

École Normale Supérieure de Lyon
Sciences humaines et sociales
Information et communication
Master Architecture de l'information

Master's Thesis

User Experience of Voice Assistants: Exploration of Anticipatory Assistance
with Augmented Sensitivity to Usage Environments' Context

by

Ka Hei Cheung*

May 2018

supervised by

Isabelle Sperano
MacEwan University

and co-advised by

Charlotte Block
denkwerk GmbH

Sandra Griffel
denkwerk GmbH

© Ka Hei Cheung 2018

* ka-hei.cheung@ens-lyon.fr

ABSTRACT

Nowadays, voice assistants are being introduced to a growing number of usage environments. However, their performance is not fully adapted to these various environments' context. This has hampered users' interaction with the assistants and thwarted an inclusive experience. Furthermore, voice assistants' development has generated services which are not always tuned to users' needs and expectation. This has endangered the role of voice assistants and their utility. Thus, there is a necessity to weigh users back in the process to create appropriate experience. In this thesis, users, usage environments and the role of voice assistants pedestal the user experience perspective. First, current voice assistant users' interaction with voice assistants is studied; their experience perceived through such interaction is investigated; improvable elements within the interaction are also analysed. Second, potential ways of augmenting voice assistants' capabilities are explored; experiments are conducted on the experience design around such augmented voice assistants while considering holistically users' aspect, voice assistant development, and usage environments. Finally, through engaging with current voice assistant users and taking a collaborative design approach, concepts related to augmenting voice assistants' performance through anticipatory assistance are proposed for enriching users' experience; prototypes focusing more specifically on the commute environment are created.

Keywords: voice assistant, voice user interface, user experience, usage environment, anticipatory assistance

RÉSUMÉ

Aujourd'hui, les assistants vocaux sont introduits dans un nombre croissant d'environnements d'usage. Toutefois, leur performance n'est pas totalement adaptée au contexte de ces divers environnements. Ceci a entravé l'interaction des utilisateurs avec les assistants vocaux et contrecarré une expérience inclusive. De plus, le développement des assistants vocaux a généré des services qui ne sont pas toujours conformes aux besoins ni attentes des utilisateurs. Ceci a mis le rôle des assistants vocaux et leur utilité en danger. Il existe ainsi une nécessité d'accentuer la place des utilisateurs dans le processus pour créer de l'expérience appropriée. Dans ce mémoire, les utilisateurs, les environnements d'usage et le rôle des assistants vocaux forment le socle de la perspective d'expérience-utilisateur. Dans un premier temps, l'interaction entre les utilisateurs actuels et les assistants vocaux a été étudiée ; leur expérience perçue à travers cette interaction a été examinée ; les éléments améliorables de celles-ci ont été analysés. Dans un second temps, les opportunités de design des assistants vocaux ont été explorées ; des expérimentations ont été conduites sur le design d'expérience autour des assistants vocaux augmentés en prenant compte de manière holistique l'aspect des utilisateurs, le développement des assistants vocaux et les environnements d'usage. Enfin, en engageant les utilisateurs actuels des assistants vocaux et en employant une méthode de design collaboratif, des concepts favorisant l'augmentation de la performance des assistants vocaux à travers l'assistance anticipative ont été proposés afin d'enrichir l'expérience-utilisateur ; des prototypes spécifiques autour de l'environnement de transport ont été créés.

Mots-clefs : assistant vocal, interface vocale, expérience-utilisateur, environnement d'usage, assistance anticipative

TABLE OF CONTENTS

ABSTRACT	i
RÉSUMÉ	ii
TABLE OF CONTENTS	iii
ACKNOWLEDGEMENTS	vi
1 INTRODUCTION	1
2 RESEARCH PROBLEM	2
2.1 A Phenomenon to Address.....	2
2.2 An Agency to Augment	2
2.3 A Development to Humanise and Make Sustainable	3
2.4 A Role to Rethink	3
2.5 Research Objectives and Questions	4
3 STATE OF THE ART	5
3.1 Voice Assistants.....	5
3.1.1 <i>Definition of Voice Assistants</i>	5
3.1.2 <i>Services of Voice Assistants</i>	6
3.1.3 <i>Working and Interaction Mechanism of Voice Assistants</i>	7
3.2 Usage Environments of Voice Assistants	8
3.2.1 <i>Definition of Usage Environment</i>	8
3.2.2 <i>Existing Usage Environments</i>	8
3.2.3 <i>Usage Environments Sensitivity</i>	10
3.3 Interaction with Voice Assistants	10
3.3.1 <i>Interaction as in Human-Computer Interaction</i>	10
3.3.2 <i>Interaction as in Voice User Interface Design</i>	11
3.3.3 <i>Limitations of Voice Interaction</i>	12
3.3.4 <i>Potentials of Voice Interaction</i>	13
3.4 Experience with Voice Assistants.....	14
3.4.1 <i>Accessibility of Applications and Connected Services</i>	15
3.4.2 <i>Accuracy in Processing Utterances</i>	16
3.4.3 <i>Affectional Influence</i>	17
3.4.4 <i>Anticipation</i>	17
4 METHODOLOGY	19
4.1 Questionnaire	19
4.1.1 <i>Contents of the Questionnaire</i>	19

