the idea of suggestions from Alexa-focused interactions, but was very negative about unprompted suggestions from non-device-focused interactions. She went so far as to describe this as "invasive."

P7's frustration with the Echo is primarily a lack of consistency in whether or not Alexa will answer a question, including Alexa not recognizing the question properly and giving a "wrong" answer, or Alexa not understanding the question being asked and responding that it is unable to answer the question. Another frustration with the Echo is that she often has trouble getting the Echo to connect to services such as Pandora, despite others being able to do so.

Overall, P7 was happy with the device, but faced significant difficulty when attempting to take her use further.

P8

> "The burden of navigation is entirely rested on the user's knowledge of what's possible, whereas with any other graphical

> interface, you have an opportunity to present things to the user and explain what the possibilities are."

P8 received the Amazon Echo as a gift, and after the Google Home was released, he bought a Google Home for several of his family, friends, and employees, as well as for himself, as a gift. He was the first participant with young children in the home, with two children younger than 10 years old. The Echo is in their bathroom, and the Google Home is in the kitchen/main living area. P8 said that the kids mostly use the Google Home as an "eliminator of dispute" in their debates, or for just general questions and trivia. He also said the Google Assistant provides more comprehensive answers than Alexa, where Alexa often will respond that it cannot answer your question or does not understand your question.

P8 uses Google for enterprise communication for his company, as well as personal communication, reminders, and tasks, and he has a Pixel (Google released) phone, as well as a Chromecast, so having the Google Home is more convenient in terms of syncing reminders, calendar, email, etc. After buying the Google Home, P8 said that he started using the Google Assistant on his phone more as well, where he had not used it much, if at all, before having the Home. He primarily uses the Echo for streaming radio and music, more as a speaker than an assistant, though he uses the Home for casting photos to his TV, for answering questions, and many other integrated uses (e.g. Google services).

In terms of frustrations, P8 said that it was difficult to know the capabilities and limitations, and it is difficult to keep up with how the devices are advancing and changing: ""I have no way of knowing that those new capabilities are there unless I spend time researching or finding out. I'd say I spent more time up front just experimenting with it, and understanding where the limits are, and I've interacted with it based on those limits." He complained of a heavy burden of navigation placed on the user to discover what is possible, as opposed to other interfaces, where there is an opportunity to read about the new capabilities perhaps more easily than receiving them on a voice interface. He suggested that the features could be released in a more piecemeal way, similar to the way the emails work, but in a more personalized way, to educate the user about a use they might actually be interested in applying. He also has issues with both the Home and the Echo mishearing or not responding to him.

Originally, P8 thought the advertisements for the Echo were a joke. The idea of a device sitting on your countertop always listening and waiting for the user to ask it questions sounded like a stretch in terms of usefulness, when those capabilities were already available on a phone (e.g. Siri or Google Now, at that point). He admitted that he underestimated how useful the smart speaker would be, since the phone is not always listening in the same way the smart speaker is, and the user might not think about using the phone in the same way.

Discussing how his children use the devices, he said, "There's this built in expectation that you can just speak to something and it's going to do the things that you're asking it to do. In [the kids'] mind, it's normal. They're very confident with it." He is worried that there will be a large gap between generations who grow up with this type of ambient, always-available technology, and generations who get left behind in the transition to connected/integrated homes with smart speakers and devices that are linked with them such as lights and plugs. P8 pointed out that his kids did not think as much about limitations and possibilities, they just experimented with it, whereas he and others he knows often had difficulty using the device naturally and knowing what is possible. His mother, who had a Google Home, had difficulty remembering to speak clearly enough for the assistant to understand her, and she would just get frustrated with it.

In reaction to the question about the device suggesting uses, P8 responded positively to the idea of the assistant tracking previous requests and interactions to provide targeted suggestions in the future. However, he said if the device was to suggest based on non-device-focused interactions, it would be

annoying, and he worried that the feature would be misused or inappropriate, since he thought it would be weird even if a human provided feedback or suggestions based on an overheard conversation.

P9

> "When I'm typing I can be very precise, and it doesn't misinterpret what I type, whereas when I speak, I am less precise, and therefore, it doesn't understand what I want."

P9 had just bought the Google Home on Amazon Prime Day (July 11, 2017), twelve days before the interview. Her son was home, and he and P10 (P9's husband) decided to buy one at Best Buy. The Home sat on a side table between the living room and kitchen/dining area, allowing access to the device from the primary space in the home.

When she first heard of the Google Home, she thought it would be interesting to talk rather than type, but she was not sure how useful it would be. By the time of our interview, her thoughts were still largely unchanged regarding the usefulness of the device, and she still did not know much about using it. Her regular uses at the time of the interview were using it to play music, though she had experimented with asking it to tell them prices for international flights from their location to where their son was living and with asking it for information like when a football game was happening, and was frustrated with the lack of availability of information the Home was able to give them. Just before our interview, she had experimented with using the timer to see how it interacted with the music feature, and how it worked compared to using her phone or other timers.

In the future, P9 said she would be interested in using it to keep shopping lists, perhaps integrating it with a security system, and having integrated lighting systems. She speculated that it could be nice if the device knew you were close to home and would turn the lights on for you, or if it would be able to lower and raise the blinds when the sun was coming in the windows, or to help them monitor power usage. She also said it would be nice to hear her schedule in the morning, but her work phone (an iPhone) was not connected to the Home.

Regarding the Home suggesting uses, P9 responded positively to suggestions from previous device-focused interactions saying that it would be interesting "as long as I can shut her up." However, if the device were to suggest uses from non-device-focused interactions, she was unsure how she would feel about it, but she said it would probably be annoying. She did not like when ads online changed based on what she had just searched or done on the internet, and it would be "really weird" if the Home started using non-device-focused interactions to inform its suggestions.

Her biggest frustration was that the device was unable to provide results for searches that worked on a Google search on a phone or computer, and that voice allowed for more misinterpretations and misunderstandings which resulted in frustration.

P10

> "It's on me as the user to ask it in a way that will give me the information that I need."

P10 bought the Google Home on July 11, 2017, twelve days before our interview. He had first heard of smart speakers from TV ads and news articles, and thought they were a novelty item, but he could see them being useful eventually—the more he heard about them, the more he was intrigued. Then on Amazon Prime Day, Best Buy offered a price match for the Google Home, offering the Home in-store at

the same price as Amazon was offering the Echo online. P10 and P9's son suggested that they would enjoy the Google Home more than the Echo, because it interfaced with what they already had and they could bring it home with them that day from the store, rather than waiting for it in the mail.

The primary uses P10 had were weather, asking about the traffic on his route to and from work, and music. He was impressed with the quality of the speakers, and he felt that it could stand on its own as a speaker. However, because of he had just bought the device less than two weeks before the interview, the novelty had not yet worn off, but he saw a lot of potential for future use. He would like it to control more things in the home, but he did express some worry about security, that too much information could become available about their use patterns.

P10 spent some time experimenting with the device by asking the Home until it responded to what he was asking, changing the words and intonation, and while it did not always work, he was happy to continue trying new things. He was still trying to find out if there were better ways to make his search more effective, such as using certain keywords to help the assistant recognize his request. He liked the voice interface's convenience, as it was often the quickest way to get an answer to a question if the assistant could answer it.

In response to the idea of the device suggesting uses, P10 said it would be "outstanding" if the device would tell you it could do something new, but he said that suggestions from non-device-focused interactions would not be something he would want. He only wanted the device listening when he asked it to do so, with the exception of a security system, where he would want a warning if there was a reason for alarm.

Overall, P10 was pleased with the device, and saw a lot of potential for future use. He liked the idea of unlimited information available immediately no matter what he was doing or where he was.

P11

> "Nobody else cares about all this new technology, and I'm tired of figuring this stuff out on my own."

P11's husband bought the Echo after a friend of his talked about how much he loved using it, and her husband bought one to give to her parents and one for them. P11's initial impression was that she and her husband did not need one in their house because they had a TV and an iPad. Despite the device having been bought in 2016, she did not use her device until she and her husband moved into a new home in 2017 because they decided to connect their home with smart bulbs and plugs.

She primarily uses it for weather, and their home has lights connected, but otherwise, she just leaves the TV on for news and music. In the future, she would like to control the TV, have radio/SiriusXM on the Echo (which it can do but she had not figured it out), having the grocery list, and calling through the Echo, but she cannot imagine doing much beyond that because for many of the uses (e.g. calling, recipes, timers, etc.) are available on her iPad.

In terms of suggestions coming from the device, she would be fine with suggestions from device-focused interactions, but non-device focused interactions "would freak me out." P11 was more comfortable if she initiated an action, but actions not initiated by her would be upsetting. However, she said she would accept the device asking permission to access certain information from the iPad or her phone or other devices.

Her biggest frustration was that it required such precision from the user to activate an action. She would like more anticipation or help from the Echo, but she likes the activation word having to wake Alexa up.

She wished she could integrate her Apple devices (iPad, iPhone) with the Echo, because she uses those much more frequently to look up information, to play music, to store lists, and many other actions the Echo offers without a screen. She enjoyed exploring down information trails, and she gave an example of a Wikipedia binge on Charles and Diana she went on after the anniversary of Diana's death, and P11 said that she would not be able to do that kind of exploratory information search on Echo. Also, she enjoys using the apps on the Apple products because she could find a quality app which will do almost anything she could think of, though in terms of discovering new uses, P11 simply said she did not have time to explore that deeply, and without anybody else in her family to get excited about the Echo with, she did not feel the motivation to use the smart speaker more.

P12

> "I haven't figured that it would be sufficiently beneficial—have you heard the term, 'The juice isn't worth the squeeze'? Sometimes after I get used to it, I will continue to use it, but I won't really go out of my way to do so."

P12 owned an Amazon Echo, which he received as a gift from his son-in-law who works for Amazon. P12 went blind about twenty years ago, and he uses Siri on the iPhone as well as Alexa on the Echo. His Echo sits on a shelf between the kitchen and living room area, so there is access to it from most of the home. When he first heard about them, he thought they were "a tremendous trifle; much ado about nothing" because he did not know what to do with it, and it was only over time that he learned any of the relevant uses. He thought the price was too high for what it offered.

His use included time, weather, definitions of words (though he said Siri has a better dictionary and generally responds to more information requests), he uses timers, reminders, weather, and he recently began using music and news. One story he gave was that he asked Alexa to remind him not to feed his helper dog the night before the dog was supposed to have surgery. He characterized interactions with Alexa as similar to having an administrative assistant, though perhaps not a very smart one. He often leaves the news or music running in the background, but P12 does not often interact with it other than to change the song or occasionally to set a timer.

P12 was a unique user in terms of how long he had been using assistive technology due to his blindness, and every interaction is essentially voice-based for him. He uses his phone to help him "see" the world around him, through apps which will tell him what color a shirt is, or what item his camera is pointing at, and his phone is portable, whereas the Echo is not. The iPhone can be slipped into his pocket, whereas the Echo must be plugged in and stationary to work.

One of his frustrations with the Echo was that Alexa interrupted him frequently. Also, he wished Alexa was less literal—if he did not get the title of a song precisely right, she would not give him the response he was looking for. He felt that 90% of the time, it worked well, but that last 10% where it did not work was annoying.

In the future, he could use the device to send texts or emails, which he currently uses his iPhone and computer for. He would like the natural voice processing to improve to the point where the assistant would pick up his words correctly and reduce the errors in word choice that he faces now with the iPhone and computer.

He would like the suggestions to help him discover the skills for the Echo, and he would like to be able to explore the Alexa app more easily. However, in terms of the Echo suggesting uses based on non-device-focused interactions, he would not care for Alexa listening and keeping recordings of interactions where he had not activated it.

He did not think using the Echo was as useful as voice over on his phone, but he did think it would be worth the work to find further uses for the Echo at some point.

P13

"The kids demand things from her. I wish she wouldn't do it unless somebody said please, you know?"

P13 owned an Amazon Echo, which she received for Christmas from her brother-in-law, who worked for Amazon. She was the second participant with young children in the home--she and her husband had four children under the age of 10. She originally thought it was "kind of fun," though she does not have a very "tech-y" family, so her nephews (11yrs old, 9yrs old, 6yrs old) helped explain to her and her family how to use the Echo. Often, she and her family would forget that the Echo could help them, so they would use their phones instead. At the price point of the Echo, she thought that the Dot was the only device worth the quality it provides as a speaker and an assistant. They kept the device in the kitchen primarily, but they would sometimes take into the basement if the family went down to the basement.

Since receiving the device about 3 years before the interview, she said her understanding of the Echo had not changed. She used the device only a little more often than at the beginning, but she thought she had not broadened her horizons very far beyond her established use. She and her family used the Echo for the following: music, spelling help, definitions, weather, time/timer/alarms, radio, directions, location, phone numbers, jokes, and music. She reported that she and her family used the device multiple times daily. One example of using the timer that was unique to this interview was that she said she would set the timer for 30 minutes

In discovering uses for the device, she mostly asks her sister (married to the brother-in-law who worked for Amazon) for help. When asked if she had used the Alexa app, she said she had installed it on her Kindle, but that she did not know where her Kindle was and was thus unable to access the Alexa app. She said she could download the app on her phone, but she knew that it would likely take more time than she wanted to spend to figure out which email she used to set up her account, to figure out her password, and then to set up the app with the device. In the time she did use the app on the Kindle, she did not like how previous requests scrolled on the front page, though she did not understand that those previous interactions being displayed were part of the way Alexa learned to interact with the user.

One of her frustrations, which I did not hear from any other participants, was the way her kids interacted with the Echo. When they interacted with the device, they would shout over the adults and bark orders at Alexa. She cited that they were not being encouraged to use their "manners," or to speak politely when asking questions or giving commands to Alexa. She worried that her kids were learning to feel "entitled" when the Echo will respond immediately no matter how the request is given.

When asked to characterize the device, she said it was "frustrating," especially when Alexa would respond that she did not understand the question. She liked having the activation word, but did not like how she had to re-activate it with successive requests. She said it was more like a robot or a limited personal assistant, because interactions were very one-sided. Another frustration was having her kids talk over her to try to help her or her husband or each other with a request/question, where Alexa would then not recognize any request.

In the future, she would like more applications for helping her kids. For example, having applications for quick exercises they could do to get up and move during homework time would be good, or relaxing breathing exercises. Mostly, she just wanted the process of finding those applications to be "easy."

Regarding the "feel" of having the device, P13 said that it was not physically intrusive, it was just "there." She would not want it to be conversational, or part of the family. She was happy with it just being a utility.

When asked about the device suggesting uses, she liked the idea of suggestions coming from device-focused interactions. However, if suggestions came from non-device-focused interactions, she said it would be "creepy." She told me that she hated using Gmail because of reactive ads to what she was writing in her emails. On the other hand, if the device would ask for permission to access use information from other apps, she said she would probably like the ease of use that would provide.

P13 was one of the only participants who commented on the fact that it was nice not to have to use a screen because it allowed her to look at her family. Using the Echo allowed her family to reduce their screen time, for example by streaming news (rather than reading an article), reading books, and answering questions (rather than having to stare at a screen to find a result).

P14

> "To do more stuff, you have to buy more stuff. Like to turn a light on, you have to buy a plug."

P14 went with P15 (his mom) to buy the Google Home for P16 (his dad) for Christmas in 2016. He first heard of them through a tech channel and thought they were cool. They bought the Home for his dad, but it was more a gift for the family. In choosing between the Echo and the Home, he chose the Google Home because it would integrate with their phones and their Chromecast.

Originally, the Google Home came with a subscription to YouTube Red (a special service from YouTube with special channels and music streaming), so he used that originally for music. Additionally, he used it for timers, weather, lights, TV control (and speaker control), and Spotify (for music). He said he used it about once or twice every other day, but that his family used it about five times a day in total.

In the time since his family got the device, he has learned more about the uses and limitations. For example, in order to integrate the home, the user has to buy new plugs, new lightbulbs, etc. that can integrate with the device. He thought he would be using the Home more than he used it in reality—for example, he thought he would be able to get the Home to play a specific episode of a show on Netflix, but instead the Home will start the show from the pilot episode of the show.

To learn to use his device, he used YouTube and Google to explore new uses, since he already spent time watching and reading tech-related news.

He would like the Google Assistant to be more conversational, and he wanted the Assistant to have a more realistic voice like the one used on the Google Assistant on Android phones. He felt that the Assistant was still limited, though.

P14 told a story about how his dad, P16, would stand up to talk to the Home, or he would raise his chin while speaking to the device, and P14 said it did not feel weird to talk to it, and he did not have to "do anything weird" to talk to it.

He was frustrated that the activation word was difficult and that the device itself was not actually very responsive to commands in his experience. He had discovered a different way of saying the activation word, "Ok, Google," to "Ok, Booboo," which was easier for him to say and still activated the device.

When asked about the device suggesting uses, P14 said that it would be "cool" if the device provided suggestions based on device-focused interactions or from granting permission to access use data from other apps. However, when asked about suggestions from non-device-focused interactions, he said it

would be “creepy” and the question made him “a little bit freaked out.” He liked the idea of “opting in” rather than “opting out” to the device suggesting uses either way.

He thought it would be nicer if it had a screen, but he did say if it had a screen, it would not be as discreet or as nice of a design (which was an aspect he liked about the current device).

P15

> “[P16] uses it all the time as a timer. It never even crosses my mind to do that. Pretty much radio and news is it. It’s completely under-utilized I’m sure.”