4.1.2	<i>Model of the Questionnaire</i>	20
4.2	User Interview	21
4.3	Participatory Workshop / Co-Design Workshop	22
4.3.1	<i>Profiles of Workshop Participants</i>	22
4.3.2	<i>Structure and Activities of Workshop</i>	23
4.4	Experience Blueprinting	29
5	RESULTS AND DISCUSSION	30
5.1	Questionnaire	30
5.1.1	<i>Responses</i>	30
5.1.2	<i>Insights</i>	30
5.2	Co-Design Workshop	31
5.2.1	<i>Elements</i>	31
5.2.2	<i>Use Cases</i>	32
5.2.3	<i>Prototypes</i>	33
5.3	Experience Blueprinting	37
5.4	Concerns and Highlights	39
5.4.1	<i>Concerns of Applied Study Partakers</i>	39
5.4.2	<i>Anticipatory Assistance and Voice User Interface</i>	40
5.4.3	<i>Assistance and User Profile</i>	40
5.4.4	<i>(Anticipatory) Assistance and Cultural Environment</i>	41
5.4.5	<i>Voice Assistants and Anticipation</i>	41
5.4.6	<i>Voice Interaction and Experience Design</i>	41
6	CONCLUSION	42
	APPENDICES	43
	Appendix A Alexa-Enabled Devices	43
	Appendix B Alexa Assistance Service	44
	Appendix C Alexa’s Working Mechanism	45
	GLOSSARY	46
	BIBLIOGRAPHY	48

LIST OF TABLES

Table 1 Categories of Alexa-Enabled Devices (Gordon and Wroclawski, 2018; McGrath and Hastings, 2018)	43
Table 2 Categories of Alexa’s assistance functions (Johnson, ‘Everything Amazon’s Alexa Learned to Do in 2017’, 2017; Activate, 2017)	44

LIST OF FIGURES

Figure 1 Radziwill and Benton’s (2017) classification of software-based Dialogue Systems	5
Figure 2 Example of anticipatory assistance by augmented voice assistant (ordering pizza delivery)	24
Figure 3 Dot-voting for identifying the how-might-we problem-to-solve	25
Figure 4 Alternatives created in Crazy 8s on the only Exquisite Corpse combination (left) and three-frame Storyboarding (right)	26
Figure 5 Gallery Board—Extraction of Snippets from Use Cases to form Scenarios and User Profiles	28
Figure 6 Participants deciding the blocks of actions (left) and filling these blocks of actions with dialogues (right)	29
Figure 7 Prototype B covers the whole school day and tends to help teenage users integrate to their new school	34
Figure 8 Prototype A focuses mainly on the period before school and on preparing the teenage users for their school day	35
Figure 9 Educational Voice Assistant (partial) utilising the time before the commute to prepare teenage users for their school day	37
Figure 10 Educational Voice Assistant inviting teenage users to bear responsibility with transport service delay notice	38
Figure 11 Discovery Voice Assistant suggesting after school activities in the neighbourhood (alternatively proposing museum visit, a different route to go home, and grocery shopping on the way)	38
Figure 12 Alexa’s Working Mechanism (Doring and Srivastava, 2017)	45

ACKNOWLEDGEMENTS

First, I would like to thank my thesis supervisor **Dr Isabelle Sperano** of Faculty of Fine Arts and Communications at MacEwan University for generously sharing her knowledge and providing guidance to me. I greatly appreciate her kindness in always maintaining an open door and making her time available for steering me in the right direction whenever I knock.

I would also like to thank my co-advisers **Ms Charlotte Block** and **Ms Sandra Griffel** of *denkwerk GmbH* for sacrificing their time and attention to my study and assisting me with their expertise. I am grateful for their investment of ideas which has enriched my personal project.

I would like to thank all the partakers of this study, my professors, my friends, my classmates, and all the professionals whose participation has—directly or indirectly—contributed to the contents and reflection upon them in this thesis.

Finally, I would like to express my profound gratitude to **Mr Ricardo Dias** for his continuous patience and understanding, steadfast support, and unconditional encouragement; and to my parents who have always supported me for my higher pursuits. This accomplishment would not have been possible without them.

1 INTRODUCTION

Voice assistants are virtual software **agents** (Radziwill and Benton, 2017). They are programmed to perform tasks or provide services over **voice commands** and/or **dialogues** (Radziwill and Benton, 2017; Pearl, 2016, pp. 1–70). They are often made for public groups but bear interaction with individual users (Porcheron *et al.*, 2017; Reeves, 2017). Mostly, they are powered by **artificial intelligence (AI)** (Canbek and Mutlu, 2016; Radziwill and Benton, 2017; Sharma, 2016; Falkner, 2018) and are accessed through certain portal hardware devices (Canbek and Mutlu, 2016; Amyx, 2014; Rowinski, 2013; Falkner, 2018).

Voice assistants are a pluridisciplinary topic. It crosses, in particular, the fields of Human-Computer Interaction (HCI) and Communication (Knight, 2016; Berg, 2014, pp. 20–35). The development of voice assistants is closely related to the advancements of such fields.

Lately, voice assistants have been making headway in development progress. **Applications**¹ are on a burgeoning number to offer sundry services; hardware devices, too, are expanding swiftly to make voice assistants accessible through various forms of portal and in different kinds of **usage environments**² (Perez, ‘Amazon’s Alexa Passes 15,000 Skills, up from 10,000 in February’, 2017; Kinsella, 2018; VoiceLabs, 2017; Lacoma, 2017; Perez, ‘Alexa Is Coming to Wearable Devices, Including Headphones, Smartwatches and Fitness Trackers’, 2018; Hetzner, 2017); as a result, this has attracted groups of adopters who are becoming diverse and significant in number.