P15 first heard of the Google Home in the summer of 2016, and she bought one for her husband for Christmas 2016. She went with her son (P14), who was also excited about the device. She was convinced to buy the Home because Best Buy (a technology store in the USA) carried the Home and plugs that could integrate with the Home, both of which she could purchase and take home, without waiting for shipping. She was also more interested in the Google Home because her family uses Google services (email, calendar, etc.), so the Home was an ideal device for connecting with everyone’s accounts. Before purchasing, she read some reviews online and went into the store to talk to someone, and once she had put time into investigating this device, she was satisfied with its functionality and its price point and did not want to investigate any further into other devices.

One of her favorite aspects of the device is the news readout, and she uses the Google Home as her primary news readout. Additionally, she enjoys using it because the Home is a much easier way to play music and radio as opposed to using three remotes to turn on the radio and the stereo in order to listen to something. Weather is another use she enjoys, especially when trying to decide what to wear that day. Also, when she is cooking, she often asks for conversions (tablespoons to cups).

She had not spent much time discovering uses, but spending time discovering uses was on her “to-do” list. If she learns about new uses, it is usually from one of her sons.

Regarding fitting into the home, she thinks it fits well into her home. However, when people use it to turn on or off the lights, she thinks about what her grandparents would say, that her grandmother would think they were lazy and her grandfather would think that the device was really cool.

She had not noticed anyone in the home changing their voice to speak to the device except P16 (her husband), who spoke loudly to the Home.

One of her frustrations is that the assistant will answer a command one time but it will not respond to the same command a second time.

In terms of future use, she would like it to respond to more variations of a command. Also, she would like it to suggest uses from previous interactions or connected apps, since she lacks the time and inclination to research on her own, but she said that the device suggesting uses from a conversation not inclusive of the device would be “creepy.”

P16

> “It does a lot of things well, but nothing outstanding.”

P16 was given the Google Home as a Christmas gift from P14 and P15. He said he likely would not have bought it for himself, but he was glad to get it as a gift. In terms of connecting multiple smart speakers, he was not interested, but he has some connected plugs turning on and off lights in the kitchen and a

Chromecast which the smart speaker is linked to. His use of the device has not advanced very far beyond what he tried when he first got it (music, lights, etc.) because he does not want to devote the time to it. P16 used a program called "If This Then That" (IFTTT) to connect the light plugs, which allows the user to specify the phrase the device will recognize to carry out actions such as activating plugs, lightbulbs, etc.

He has used Sync on Ford vehicles (another IPA but embedded in the vehicle) for about 7 years, so it was natural to start using the Google Home. He said it is very handy to do things while you have your hands full, such as starting a timer, changing the lighting, or changing the thermostats. However, if you want it to read out a recipe while you're cooking, it cannot read out step-by-step recipe instructions, and P16 said that using a tablet would be more convenient.

One of his frustrations is that the Home has trouble recognizing commands, and that the device recognizes certain voices better than others because those users have more interaction with the device. Additionally, he was frustrated that the device does not recognize variations of phrasing for turning on and off the lights that he has programmed through IFTTT.

In terms of future use, he can see the Home being a connected home hub. He liked the idea of using it for a security feature to program lights to turn on and off while you are away from your home, and he liked the idea of being able to turn the thermostat on/off while you're out of your home.

When asked about the Home suggesting uses, he said, "I want it to be stupid. I don't want it to be too smart. […] I want a device to do what I tell it to do. I want to pull information, I don't want it to push information." He spoke negatively about the device pushing ads the way websites do currently. He used the term "invasive" to describe the potential of suggestions coming from the device unbidden.

His favorite aspect of the device is the hands-free ability to interact with it, and he liked having the ability to play music from it.

P17

> "Where's the line between helpful assistant and taking over, and surrendering your control, your everything, to this device, this non-living thing?"

P17 owned an Amazon Echo, which she keeps in her bedroom, and an Echo Dot, which she keeps in the main kitchen/living area. The Echo was purchased when the devices first came out in early 2015 because Amazon offered a significant price discount for early adopters. She said she would be interested in the Google Home because she uses Google services (e.g. calendar, email, etc.), and she does not use the Echo for any integrated personal services such as lists or calendar integration because she is satisfied with using her phone or computer to carry out those tasks.

P17 used her device for news, weather, traffic, and as more functionality was added, alarms, time, weather, and general entertainment. She used to check the skills on the app, and read the emails, but that now there are "so many" so she could not keep up with the developments. She edited her newsfeed (a readout of the day's news from a user's chosen news sources, such as Fox News, BBC, NPR, etc.) more regularly than she edited any other aspect of her interaction with the device to include more relevant news sources. Originally, P17 thought she would be using the device much less, or in a more limited way, than she did at the time of the interview. She said she uses the device to help her overcome her insomnia: she did not want to use her phone because the screen's light can interrupt sleep patterns, so she turns on ambient noise (e.g. bird noise, waterfall, white noise) or plays 20 Questions (a simple guessing game) until she is tired enough to sleep. P17 enjoys interacting with the device, because she likes the Alexa voice, and some of the idiosyncrasies of Alexa make her laugh. She also likes how Alexa's

presence "fills the space," so she will often have old radio shows playing in the background. She also enjoyed pranking her dog by asking Alexa to bark or to meow like a cat.

P17 appreciated that the device was not flashy, and that it could sit well in the background without taking over the space. She also enjoyed using it, especially because you could customize it to allow you to do what you want to do.

In learning to use the device, P17 noticed especially with music, it could be frustrating to get Alexa to play the correct artist, or to play more than one song when the artist has multiple songs on Pandora or Spotify. Another of P17's frustrations was that the Echo will drop the Wi-Fi signal, and she has learned that she can ask Alexa whether or not the Echo is connected, and Alexa will tell her if the Echo is connected.

When asked to characterize interactions with the device, P17 stated that the interaction was not at all like a conversation. She responded negatively to the idea of Alexa being more conversational, "because a device or a tool shouldn't replace a human. It feels too creepy-Sci-Fi talking to this little tube thing. […] It's just a device. I don't feel comfortable with the whole aspect of artificial intelligence. […] Where's the line between helpful assistant and taking over, and surrendering your control, your everything, to this device, this non-living thing?" When asked about why she was comfortable surrendering her data to this device to allow for personalization, she said, "I'm still in control, I can tell it what skills to do or not to do, and an AI would make those decisions for me. I'm not comfortable with that." Later in the interview, when P17 mentioned that she would like the device to read to her, she said she would like Alexa to have more natural inflection when reading information off the internet rather than a prepared script, but that she still did not want it to cross the line of sounding too much like a human.

In response to a question about the device suggesting uses, P17 was happy with the device suggesting uses based on previous interactions with the device. In terms of the device suggesting based on access to apps on the phone, she wanted to still be in control of what the AI has access to, but would generally accept this level of personalization. When asked about the device suggesting uses based on conversations it hears in the home, she was uncomfortable losing control of the device. She said she would need time to become more comfortable with that, because at the time of the interview, she would want to ask Alexa before Alexa responded to something like a conversation in the home.

For future developments, P17 suggested that the device could become a therapy tool. For users with bipolar disorder or PTSD, for example, who might need intervention and do not necessarily always have access to a therapist right away, the device might help these users by noticing if the user is beginning to show symptoms that would need to be addressed. For example, if a user with PTSD woke up in the middle of the night in a panic attack, and had an integrated smart watch with a heartbeat monitor which could tell if the user had an elevated heart rate, the device could enter a "therapist mode" to help the user with management techniques such as breathing exercises until the user could access a real person.

P18

> "I didn't know *what* I would be doing with it, but I thought I'd be doing more than what I'm doing with it." "I'm all about instant gratification, so if I'm not getting the answer right away, I'll give it one or two shots."

P18 owns an Amazon Echo Dot, which was located in the family room, an area he called the main living area of the home. His initial impression of the device, before owning it, was that the access to music, the trivia, and the device itself were "cool." He received the device as a gift for Father's Day. He described the first experiences with the device as "things that you wouldn't think would be too difficult from a Google

search perspective, and the device kept coming back with 'I don't know what you're asking.' But it was not complex, what I was asking it to do." After some initial exploration, he said that his understanding had changed for the negative, "not quite as cool as I thought it was," though he claims that part of that negative impression could be his fault. He talked about how the app is difficult to navigate in terms of discovering new uses: "You've gotta be in there pretty hard, but there should be more information readily available about what you want." He expected the experience of asking the device for information to be more similar to a Google search, where he could ask information such as "top ten restaurants in Detroit, Michigan," but the device was unable to either understand or deliver information on that request. His mobile phone is Android, and he is a Google services user.

His primary uses are music, news, weather, and sports, and from time to time he will try to ask the Echo trivia questions. On average, his reported use was "once a day, tops." In discovering uses, he said that he has been on the app, and he knows he could go onto YouTube to figure out more uses, but he claims laziness as his excuse for not digging deeper. For example, he has a wireless speaker system in his home office, and he knows he can pair the Echo to pair up with his Bluetooth speakers, but, in his words, "It's almost like I'm a little bit afraid to, because I don't want to mess up what I already have set up in my entertainment system. There's got to be an easy way to do it, but I just haven't yet. […] I don't want to screw up a good thing." His fear of ruining his wireless speaker configuration by integrating with his Echo Dot contrasted with his knowledge that integrating the Dot with his wireless speakers would improve the sound quality of the music he listened to on the Echo Dot. In terms of other integrations, P18 said that he and his wife have been in their home for 20 years, and he was done making investments in it because they would likely move soon. In his future home, P18 said he would be likely to look into home integrations such as plugs, lights, etc., but he said "having a smart home at that level doesn't interest me. I'm okay flipping a switch." He doesn't use the lists feature, because he manages information that could go on lists in other ways that he is happy with.

Regarding learning how to use his device, he said that he used trial and error and that Alexa is responsive, and he said the main problem was that Alexa could not answer the question "98% of the time." P18 stated that he had not explored the Alexa app very thoroughly, and he stated that while he received and looked at the emails from Amazon about the Echo, he paid little attention to the emails.

The device for P18 does not provide a conversational interaction—it is a command-based interaction, where there is no turn-taking, or "back and forth" interaction.

When asked about the fit of the device in the home, both in terms of design and ease of interaction, P18 stated that the device fit nicely into the home. The Amazon Echo Dot sits in P18's home on a black shelf, and the device itself is black and small, so the device blends in well with its setting. In P18's home, there are often guests in the home, and P18 said there was no issue using it with other people around, or having guests use it. P18 said that the voice-based, command interaction was "bizarre" and "unique" from interacting with a screen, but that the interaction eventually became normal and "you get used to it." He also said that his voice is "firmer," meaning "louder," when he speaks to Alexa, to ensure that he is heard. He expressed that he was sure he did not need to "shout" at the device, but he speaks loudly when interacting with Alexa regardless.

One of his frustrations is that he wants to use the Echo like a Google search but that the device does not answer the questions he asks when he phrases them as he would in a Google search. He would like Alexa's ability to answer questions like Google to improve.

In terms of improving the device, P18 would like more information in the box on what the device can do, as opposed to a small pamphlet that primarily serves to direct the user to the Alexa app. He was open to the idea of the device providing suggestions based on previous interactions, but he would want to grant permission piecemeal rather than providing Alexa access to all his apps on his phone. He was also open to the idea of the device listening to conversations in the home (including interactions not directed at the

device) and providing suggestions from that context, and he suggested taking that interaction one step further. P18's suggestion was to take a conversation about going to a baseball game (for example) and allowing the Alexa to offer the option to purchase tickets.

P18 likes the fact that interaction with the device is voice based, and he enjoys the fact that he "can just sit in [his] chair and yell at it." His favorite use for the device is playing music: he knows what the news for the day is by the time he comes home in the evening, so the Echo playing music is a nice way to relax after work.

P19

> "I have a friend who knows 90% of what his device can do, and I might know 20% of what he knows. So, it's like a musical instrument—I can only play 'Mary Had a Little Lamb' and he can play 'Mozart' because he knows all the features. I'm just beginning to learn what the Echo can do. There's so much more to know."

P19 received this device as a Father's Day gift, in June 2017. His son helped him to set it up, and they spend time playing with it together. When he first heard of them, he thought they were cool, but not really for him. He is interested in discovering uses, and he has integrated with his home with a Wi-Fi plug for lights in the family room, and his white noise machine in his counseling office linked to the second Amazon Echo (full device, not a dot) that he bought for his office. He is interested in integrating his home with a camera, bulbs, and a thermostat which can link up to the Echo. He uses the device for news, weather, traffic, and music. He explored a little bit with the skills, and while he said he had not explored the Alexa app in a while, he knew it was there for him to use.

He characterizes his interactions as entertainment-based and informational, but not conversational. If it was integrated with his calendar, he said it would be nice to have that option to sync his and his wife's calendars to receive reminders. Beyond that, he was a new user and felt that he was unsure of the conversational nature of the device. During the interview, we explored a recipe app.

When asked how it felt to use the device, he said it felt good. One of his anecdotes was that he helps put his wife in a good mood by putting on her favorite song while she is cooking. He also was excited about the voice-controlled volume control, though he was aware that a louder volume on the device meant the user must raise their voice louder than the device's volume to be heard.

One of his frustrations is that he faces issues while integrating his Apple phone's calendar—he would like to be able to have his and his wife's calendar synced with the device, but he reported that the codes given by the Alexa app and by Apple did not match up, preventing him from completing the integration with his calendar. Another frustration comes from asking the Echo for songs and getting a radio station for the artist or song, or shuffling the songs by the artist if he provides the artist first and the song title last. P19 is used to using the iPhone, where providing a broken or partial command still results in Siri providing the user with options for action. He claims, however, that his Echo is better than Siri, because he can play more songs on the Echo with his monthly music subscription.

Another frustration, common to many of the interviewees, is that you must be specific with which words you choose, and you have to be ready to give the full command when you start the command. Also, during the interview, we interacted with the device, and Alexa interrupted a command midway through, and that made the interaction frustrating. P19 commented that the frustrating interactions, while the

user learns to speak with the device, could be a significant barrier to someone with less patience during what can be a steep learning curve.

Regarding the amount of information Alexa provides in a response, P19 said that if the device gives him too much information, he cannot remember it all, and he must ask it to repeat the information, delaying him from further action and causing frustration.

In terms of future developments for the device, P19 brought up the comedy show Saturday Night Live's skit on the Amazon Echo Silver, a made-up version of the Amazon Echo for elderly adults which responded to more abstract commands, which seemed better able to address a natural user command or request. Following this, P19 also brainstormed about what the device would learn about him, such as playlists for a certain time of day, or a form for autocomplete (similar to the autocomplete text suggestions a phone can provide). When the idea of suggesting uses came up, he responded positively to the possibility of the device suggesting uses based on previous use and patterns of use.

Despite his awareness of the frustrations, he is still very positive and curious about using the device.

7.3. Location of the smart speaker in the home

The smart speakers were generally placed in a few different areas in the home. If there was only one smart speaker in the home, it would likely be in an area where the user could access it from the kitchen and living/family room area. Any further smart speakers would likely be placed in the home office, the bathroom, or the master bedroom, or a combination of these if there were more than one. Due to the cost of the devices, few users who I spoke with had smart speakers spread throughout the home, beyond the kitchen/living area, bedroom/bathroom, and/or home office.

Below, I have included three figures illustrating a smart speaker's likely location in the home. The first figure shows the location of a smart speaker in a basic one-level ranch home where the kitchen, dining area, and living room are adjacent and provide access to the smart speaker without too much acoustic interference from walls or doors. The second figure shows photographs taken by the users in their home of their device in a main space in their home, usually a combination kitchen/dining/living area. The third figure shows photographs of these smart speakers in a more private space in the home, such as a home office, bathroom, or bedroom. All photographs of the devices were taken by the users.

Figure 8 One example of the location of a smart speaker (red dot) in a small, single-level home, placed in an area where users in the kitchen/dining area (highlighted in grey) as well as the living room (highlighted in green) can be heard by the smart speaker.

Figure 9 Smart speakers in a main space in the home. Clockwise from top left: (1) Amazon Echo on kitchen counter adjacent to dining area, living room, and family room; (2) Google Home on a table in the living/family room adjacent to dining area and kitchen; (3) Amazon Echo on countertop between kitchen, dining, and living/family rooms; (4) Amazon Echo in kitchen area adjacent to dining and living/family room; (5) Amazon Echo in living room with Echo pamphlet underneath

Figure 10 Smart speakers in private areas of the home. Clockwise from top left: (1) Amazon Echo Dot on a bedroom nightstand; (2) Amazon Echo Dot on a desk next to other small personal items; (3) Amazon Echo Dot on a vanity counter in an area between the master bedroom and the bathroom; (4) Amazon Echo Dot next to a speaker in a private room in the house; (5) Amazon Echo Dot on a nightstand next to a bed.

8. Analysis

8.1. Thematic analysis

In processing the interviews, several themes emerged. In order to more easily visualize the relationships between these themes, I sorted them into a map of three categories: 1) actions carried out with the device which were routine or unconscious; 2) actions that were described aloud; and 3) activities that made actions meaningful. This map can be seen in **Error! Reference source not found.**.

The first category of themes, actions that were routine or unconscious, included the practice of speaking to the device (the volume of the command, the structure of the command), and learning about the device. Establishing initial knowledge and building knowledge about using the smart speaker were the two phases of routine user actions involved in users learning about their smart speaker. These could include reading emails (a passive method of discovery compared to actively searching the internet for uses), trial and error, and repeating the action/command.

The second category of themes, described actions, includes the types of actions, the content of actions, and the rules of actions. The types of actions include entertainment (with examples as the child nodes), assistance, and informational search. The content of actions node contains child nodes which are the items included in the commands. The rules of actions node contains what the participants described as the rules they needed to follow in order to be not only heard by the smart speaker but also to make the smart speaker more likely to respond—this includes precision of speech and word choice, the flow of the command (if the device can listen to more than one command in a row), and the physical interaction with the device (the conscious control of the volume of the command, and the knowledge of where the device is located in the home).