While the development of voice assistants depends greatly on HCI and Communication on the one hand, the growing users are turning into a stark weight on the other. This creates a practical interest to examine the subject from a User Experience (UX) perspective—to find out what voice assistant users appreciate and desire, to discover potentials in the current and appearing usage environments, to maximise the experience of their assistance services, and to humanise the development of voice assistants.

In this thesis, I shall first describe in CHAPTER 2 RESEARCH PROBLEM the observed happenings around voice assistants and demonstrate the reason why a UX probe is necessary into the topic. Then, I shall present how user experience is related to voice assistants in CHAPTER 3 STATE OF THE ART by defining voice assistants, discerning the interaction, and dissecting the experience. After that, I shall describe my chosen methods in CHAPTER 4 METHODOLOGY for my applied study. In the end, my findings are explained and issues which are worth the attention are raised in CHAPTER 5 RESULTS AND DISCUSSION.

Since the methods adopted in this research are iterative—the results of each step of the chosen methods had an effect on the subsequent steps; often are the succeeding methods evaluated and adjusted before being continued—it is *recommended* to refer to the relevant SECTIONS in CHAPTER 5 while reading CHAPTER 4. For convenient navigation, cross-reference links have been added to these SECTIONS.

N.B. There can be a nuance between meanings of words in a general, social context and voice assistant vocabulary. When the nuance exists, in this thesis, words are left plain when they refer to the former, capitalised when they refer to the latter. For instance, ‘utterance’ refers to a spoken work or a voiced statement, whereas **Utterance** refers to a recognised text phrase transformed from the streamed **voice command** or **request** by **Automatic Speech Recognition (ASR)**. For clearer understanding, particular words which can carry specific meaning are put in bold type at their first appearance and defined in the glossary at the end of this thesis. Stresses and verbatim quotes are put in italic type. Words are put in small caps when they refer to the content structure of this thesis (e.g. SECTION).

¹ A **voice assistant application** is often an additional software program which can enhance voice assistants’ services. This is often done manually by the users.

² (**Usage**) **environment** refers to the space in which voice assistants are being accessed. For more details, (see SECTION 3.2 Usage Environments of Voice Assistants below).

2 RESEARCH PROBLEM

Voice assistants have been a subject across several disciplines; notably, it has been extensively studied in Human-Computer Interaction (HCI) regarding the action and influences between humans and machines, and in Communication concerning the exchanged information and language (Radziwill and Benton, 2017; Berg, 2014, pp. 20–70). Lately, there are also more publications appearing on the pedestal of **Voice User Interface Design** which focus mainly on voice assistants' usability (Harris, 2005; Pearl, 2016). The analyses of voice assistants with the lenses of User Experience tend to appear, however, less abundant and united. Yet, from the observation of happenings evolving around voice assistants, a growing significance appears to point towards User Experience Design, with respect to the link that voice assistants and users create during interaction and the experience perceived by users of such products. These happenings could be summed up in four aspects as follows.

2.1 A Phenomenon to Address

The proliferation of supporting devices (**voice assistant-enabled devices**) has raised the availability of voice assistants and diversified their possible **usage environments** (Perez, 'Amazon's Alexa Passes 15,000 Skills, up from 10,000 in February', 2017; Kinsella, 2018; VoiceLabs, 2017; Lacoma, 2017; Perez, 'Alexa Is Coming to Wearable Devices, Including Headphones, Smartwatches and Fitness Trackers', 2018; Hetzner, 2017). In the ideal situation where the voice assistants could perform fully to its capacity, both the hardware and the software part ought to complement each other. However, this is often not the case in the reality (O'Donnely, 2016); whilst voice assistants are being introduced into new enabling devices at a rapid rate, the software side of the assistants themselves progress more slowly. This generates a risk which voice assistants can become incompatible with the devices through which their assistance is meant to be accessed and offered to users—in other words, users can manoeuvre voice assistants in some new environments but the voice assistants behave indifferently to the environment context. An example is the introduction of Amazon Alexa into iOS and Android devices, which allows the voice assistant to become accessible in a mobile environment, synonymous to any places where the user has Internet connection at their disposal (Perez, 'Amazon Adds Alexa to Its Main Shopping App', 2017; Perez, 'Alexa Is Coming to the Amazon App on Android, Starting This Week', 2017). Nevertheless, when the assistant is asked for restaurant suggestions near where the user is, since the voice assistant does not support real-time location detection, it still relies on the address reference point which the user had registered in his/her Amazon account (Johnson, 'Amazon Alexa Skills Can Now Access User Location for Services Such as Food Delivery', 2017). Thus, the given suggestions would not be fully accurate. This creates a certain gulf between what the hardware support aims and what voice assistants could offer. This gulf could fail users' expectation, undermine the repute of 'smart, personal assistants', and eventually jeopardise the usefulness of **artificial intelligence (AI)**.

Besides the individual **extended usage environments**, it has been observed that a significant group of users access voice assistants in multiple **environments** and/or bring their voice assistant-enabled devices across different **usage environments** (Perez, '42 Percent of Smart Speaker Owners Have Bought a Second Device [or More]', 2017; Gaines, 2017; Wollerton, 2016; Goode, 2017). The compatibility issue is, therefore, not only a dissonance between the hardware and software development of voice assistants to equilibrate, but also a users' behavioural need to address.