The third category of themes, activity (what makes an action meaningful), includes child nodes of searching for data, interacting with assistant, feelings about using the device, troubleshooting, and referring to the device. Each of these nodes include activities, sorted into the way these activities provide meaning to the actions. All the work done by the user—to learn about the device, speak to the device, to know what the device does, and to know the rules of actions—is only worthwhile if the user perceives there to be meaning in interacting with the device. The user might interact with the smart speaker in order to search for data on weather, news, traffic, their calendar, or other data “skills” which provide the user with simple information which can inform how they move through their day. The user might also interact with the assistant, in terms of activating lists or timers, or asking the assistant to set a reminder or alarm. When interacting with the device, the user will also experience feelings about their actions, such as frustration that the device is not as effective as they thought or that it cannot hear them; enjoyment that their action makes it easier for them to carry out an action in their daily life; a feeling of routine, that using the device is normal or natural in the way they have chosen to do so; excitement when discovering new uses or having success when using the device; or worry that they are not using the device to its fullest ability, or that they are being listened to by the company or the government. These feelings give meaning to how users interact with their smart speakers. Additionally, there are actions involved in troubleshooting—these are divided into user error or machine error (which could also be called assistant error, depending on the actual source, but the point is that the device is the problem rather than the user). This type of activity drives the trial-and-error and repeating actions listed in the first category of themes. Lastly, the terms the user employs to refer to their device—whether they use “she” or “it”—was a theme I noted, but which did not give much meaning in terms of further analysis other than that “she” was more commonly used with the Echo and Alexa (perceived as a female name) and “it” was more commonly used with the Google Home and Google Assistant. Both terms were used to refer to both devices, perhaps because they are machines with female voices. This could be an interesting path of research for another project, but it falls outside the scope of this one.

Before moving on, I feel that it is important to address that this theme map was inspired by activity theory. Activity theory was not used as a theoretical approach for this paper, as it serves more to benefit an interaction designer than to serve an information architecture designer due to its focus on a subject's interaction with objects (via a tool or set of tools) and the subject's motivation for interaction (the desired outcome). However, this theoretical framework can still provide inspiration for understanding the complexity of real-life interactions and relationships between the user (subject) and the smart speaker (tool) in terms of the user's motivation to find and manage information (desired outcome). The area of activity theory which inspired the theme map in Figure 11 is the three-level hierarchy model of activity. In activity theory, an activity is generally understood as a purposeful interaction with the world carried out by a subject through an object. In the field of human-computer interaction, the meaningful interaction always occurs between a human and a computer. Therefore, as applied in this project, the subject is the human, and the smart speaker and the assistant both play as object (when the desired outcome is to use the smart speaker) and tool (when the desired outcome is finding and managing information). The human side of this dynamic lies more in the realm of interaction designers and the field of human-computer interaction, while the finding and managing of information lies more in the realm of information architecture (of course there can be overlap, but this project's focus is within information architecture).

The three-level hierarchy of activity was first described by Leont'ev (1981) and was elaborated on by Kuutti (1996), but a simple summary can found in Bødker and Klokmose (2011, pp. 319-320):

- Activity: "To describe an activity at the activity level means to focus on the social and personal meaning of activity and its relation to motives"
- Action: "To describe an activity at the action level means to focus on what the subject does and on possible goals, critical goals, and particularly relevant subgoals of subjects"
- Operation: "To describe the activity at the level of operation means to focus on how the activity gets carried out, the concrete way of executing an action in accordance with the specific conditions surrounding the goal"

In the sense that these terms inspired the map shown in Figure 11, the branch "Actions (routine or unconscious)" could broadly be described as operations in AT; the branch "Actions (described)" could broadly be described as actions in AT; and the branch "Activity—what makes actions meaningful" could broadly be described as activities in AT.

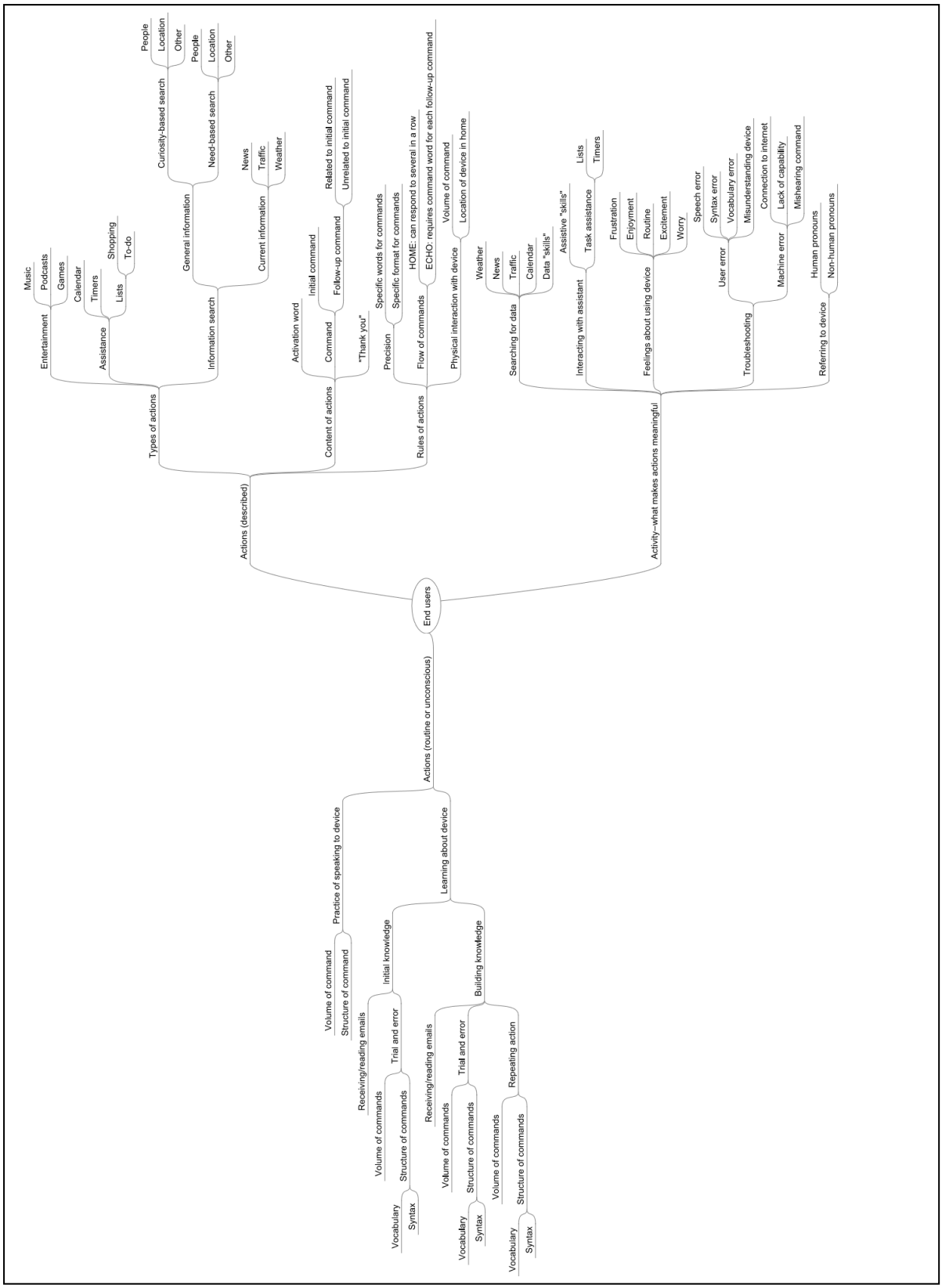


Figure 11 Map of themes in data

8.2. Stages of use

After spending some time with the users and with the interview data, three phases of use became clear. These can be seen in Figure 12, where the phases of experimenting, learning, and establishing use are shown. The first phase, experimenting, involves the user getting to know their own wants and needs for

the device, as well as the abilities and limits of the device. The second phase, learning, involves the user learning to what extent their wants and needs and the devices abilities overlap. The third phase, establishing use, involves the user establishing a routine set of uses with the device, where the routine uses have combined with what they know the device can do.