2.2 An Agency to Augment

As the number of usage environments in which voice assistants can be reached increases, by both the proliferation of supporting devices and their portability (Lacoma, 2017; Perez, 'Alexa Is Coming to Wearable Devices, Including Headphones, Smartwatches and Fitness Trackers', 2018),

allowing higher **sensitivity** to **usage environments** can improve voice assistants' **agency**³ by adapting their assistance to the environment in use (McMillan *et al.*, 2015; Chachada and Kuo, 2014); relevancy, coherence and utility of their artificial intelligence aid for users can hence be raised. This could give us a wide field of imaginable benefits—for example, voice assistants could understand more the situation context in the usage environment and offer more suitable and holistic assistance service (Novak *et al.*, 2016; Lyonais, 2017), smoothen the **dialogue** interaction with users by *anticipating* and making use of the relevant environment context (Pearl, 2016, pp. 195–204; Johnson, 'Alexa and Google Assistant Should Tell You When the next Bus Is Coming', 2017), adapt their behaviour regarding environment context and users' preference (Turunen and Hakulinen, 2000; Turunen, 2001), and even create a more engaging and surprising experience with their users by showing that they can actually 'understand' and perform *for* them (Moorjani, 2018). Furthermore, unlocking this sensitivity could also mean a springboard for development of other **applications**; this leads us to envision the possibility of enabling various potentials where usage environment context can take part as a perimeter in providing services, thus bringing extra value added to voice assistants.

2.3 A Development to Humanise and Make Sustainable

Although the development of supporting hardware can be a mutual co-operation between voice assistant companies and third-party manufacturers, the development of the assistant AIs still remains fairly reserved to the voice assistant companies (VoiceLabs, 2017; Activate, 2017). Application developers often have to accommodate themselves to the ways which voice assistant companies have paved and build their services around them (see SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants below and Appendix C Alexa's Working Mechanism). To use voice assistants and access desired services, end users have to further depend on what application developers have constructed (Anders, 2017; Reeves, 2017; Earley, 2016). This cause-and-effect interlinkage has created a hierarchy which subjects users' experience with voice assistants to a top-down provision, which does not always ensure the best and/or correct services and/or experience for users (in terms of voice assistant interaction performance and the experience offered to users; see SECTIONS 3.3 Interaction with Voice Assistants & 3.4 Experience with Voice Assistants below). This can lead us to ponder over the helpfulness of the User-Centred Design approach to address this issue, find out how to augment voice assistants' agency in their users' favour, and invert the chain by starting at the end with them.

Conducting an independent user-centred study on voice assistants can allow reviewing current voice assistants' services and learning any missed issues concerning their services in their usage environments. By finding out what users' current delight and frustration are, and hence by improving voice assistants' functions and services regarding those aspects, the performance could be raised and the tie between users and voice assistants could be strengthened. Returning to users could also lead to insights into methods to resolve problems and into new opportunities for service expansion. Furthermore, bringing users in the design process lets them have their say in what they use, and even co-construct their own *experience* while bearing no business burden to conform. This can help guarantee a better usability, accessibility, and enjoyability of interacting with voice assistants in the environments where they use them, promise better the experience demand, return the development of voice assistants to users—while opening it up and supporting a better sustainability of use.

2.4 A Role to Rethink

In addition to User-Centred Design approach, the role of voice assistants is also worth the contemplation. While voice assistants are being expanded to different hardware devices and usage environments, some introductions have become gasping—if not absurd. To name a few examples,

³ **Agency** is defined as the ability to act *effectively* with respect to goals (Novak *et al.*, 2016). See SECTION 3.3.1 Interaction as in Human-Computer Interaction below for further details.

voice assistants have been brought to bathroom environment (via mirrors, toilets, and showers) (Crook, 2018); to baby and pet care environment (via dietary devices and baby monitors) (Kelly, 2018); even to garden environment to offer interaction and/or experience whereof the intention and benefit for users are unclear (e.g. on parasols) (Kelly, 2018). The necessity of voice assistants on these devices and in these environments have been raised as a question (Baig, 2018; Kelly, 2018).

Contrary to the ‘*dot-com* bubble’ strategy⁴, we could ask whether the insertion of voice assistants by manufacturers and voice assistant companies is purposeful. As Johnson (‘Alexa and Google Assistant Should Tell You When the next Bus Is Coming’, 2017) instates, voice assistants do not have to do everything in users’ lives but help them in the necessity in a smooth way. This can be realised by relating users and usage environments where voice assistants are available, and by instantiating voice assistance to the particular tripartite combinations. By correlating the three components of the service of voice assistants, i.e. users, usage environments, and the role of voice assistants, a more *holistic* evaluation and consideration could be set for designing the assistants (Lyonnais, 2017). Moreover, by exploring, rethinking about, and breaking down current usage environments, potentials for possible inclusive services which voice assistants could offer to their users could be found out (Novak *et al.*, 2016; Lyonnais, 2017); the experience through interaction could also similarly be adapted and enriched by *anticipation* (Moorjani, 2018).

2.5 Research Objectives and Questions

In the demonstration of current problems with voice assistants, the importance of resolving these issues, and the argued prospect of voice assistants above, users, usage environments and the utility of voice assistants appear to be the core stake of the subject. This incites us to think about the necessity of consolidating User Experience within voice assistants in order to:

- Maintain the usefulness of voice assistants around how users *can* use them;
- Elevate their performance concerning the sensitivity to the *environment(s)* where they are used;
- Keep users at the *nucleus* of the design and development;
- Address voice assistant services as what they *mean* to their users in their usage environments *smoothly*; and
- Raise the quality of the experience of voice assistants by enhancing their *role*.

In order to properly apply a User Experience approach to study the subject, it is important to look into the use of voice assistants to first understand their users. Research questions raised for this study can be summarised as:

- How do voice assistant users use their voice assistants?
- In what ways is users’ experience with voice assistants perceived and enrichable?
- What are the desired but missing functions of voice assistants regarding their usage environments?
- How are the possible ways to improve voice assistants performance and/or services while addressing users’ aspect, voice assistant development, and usage environments altogether holistically?

⁴ ‘*Dot-com* bubble’ refers to the excessive speculation which occurred in the late 20th century. The strategy is linked to putting the ‘.com’ suffix to companies’ name unconditionally so as to raise capital. This phenomenon is used as a satirical metaphor for the voice assistant bloom amongst hardware devices (see SECTION 2.1 A Phenomenon to Address above).

3 STATE OF THE ART

In this chapter, I aim to first determine what voice assistants and usage environments are, and how my study includes them in the study scope and study focus. Then, with the aid of human-computer interaction and **voice user interface** design perspectives, I discern the interaction between users and voice assistants, analyse the usage environments of voice assistants, identify limitations in the current interaction, and distinguish potentials which could resolve these limitations and/or bring forward the interaction. After that, I attempt to show how the experience with voice assistants is interpreted on users' mind, comprehend the elements which contribute to this experience, and identify how this experience could be enriched.

3.1 Voice Assistants

There is a wide range of terms which are attributed to describing voice assistants. Some can overlap with each other in one context whereas they can invert the envelope in another. What voice assistants can achieve also can affect how they are being addressed. Since there are different types of voice assistants—which offer dissimilar fans of services—existing in the market, before going deep in the subject, it is vital to determine what voice assistants and their services are, and how they are considered in this study.

3.1.1 Definition of Voice Assistants

In general, 'voice assistants' could be interpreted as **agents** which bear the objective of offering help over voice (Novak *et al.*, 2016; Pearl, 2016, pp. 1–12). To further specify the definition of 'voice assistants', Radziwill and Benton (2017) have established an extensive classification about Dialogue Systems.

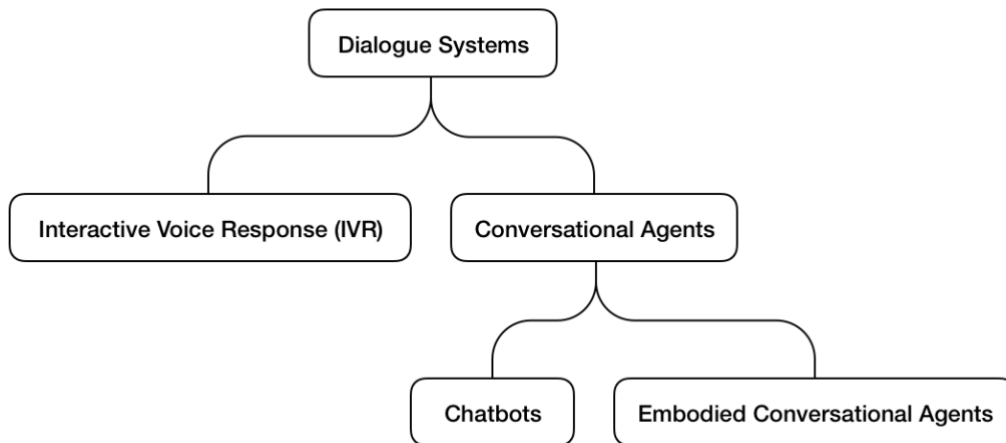


Figure 1 Radziwill and Benton's (2017) classification of software-based Dialogue Systems

Radziwill and Benton (2017) state that conversational agents are software systems that mimic interactions with real people. All conversational agents, together with Interactive Voice Response (IVR)⁵, originate from Dialogue Systems. 'Conversational agents' encompasses chatbots and embodied conversational agents. The former ones are described as programmed robots which communicate in a mimetic manner via text; they are generally found online and in some Internet of Things (IoT) devices. They could sometimes be embodied, but often less of an integral impersonated role when compared to embodied conversational agents. This latter category further

⁵ Interactive Voice Response (IVR) is a technology which allows interaction between human and a computer system through Dual-Tone Multi-Frequency signals (Harris, 2005, p. 535). It is widely used in automated telephone hotlines to let customers choose their desired services by tapping the corresponding inputs on their keypad (Radziwill and Benton, 2017; Pearl, 2016, pp. 13–16).

includes a range of subclasses, all depending on the complexity of the programmed role and capabilities, varying from personal assistant, intelligent virtual agent, to companion.

In this thesis, borrowing the definition by Radziwill and Benton (2017), ‘voice assistant(s)’ denotes the ‘embodied conversational agents’ whose goals are, additionally, to offer assistance to human(s). In summary, ‘voice assistants’ are the agents which:

- i) Have **voice interface** as the principal way of access and interaction with their users;
- ii) Hold the capability to partake in a dialogue with human(s); and
- iii) Carry the role of offering assistance to human(s).

The term ‘voice assistant(s)’ thus, in theory, covers the subclass range of ‘embodied conversational agents’—personal assistants, intelligent virtual agent, and companion—as long as these three above-mentioned categories are fulfilled.