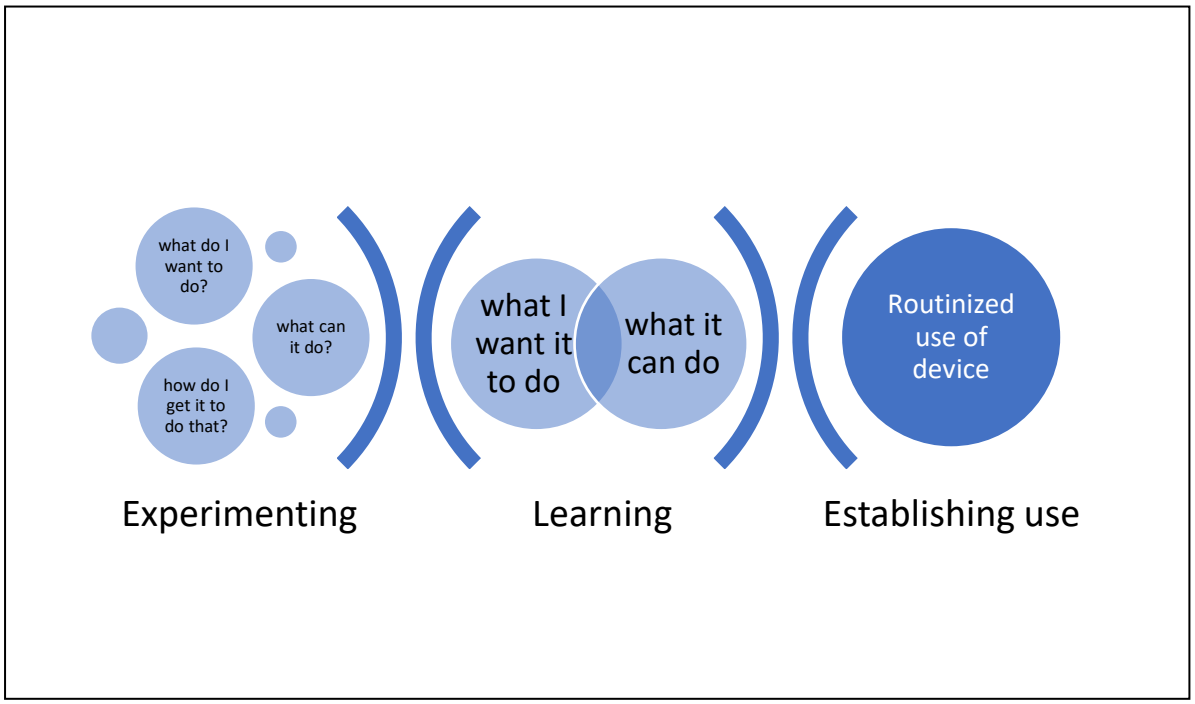


Figure 12 Phases of smart speaker use

These phases are not as cut-and-dry as the figure shows, because these devices are changing and growing over time. For example, a user may have long periods without discovery because they have not had time to spend exploring the developments available, only to have a friend show them a new use they might like, thus throwing them briefly from the establishing use phase to the experimenting phase if the user chooses to try the use out. This may lead to a period of experimenting with new uses and exploring new uses, then learning to implement the uses in their daily life, and finally routinizing the new uses. If the frustration expressed by the participants at not knowing the abilities or limits of these devices can be tempered by the need for patience with the changing abilities of the devices acknowledged by some of the participants, this may become more of an ongoing cycle rather than the current trend of linear progression through these phases.

9. What do users of smart speakers experience with respect to finding and managing information?

Overall, users seemed happy to have their smart speaker for the conveniences it offered and the possibilities for future use in terms of finding and managing information. Users primarily looked to the device for entertainment purposes (e.g., music), assistance (e.g., with calendar or timers), and information searches (e.g., satisfying their curiosity about a topic, or getting information about traffic or weather). However, many users expressed that they had become routine in their use of the device, and that their use of the device had not grown as much as they thought it would. In this suggestion, I provide two areas of focus where information architect designers might improve the user experience the most.

9.1. Easing the user's burden

Once themes were categorized, understanding the areas for improvement becomes easier. One of the key themes that overlapped all the themes above is that the burden for discovering uses, for learning the rules of use and understanding the limits and abilities, all that work relies on the user taking the initiative to do so. The burden of the user can be seen in the following areas:

- Fear of changing the status quo (e.g. with P18 and his speaker system)
- Laziness (not wanting to take on a project where the benefits would not be worth the time and energy put into it)
- Learning to speak to the device
- Learning the abilities and limits of the device
- Finding uses out of the many possibilities that are relevant to the individual user
- The users training themselves to remember to and remember how to use the device

These aspects of the user's burden are referenced in multiple interviews. In many cases, the participants themselves were happy to reflect on ways to improve their experience.

The first user-provided improvement was to allow the device to provide more intuitive assistance to the user. For example, many users responded positively to the idea of adding personalized suggestions from the smart speaker itself. To some extent, the Alexa app already does this, and the Echo will provide suggestions if asked. However, these suggestions do not ease the user's burden in a meaningful way. They are not personalized to the user's needs, and while the current suggestions for use might be fun, they do not contribute to easing any of the areas of burden above. The user will still have to put energy and time into learning the use, discovering whether the use is relevant to their needs or not, and training themselves to make that use routine. One caveat of this improvement is that, while most users were enthusiastic (or at least accepting) of receiving suggestions based on previous interactions with the device or based on access granted to use information from other apps or browsing data, nearly all users reacted negatively to the device listening to their conversations and responding without being addressed in the context. Users worried about losing control of their device, or about this interaction feeling "creepy" or "inappropriate," or for this interaction feature to be "misused."

Another example of improving intuitive assistance was to add functionality to the physical device itself—to some extent, this is already beginning to happen with the Echo, where a yellow ring around the top indicates that there are notifications for the user. However, this type of functionality means that the user must look at the device in order to notice that there are notifications. One user (P5) suggested that the device might even notice when someone has come home and, after greeting them, ask if the user wanted to hear their notifications. This would allow for ease of use in terms of alleviating the user's need to remember to ask the device for notifications.

The second user-provided improvement was to provide a more informative introduction to the smart speaker. This introduction at the time of this paper primarily includes a small paper included with the smart speaker which details a few phrases the user can try and then directs the user to the app for more information. While there is more information online both on dedicated Google Home and Amazon Echo support pages and on tech review sites, many of the interviewed users either left the navigation of discovery to a more enthusiastic user (e.g. their spouse) or just did not see the point in investing the time and energy it would take to discover new uses that were relevant beyond the basics (e.g. music, weather, news, timers, lists) that made their way into routine use. Nearly all users expressed an uncertainty in how to use their smart speaker, and a feeling that they were failing to use it to its full capability. A more informative introduction to the abilities of the smart speaker at the start of the user's experience might prove to help alleviate some of this uncertainty and instill in the user a confidence that would carry them through implementing some of the more difficult (or boring) uses that would, in the end, help the user engage with their device more frequently in order to make their own lives easier.

The third user-provided improvement involved releasing information in manageable chunks. Some participants expressed a frustration that they felt they had to spend a lot of time and effort browsing the web for relevant uses. While this could be partially alleviated by the personalization of suggestions mentioned above, the key here is that the users might feel overwhelmed by the number of apps or skills or applications for their smart speaker. Splitting these personalized suggestions up into manageable chunks, similar to the way the emails from Amazon about Alexa already do, would help the users be more inclined to read them—these information chunks would have a higher chance of containing at least one relevant use to carry forward into routine use, or at least help the user be more well-informed about the developments in progress for their device.

9.2. Understanding the importance of location

The motivations for seeking information can be divided to some extent into which space the device occupies. While many users will activate music, news, or other ambient audio in multiple rooms, certain spaces are associated with specific uses. Additionally, these uses might only occur at a certain time of day. As the devices develop and grow in their ability to process natural language, and in their abilities in general, they come closer to the goal of providing intuitive, non-invasive, personalized suggestions and interactions with the user. These suggestions and interactions can be informed by the device's location in the home: in the bedroom, informing the user there is an alarm clock feature or suggesting the user try to ask about the weather; or in the kitchen, asking if the user is interested in trying to have Alexa guide them through a recipe step-by-step. The location provides a point for the information architect designer around which to focus in order to produce an appropriate design for the space.

10. What might an information architecture designer do to improve user experience?

The goal of this project was to answer the research question: What do users of smart speakers experience with respect to finding and managing information, and what might an information architecture designer to do improve user experience? The previous section dealt with the experiences of the participants in terms of finding and managing information. This section deals with what an information architecture designer might do about improving user experience with the smart speakers.

The idea that users are responsible for knowing or discovering what to do with respect to their environment (or in this instance, with their smart speaker) is reflected in the concept of "affordances", which originated in Gibson (1979), where he provides the term "affordance," stating, "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" (p. 127). This term "affordances" was further developed by Norman (2002) (original edition published in 1988), specifying that, in the context of human-machine interaction, affordances referred to only possible actions readily perceived by an actor (e.g. a handle on a mug indicates that the handle can be used to assist the actor in lifting the mug).

Taking this concept of perceivable affordances one step further, Gaver (1991) introduced three categories of affordances: the perceptible affordance (e.g., a button that, when pressed, allows the actor to carry out an action); the false affordance (e.g., a button that does not, in fact, have any real function); and the hidden affordance (where there is no perceptual information indicating an existing affordance, and it must be inferred from other evidence). He also introduces the concept of sequential affordances, which he uses "to refer to situations in which acting on a perceptible affordance leads to information indicating new affordances" (Gaver, 1991, p. 82)—one example given by Gaver is that of door handles—it is perceivable that the door handle can be grasped, and once grasped, turned, and once turned, used to open the door, each affordance revealing itself in sequential order. The affordance of opening a door is not necessarily perceived immediately, and perhaps the user only discovers this affordance after an exploratory or random push or press which activates the handle's movement.

The idea of affordances as applied to smart speakers matches well with the desires of the users to have their device be intuitive and personalized. Let me lay out a scenario where, for simplicity's sake, the user in has an integrated home and a set of Amazon Echo devices.

In the morning, an Echo Dot in the bedroom uses the alarm to wakes them at 7:00. They ask their assistant to turn on the bedroom lights. Then, as the user dresses for their day, they can ask about the weather in order to determine if a long-sleeve shirt and pants would be better suited to the temperature outside than the short-sleeved dress they had planned on wearing. The user might also have their calendar synced with their assistant, and they ask what time their first appointment is, only to discover that their first appointment has been cancelled and they do not have to come to the office for another hour.

They now have time to go into the kitchen, make breakfast ("Alexa, how many teaspoons in a tablespoon?" they ask, as they make cinnamon-sugar French toast), listen to the news or their favorite radio station, and before they leave for work, they ask their assistant to turn the temperature down on their thermostat. While they are gone, nothing much happens with the device. However, they have programmed it to turn the temperature on the thermostat up/down (depending on the weather) at 16:30, 15-30 minutes before they get home, to let them come back to an appropriately warm or cool home.

As they walk in the door after their busy day at work, the Echo in their main room greets them: "Good afternoon. While you were gone, you received two notifications. Would you like to hear them?" When they respond, "Yes," Alexa reads their notifications aloud, allowing them to react when necessary (if, for example, a package shipped, and they want to track it). Once the notifications are done, Alexa asks if the user would like their usual after-work playlist to start. The user says no, because they know they have a little bit of work left to finish at home.

Once their work is done, they ask Alexa to turn on Netflix and pick a show. Unfortunately, none of the shows catch the user's attention, so they decide to experiment with Alexa instead. "Alexa, what's new?" they ask.

"Well, I noticed you installed the FitBit app on your phone. Would you like me to turn on Echo syncing and integration features for your FitBit?" says Alexa.

"Hm. What are the features?" queries the user.

"If you open the Alexa app on your phone, I can take you to the skill page," offers Alexa.

"Okay, sure," replies the user.

Then, the user can browse at will all the new features that FitBit has integrated with Alexa: the assistant can tell the user how well they slept after the alarm wakes them, it can tell them to get up and move if they have been sedentary for an extended period, and more. They activate the skill and connect with their FitBit activity tracker.

After this helpful interaction, the user is left with a sense of success with discovering a relevant new use for Alexa without having to spend too much time or effort finding or installing it.

"The End"

In the scenario above, the user had no problems of being misunderstood by their device, they did not have to spend hours exploring apps or websites to find new uses that were relevant to their life, and the device provided intuitive help to access notifications and new uses. While I did not show an introduction to the device, there is no reason it could not be as easy as the interactions in the above scenario.

This project's goal was to provide suggestions of design guidelines for information architects when approaching smart speaker-user interaction. Here, I lay out a set of guidelines, as drawn from the analysis of participant data and from results of papers also focused on the design of IPAs:

1) The users must feel understood and heard by the device.

This is developed through successful interactions with the device, where the device could allow for repeating and refining their query (Porcheron, Fischer, & Sharples, 2017) to help the user find and manage information more easily, and reduce frustrations based on usability breakdowns. Weizenbaum's reflections on how people were willing to allow more error from their anthropomorphized partner after feeling understood and heard might also provide some area for reflection here.

2) The users must trust the device.

In order to personalize the device to a point where intuitive search aids such as auto-complete are possible, the user must trust the device. The building of trust is a tricky business, but it can be helped by the devices being more consistent in the perceived quality of their responses to users' queries. Looking at Vtyurina et al.'s work might help with creating more consistent responses to user's information searches.

3) There must be minimal breakdowns in understanding between user and device.

This idea of avoiding breakdowns is mentioned in Winograd and Flores, where they state that breakdowns undermine a system's usability. Users in this study were frustrated with breakdowns in understanding between them and their smart speaker, and reducing this frustration would allow for more motivation to explore new uses and deepen exploration and understanding of the device.

4) Users must be involved in the process of designing information architecture.

This is the viewpoint that guided this project, emphasized by Suchman earlier in this paper. It is difficult to understand and build from user's frustrations if users are not involved in the process of designing the information architecture. Interviewing users is the first step, but I suggest that designers of information architecture approach their work from a thoughtful design standpoint. In a book by Löwgren and Stolterman (Thoughtful Interaction Design, 2007), they propose the idea of the thoughtful designer, which is a designer who reflects on the contexts and consequences of the tools they design. This concept of being a thoughtful designer lines up well with Suchman's point that the programmers and designers must consider the politics of the systems they build. These authors also state that, even though ceratin aspects of a digital artifact might be independent of the context, the most crucuial qualities are almost always deeply context-dependent. Part of this is that the artifact is and must be judged according to the users' competences and skills, and that "good" design is always relative to political, ethical, legal, and societal contexts. This is a very good justification for designing together with users, as only users can show what their competences and skills are. A workshop would help bring light to the users' competences and skills, and what they might value in an interaction.

One version of thoughtful, user-inclusive design is participatory design. Muller and Druin (Muller & Druin, 2012) define participatory design as "a set of theories, practices, and studies related to end users as full participants in activities leading to software and hardware computer products and computer-based activities" (Muller & Druin, 2012, p. 1125). A key part of Muller and Druin's contribution to the concept of participatory design is that there are three spaces in participatory design: the first is the perspective of the users, and their environment; the second is the perspective and environment of the developers (designers); and the third is a blend of the two which allows new relationships and new understanding to emerge. This is a useful concept to take forward in any workshop to design information architecture for these smart speakers, as the designers and users will likely have very different understandings of how a smart speaker and IPA might fit into the user's life.

In terms of future research, there are many directions that information architects can explore. Developing the physical appearance of the device or the accompanying apps to indicate affordances and allow for fewer "hidden" affordances could help alleviate the burden on the user to hunt for new applications. Also, the idea of making the introduction to the device more like a game, training the user in a more entertaining way, could work well with the fact that many of these users bought their smart speakers as "toys" or "gadgets." Regardless, there is much more work to be done in terms of developing how much the users explore and use their smart speaker, and this paper is arriving at what I am sure is just the beginning of the development of smart speakers and research surrounding them.

11. Conclusion

This project focused on the users' experience with finding and managing information, and how information architecture designers could address this. After interviewing users, I found that users enjoyed using their device for the most part, but that they were frustrated by a significant burden placed on them to learn about using their device, to discover new uses for their device, and to understand the limitations and possibilities provided by the smart speaker and IPA. From this user data and from results of papers in the literature review, I outlined some guidelines for designing a successful information architecture: (1) the users must feel understood and heard by their device; (2) the users must trust the device; (3) there must be minimal breakdowns in understanding between the user and device; (4) users must be involved in the process of designing information architecture.

I concluded that thoughtful, user-inclusive design would allow information architect designers to better address the frustrations of the user in terms of information finding and management. Participatory design allows designers to understand the worldview and perspective of the users, which provides designers with the knowledge and understanding of the users required to create a program which would be most likely to permit the development of common ground with as few breakdowns in understanding as possible. After all, one of the reasons information architecture is so vital is that it supports a user's information needs and actions based on that need in an information space. In the case of smart speakers and intelligent personal assistants, I have concluded that this is best approached by having users participate in the design process.

References

Amazon. (2015, June 23). Amazon Echo - Now Available. Retrieved May 23, 2017, from https://youtu.be/FQn6aFQwBQU

Amazon Lex. (2017). (Amazon) Retrieved May 20, 2017, from Amazon Web Services: aws.amazon.com/lex

Bloor, M. (1997). Addressing social problems through qualitative research. In D. Silverman (Ed.), *Qualitative Research* (1st ed.). London: Sage.

Bødker, S., & Christiansen, E. (2012). Poetry in motion: appropriation of the world of apps. *Proceedings of the 30th European Conference on Cognitive Ergonomics - ECCE '12* (pp. 78-84). New York, NY: ACM. doi:10.1145/2448136.2448152

Bødker, S., & Klokmose, C. N. (2011). The Human–Artifact Model: An Activity Theoretical Approach to Artifact Ecologies. *Human-computer interaction, 26*(4), 315-371. doi:http://dx.doi.org.zorac.aub.aau.dk/10.1080/07370024.2011.626709

Bonsch, A., Vierjahn, T., Shapiro, A., & Kuhlen, T. W. (2017, March). Turning anonymous members of a multiagent system into individuals. *Virtual Humans and Crowds for Immersive Environments* (pp. 1-4). IEEE.