3.1.2 Services of Voice Assistants

There are four major voice assistants existing in the market: Amazon Alexa, Google Assistant, Siri by Apple, and Cortana by Microsoft. Services provided by these voice assistants depend on the range of developed **applications** for each of them. These applications can be developed by the voice assistant company and by third-party developers (through the release of ‘developer kits’, [Isbitski, 2015; Johnson, ‘Google Assistant SDK Launches for Developers and Device Makers’, 2017; Bouchard, 2016; Pall, 2017]); often it is that a voice assistant company enters in co-operation with a third-party company to connect their services to the assistant platform and to make them available to voice assistant users (e.g. *Todoist* cooperated with Amazon to natively integrate their to-do list service with Alexa in October 2016 [The Todoist Team, 2016] so that Alexa users could synchronise to-do items across the voice assistant, compute, and smartphone platforms). Depending on the openness of the policy of the voice assistant company, individual developers could also build applications to bring services to the voice assistant platform. Amongst them all, Amazon Alexa holds the greatest number of available applications (**Alexa Skills**), thanks to its early and open development⁶. Debuted in the United States of America in November 2014, the number of Alexa Skills has topped 30 000 in the Amazon Skills Store US as of late March 2018 (Kinsella, 2018). These applications cover a fan of assistance, namely but not exclusively, personal organisation, information, ordering, entertainment, home automation, and communication (see APPENDIX B Alexa Assistance Service). While it is possible that the same assistants can be accessed through different **voice assistant-enabled devices** (see APPENDIX A Alexa-Enabled Devices for device examples), the services of voice assistants remain universal and consistent as long as they are available, i.e. asking for information on an Echo speaker will lead to the same information offered by Alexa if the **request** is made on the smartphone Alexa app (Justin Nikolaus, personal interview, *Interaction 18* conference week, 2018).

While it is hard to estimate accurately the actual, practised market share between major, existing voice assistants, it is still viable to state that Amazon Alexa is leading the industry based on the accountable market penetration of their smart speakers Echo⁷. Nevertheless, it is worthwhile to mention that some other voice assistants can outperform Alexa in other aspects. For example, both Google Assistant and Siri have a lower **Natural Language Understanding (NLU)** error rate

⁶ Amazon has first released their Alexa Skills Kit to enable individual, independent developers to build customised applications in June 2015 (Isbitski, 2015). This is years ahead of other major voice assistant companies.

⁷ Apple and Google have introduced voice assistants in their smartphone operating systems (OS) as early as in October 2011 and July 2012 (Google Now, the former voice assistant preceding Google Assistant) respectively. Yet, there has not been publicly released figures on the number of users who have actually *utilised* these voice assistants, which are available amongst other various services, on their smartphone. Amazon Alexa was first launched inside smart speakers Echo, of which the service is only limited the voice assistant. According to Kinsella and Mutchler’s (2018) report, 47.3 million (as of January 2018) American adults live in a household with access to one or more smart speakers; with Echo speakers holding 71.9% of smart speaker market share. It can be concluded that Amazon holds a leading part in the voice assistant industry.

in processing users' voice commands/**utterances** (Coucke *et al.*, 2017) (refer to the following SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants). This could imply that Google Assistant and Siri are running on more advanced artificial intelligence (AI) algorithms.

For the reasons of the foremost number of applications, the significant *accountable* market penetration, and the leading open development, Alexa is often taken as the voice assistant example in this thesis.

Apart from the services connected to voice assistants through their applications, the voice assistants themselves and their specialisation also differ, depending on the voice assistant companies' strategy. Activate (2017) reports that, according to the core business share of the voice assistant companies' revenue, Google orients their Assistant more towards search service, Amazon inclines Alexa towards e-commerce and product discovery, whereas Apple and Microsoft lead their voice assistants towards the development *with* devices, software, and operating systems (OS).

3.1.3 Working and Interaction Mechanism of Voice Assistants

Regarding the four major voice assistants in the market, they are all powered by artificial intelligence (AI) and **Cloud services** (Headrick, 2018; Burgess, 2017). These voice assistants work on voice interfaces. Users can interact with them in **natural language** through turn-by-turn **voice command dialogues** (Shakeel *et al.*, 2017). In general, a voice assistant can be accessed through the hardware device on which the particular assistant is enabled, i.e. a **voice assistant-enabled device**. The assistant is usually not an incorporate part of the hardware device but exists in the remote server of the company which owns, develops, and hosts the voice assistant service (Moynihan, 2016); this is thus usually referred as the **Cloud service**. The voice assistant-enabled device, hence, acts as an access point to the remote voice assistant and bridges the information transmission between the user and the assistant service.

Depending on the voice assistant-enabled device, a microphone which constantly captures sounds in the environment around the device is usually installed so as to locally detect the **wake word** (e.g. 'Alexa' by default for Amazon Alexa, 'Hey Siri' for Siri, and 'Okay Google' and 'Hey Google' for Google Assistant, and 'Hey Cortana' for Cortana). When the wake word is captured, the audio recording is triggered. The audio, dubbed as the **request** or **utterance**, is streamed to the Cloud server for processing (refer to APPENDIX C Alexa's Working Mechanism for a graphical representation) (Moynihan, 2016; Doring and Srivastava, 2017). The device detects the end of the user's speech and closes this recording-transmission window. When the always-on microphone feature is not available on the device, this part of the procedures can often be alternatively done by pressing an action button on the device, or by pressing and holding a microphone button on any enabled remote control (see APPENDIX A Alexa-Enabled Devices for examples on start-on mechanism) (Amazon, 2014). To advise the user that the device is streaming audio, visual indications are usually provided (e.g. a blue ring or line for Amazon Alexa, four floating dots for Google Assistants, fluctuating sound waves for Siri, and a varying blue circle for Cortana).

The streamed audio recording is then treated by a series of processing on the Cloud server. The concrete steps vary according to the assistant services, but usually comprise **Automatic Speech Recognition (ASR)** which transform the streamed audio recording into words, Natural Language Understanding (NLU) which detects keywords for identifying the user's **Intent** and information (**Slot**) expressed in the request, and Machine Learning which nourishes the algorithms of such processing mechanism (Pearl, 2016, pp. 129–160; Doring and Srivastava, 2017). These calculations often compare and take account of the enabled applications which are linked to the user's account for the voice assistant service and determine whether any third-party services are necessary for realising the identified Intent (Pearl, 2016). If it is necessary, communication and information exchange would be set up with the server of the third-party service provider to gather information and/or realise the identified Intent (Doring and Srivastava, 2017). When the calculations are done, the Cloud service often generates a textual response, transforms it into an audio response by **Speech Synthesis**, and streams it back to the voice assistant-enabled device to

play an audio back to the user (Doring and Srivastava, 2017). This can be an affirmative response which closes the dialogue, or a follow-up question which continues to another round of dialogue in order to, for example, solicit extra information from the user before completing an action in response to the identified Intent (Reeves, 2017).

It is worthwhile to state that, despite the above-mentioned prevalent pattern, there are other existing possible working mechanisms for voice assistants. Notably, *Snips* offer a non-Cloud-based AI platform which provides pre-trained algorithms for building voice assistants that can work individually on voice assistant-enabled devices alone, even without Internet connection (Coucke *et al.*, 2017). Contrary to the major, universal, all-in-one voice assistants, though, this creates voice assistants which are more specific, service-to-purpose, and decentralised.

3.2 Usage Environments of Voice Assistants

3.2.1 Definition of Usage Environment

‘Usage environment’ is a term which can comprise different context. As one of the focuses of this study is to instantiate voice assistants in the usage environment and consider holistically its effect can act on and bring to users, it would be wise to take an all-round consideration of the usage environments of voice assistants. To achieve this, the asset of holistic experience design could be a reference:

‘Holistic design takes into account the person, the device, the moment, the ethnographic environment, the physical space as well as human behaviour and psychology, i.e. thinking, attitudes, emotions, motivations, abilities, triggers, etc., and aims to deliver an optimal experience. At times the entire experience (with a product or brand) is not limited to digital devices but is a mix of digital, real-world brick-and-mortar, and human-to-human interactions.’ (Lyonnais, 2017)

In this thesis, ‘usage environment’ takes the broad, general meaning. It refers to the space in which voice assistants are being accessed. It carries two layers of meaning in this thesis. It comprises:

- i) Tangible data and metadata of the environment which the voice assistant is being accessed, e.g. time, location, and users’ profile; and
- ii) More intangible information of the background context which is associated to the moment when the voice assistant is being accessed, e.g. context of the surrounding space, users’ emotion, and users’ preferences.

3.2.2 Existing Usage Environments

As explained in the SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above, most voice assistants are accessed through their designed voice assistant-enabled devices. The designated usage environments of these mechanical devices and their availability in spaces *predetermine* the usage environment of voice assistants. Whilst voice assistants were initially introduced into a limited number of usage environments, they have become soon more accessible and in sundry usage environments as the hardware support grows rapidly (VoiceLabs, 2017; Activate, 2017; Lacoma, 2017; Perez, ‘Alexa Is Coming to Wearable Devices, Including Headphones, Smartwatches and Fitness Trackers’, 2018; Hetzner, 2017).

For Amazon Alexa, it was first launched in November 2014 within the chassis of ‘smart speaker’ Echo, which was designed to be used in interior environments and requires cabled power supply (Amazon, 2015). Over the years, the company enriched the official product line of Alexa-enabled devices (see APPENDIX A Alexa-Enabled Devices), making the assistant present in various forms of hardware and diversifying the possible usage environments which it can be accessed. From the official product line from Amazon, these environments now further cover more personal indoor environment (via tabletop accessories, and cameras), indoor common place environment

(via home entertainment equipment), portable/outdoor environment (via portable speakers), and smartphone-tablet environment (via smartphone operating system [OS] applications) (Gordon and Wroclawski, 2018). By entering in co-operation with third-party manufacturers, the assistant has become further accessible, typically on indoor multi-room environment (via smart home controls, household appliances, furniture and fixtures), driving environment (via car integration), and ubiquitous environment (via wearables) (Gordon and Wroclawski, 2018; McGrath and Hastings, 2018).

Google Assistant (formerly Google Now), Siri, and Cortana have a different origin compared to Amazon Alexa. They were first introduced inside their corresponding native smartphone operating systems (OS), namely Android OS, iOS, and Windows Phone OS (Needleman, 2012; Velazco, 2011; Kapitall, 2014); they are thus originated in the smartphone-tablet environment. Later, with the expansion of services, both Siri and Cortana were brought to work environment (through their computer OS) (Statt, 2016; Osborne, 2018), with Google announcing a similar intention on development (Maring, 2018); smart speakers—similar to Echo—with only their voice assistants installed were also released, thus expanding to interior environments (Bushnell and B2B Tech Writer, 2018). All these three voice assistants were further brought to the portable, outdoor and ubiquitous environment with the launching of smartwatches, bracelets, and earbud devices (Pearl, 2016, pp. 228–231). It can be concluded that all the major voice assistants are becoming portable and closer to being ubiquitous, i.e. whenever users are and whenever accessible. This **extends** greatly possible **usage environments** of voice assistants.