Braun, V., & Clarke, V. (2006, January 01). Using thematic analysis in psychology. *Qualitative Research in Psychology, 2*, pp. 77-101. doi:10.1191/1478088706qp063oa

Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford, United Kingdom: Oxford University Press. Retrieved 2017

Butow, E. (2007). *User interface design for mere mortals.* Pearson Education.

Cicció, J. A., & Quesada, L. (2018). Framework for Creating Audio Games for Intelligent Personal Assistants. In F. C. Ahram T. (Ed.), *Advances in Human Factors in Wearable Technologies and Game Design. AHFE 2017. Advances in Intelligent Systems and Computing. 608.* Springer, Cham. doi:https://doi.org/10.1007/978-3-319-60639-2_21

Clark, H. H. (1985). Language use and language users. In G. Lindzey, & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 179-231). New York: Harper and Row.

Clark, H. H., & Wilkes-Gibbs, D. (1986, February). Referring as a collaborative process. *Cognition, 22*(1), 1-39. doi:10.1016/0010-0277(86)90010-7

Druga, S., Breazeal, C., Williams, R., & Resnick, M. (2017). "Hey Google is it ok if I eat you?" Initial explorations in child-agent interaction. *IDC 2017 - Proceedings of the 2017 ACM Conference on Interaction Design and Children* (pp. 595-600). Association for Computing Machinery, Inc.

Extending the Google Assistant. (2017, May 17). (G. Developers, Producer, & Google) Retrieved May 20, 2017, from Actions on Google: https://developers.google.com/actions/extending-the-assistant

Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988, April). Computer systems and the design of organizational interaction. *ACM Transactions on Information Systems (TOIS), 6*(2), 153-172. doi:10.1145/45941.45943

Gaver, W. W. (1991). Technology affordances. *Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91*, (pp. 79-84). doi:doi: 10.1145/108844.108856

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Google. (2016, October 4). Google Home: Hands-free help from the Google Assistant. Retrieved from https://youtu.be/nWiIWyCeZso

Griol, D., Molina, J. M., & Callejas, Z. (2017). Incorporating android conversational agents in m-learning apps. *Expert systems , 34*(4). doi:10.1111/exsy.12156

Herbert H. Clark. (2017, January 8). Retrieved May 23, 2017, from Wikipedia: https://en.m.wikipedia.org/wiki/Herbert_H_Clark?wprov=sflal

IBM. (n.d.). *IBM archives: 1980s*. Retrieved from IBM: http://www-03.ibm.com/ibm/history/history/year_1987.html

Isaacs, E. A., & Clark, H. H. (1987, March). References in conversation between experts and novices. *Journal of Experimental Psychology, 116*(1), pp. 26-37. doi:http://dx.doi.org/10.1037/0096-3445.116.1.26

Kuutti, K. (1996). Activity Theory as a potential framework for humancomputer. In B. A. Nardi (Ed.), *Context and consciousness : activity theory and human-computer interaction* (pp. 17-44). Cambridge: MIT Press.

Leont'ev, A. (1981). *Problems of the development of the mind.* Moscow: Progress.

Lester, J., Branting, K., & Mott, B. (2004). Conversational Agents. In M. P. Singh, *The Practical Handbook of Internet Computing (Chapman & Hall/CRC Computer and Information Science Series)* (pp. 1-17). CRC Press. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.9557&rep=rep1&type=pdf

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: a guide to qualitative observation and analysis* (3rd edition ed.). Belmont, CA: Wadsworth.

Löwgren, J., & Stolterman, E. (2007). *Thoughtful Interaction Design.* MIT Press.

Luger, E., & Sellen, A. (2016, May 7-12). 'Like having a really bad PA': The gulf between user expectation and experience of conversational agents. *CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286-5297). New York: ACM. doi:10.1145/2858036.2858288

Martin, E. (2017). How echo, Google home, and other voice assistants can change the game for content creators. *EContent*.

Moggridge, B. (2006). *Designing Interactions.* (B. Moggridge, Ed.) Cambridge, Mass., USA: MIT Press. Retrieved May 07, 2017

Müller, J., Polansky, D., Novak, P., Foltin, C., Polivaev, D., & et al. (n.d.). FreeMind (Version 1.0.1) [Computer software]. Retrieved from http://freemind.sourceforge.net/wiki/index.php/Download

Muller, M. J., & Druin, A. (2012). Participatory design: The third space in human-computer interaction. In J. A. Jacko, *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition* (3rd, illustrated ed., pp. 1125-1153). CRC Press.

n/a. (2017). *Usability*. Retrieved 2017, from Interaction Design Foundation: https://www.interaction-design.org/literature/topics/usability

Noda, K. (2017, August 22). Google Home: smart speaker as environmental control unit. *Disability and Rehabilitation: Assistive Technology*, 1-2. doi:10.1080/17483107.2017.1369589

Norman, D. (2002). *The design of everyday things* (Elektronisk udgave ed.). New York: Basic Books.

Porcheron, M., Fischer, J. E., & Sharples, S. (2017). "Do animals have accents?": Talking with agents in mutli-party conversation. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17* (pp. 207-219). New York NY: ACM. doi:10.1145/2998181.2998298

Porcheron, M., Fischer, J., Mcgregor, M., Brown, B., Luger, E., Candello, H., & O'Hara, K. (2017). Talking with Conversational Agents in Collaborative Action. *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 431-436). New York, NY, USA: ACM. doi:10.1145/3022198.3022666

Purington, A. T., Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & TAylor, S. H. (2017). "Alexa is my new BFF": Social roles, user satisfaction, and personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (pp. 2853-2859). Association for Computing Machinery. doi:10.1145/3027063.3053246

Resmini, A., & Rosati, L. (2011). *Pervasive Information Architecture: Designing Cross-Channel User Experiences.* Burlington, MA: Morgan Kaufmann.

Rosenfeld, L., Morville, P., & Arango, J. (2015). *Information Architecture for the Web and beyond* (4th ed.). O'Reilly Media.

SmartMob. (n.d.). Smart Voice Recorder (Version 1.8.0) [Mobile application software]. Retrieved from https://play.google.com/store/apps?hl=en

Suchman, L. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication* (illustrated, reprint ed.). Cambridge University Press.

Suchman, L. (1995). Speech acts and voices: Response to Winograd et al. *Computer Supported Cooperative Work, 3*, 85-95.

Vtyurina, A., Savenkov, D., Agichtein, E., & Clarke, C. L. (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* (pp. 2187-2193). ACM Press. doi:10.1145/3027063.3053175

Webb, E., Campbell, D., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures; Nonreactive measures in the social sciences.* Chicago: Rand McNally.

Weiser, M. (1998, January 1). The future of ubiquitous computing on campus. *Communications of the ACM, 41*(1), pp. 41-42. doi:10.1145/268092.268108

Weizenbaum, J. (1966, January). ELIZA--A computer program for the study of natural language communication between man and machine. (A. G. OETTINGER, Ed.) *Communications of the ACM, 9*(1), 36-45.

Weizenbaum, J. (1976). *Computer power and human reason.* San Francisco: W.H. Freeman.

Wilkes-Gibbs, D., & Clark, H. H. (1992, April). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*(2), pp. 183-194. doi:https://doi.org/10.1016/0749-596X(92)90010-U

Winograd, T. (1987-88). A language/action perspective on the design of cooperative work. *Human-Computer Interaction, 3*(1), 3-30. Retrieved from https://hci.stanford.edu/winograd/papers/language-action.html

Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition: A new foundation for design.* Norwood, New Jersey, USA: Ablex.

Appendix

Consent for Participation in Interview Research

I, _____, volunteer to participate in a research project conducted by Bailey Smith, a student from Aalborg University. I understand that the project is designed to collect data about my use of the Google Home and/or the Amazon Echo smart speaker device in my home. Bailey Smith is collecting interviews of smart speaker users in order to identify key challenges faced by information architects when designing for these devices, in order to complete her Master's thesis in Information Architecture.

1. My participation in this project is voluntary. I understand that I may receive some small compensation in return for participation. I may withdraw and discontinue participation at any time without penalty.

2. I understand that if I feel uncomfortable in any way during the interview session, I have the right to decline to answer any question or to end the interview.

3. Participation involves being interviewed by a student from Aalborg University. The interview will last approximately 30-60 minutes. Notes will be written by the interviewer during the interview. An audio recording of the interview and subsequent dialogue will be made. Photographs will be taken by me of the device's location in my home. If I do not want to be recorded, I will not be able to participate in the study.

4. I understand that the researcher will not identify me by name in any reports using information obtained from this interview, and that my confidentiality as a participant in this study will remain secure unless I give explicit permission for this below.

5. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.

6. I have been given a copy of this consent form, signed by the investigator.

_____ _____

Participant's Signature Date

_____

Signature of the Investigator

Figure 13 Consent form for interviews