Besides the mobility, there are several noteworthy and emerging usage environments in which voice assistants are particularly useful. *Driving environment* is one of them thanks to the arguable exclusive benefit of hands-free control⁸. Even before the official support is launched, it has been reported that Alexa users had tried bringing their Echo Dot smart speaker out and installing it to their car for infotainment purposes (Wollerton, 2016; Gaines, 2017). Several third-party manufacturers seized this opportunity and entered in co-operation with Amazon to develop driving environment-oriented hardware support, for instance, the GPS navigation gadget Garmin Speak for route planning and navigation over voice interaction, and the incorporation of Alexa as the inbuilt car assistant in BMW, Ford, and Volkswagen vehicles (Hetzner, 2017). Another usage environment which is relatively new and advantageous is the *transport-related environment*. Johnson (‘Alexa and Google Assistant Should Tell You When the next Bus Is Coming’, 2017) suggests that voice assistants should provide habit-adapted transportation information, work with public transport service providers’ data⁹ to increase the accessibility of public transport, enhance the environmental benefits, as well as make the transit experience more pleasant. Shenoy *et al.* (2017) propose the use of voice assistants in transport to assist visually impaired people in *getting around* in their daily life. They carried out a scientific study, created a prototype, and tested this use on mobility assistance, security, social, household and employment facets. The results show that the combination of voice assistant and visual recognition system is highly feasible and positive in *transport* and *outdoor environments*. Additionally backed by some findings of this study (see SECTION 5.1.2 Insights below), this *outdoor transport-related environment* is later chosen as the

⁸ While some suggest that voice assistants in driving environment can offer hands-free infotainment to drivers and keep their hands on the steering wheel as they are supposed to (Liszewski, 2017; Ong, 2017; Gaines, 2017; Wollerton, 2016; Gitlin, ‘Dragon Drive Is the Best Car Voice Activation System We’ve Spoken to’, 2016), others argue that the interaction with voice assistants still create extra cognitive load to drivers and can cause extra disturbance (Gonzalez, 2017; Gitlin, ‘Infotainment Bad, Blind Spot and Lane-Departure Warnings Good, Studies Say’, 2017).

⁹ Johnson (‘Alexa and Google Assistant Should Tell You When the next Bus Is Coming’, 2017) quoted that transportation agencies in Dallas and Los Angeles started working with ride-sharing service providers such as Lyft and Uber on data exchange in order to resolve the ‘last-mile problem’ in inhabitant movement patterns. The motive behind this cooperation was described as to encourage inhabitants in a metropolitan area to favour public and/or shared transport over private cars when they have to travel but live in a short yet ‘inconvenient’ distance between their household and the nearest available spot of public transport.

usage environment focus (as ‘**commute environment**’) because of its novelty and potential advantages.

3.2.3 Usage Environments Sensitivity

Whilst at the current stage the **sensitivity** of voice assistants to **usage environments** remains fairly limited to what users actively put in as the environment and information context (Pearl, 2016, pp. 49–53, 198, 200–202; Berg, 2014, pp. 66–67) (see SECTION 3.3 Interaction with Voice Assistants below), the assistance services can sometimes turn into rounds of cumbersome dialogues before functions are carried out (Johnson, ‘Alexa and Google Assistant Should Tell You When the next Bus Is Coming’, 2017). In this way, accuracy in capturing user input aside, voice assistants can be rather passive. *Augmenting voice assistants’ sensitivity to usage environment* could help solve the problem. It could raise their ‘sensing’ capability—especially when information is *unspoken* (McMillan, 2017; Chachada and Kuo, 2014)—and initiate environment modelling and context treatment *proactively* (see the following SECTION 3.3.1 Interaction as in Human-Computer Interaction below). Voice assistants’ qualities could also be raised by proposing suitable assistance to the users *anticipatorily* (Moorjani, 2018) (see SECTION 3.4.4 Anticipation below). Referring to Lyonnais’s (2017) quote above, this could enable voice assistants to offer services in a more holistic manner.

3.3 Interaction with Voice Assistants

With ‘voice assistants’ defined, their functioning mechanism discerned, and their usage environments presented, understanding how users interact with voice assistants is then essential. As voice assistants are driven by artificial intelligence (AI) (Radziwill and Benton, 2017; Berg, 2014, pp. 20–70) and accessed by voice, the interaction between users and voice assistants laps over the domains of Human-Computer Interaction (HCI) and Voice User Interface (VUI) design.

3.3.1 Interaction as in Human-Computer Interaction

In human-computer interaction, an **agent** is as one which acts with a goal-oriented purpose or purposes, and **agency** as the ability to act *effectively* with respect to goals (Novak *et al.*, 2016); over time, an agent would develop its own agenda and be capable of effecting what it senses in the future (Franklin and Graesser, 1997, p. 25). The capability of an agent to act depends on the understanding of the environment in which it practises its agency and the assimilation of information which it receives into its understanding of the environment. As an agent which offers assistance to users, a voice assistant’s agent model can be briefly represented with the framework suggested by Novak *et al.* (2016) as:

- Sensing – The voice assistant receives information (explicitly from voice commands, and implicitly environment metadata and personal data pulled from the associated user account and third parties’ services) about the environment;
- Environment modelling – The voice assistant incorporates the information into its understanding of the environment;
- Goal formation – The voice assistant calculates and generates intentional plans of its assistance to realise users’ intent(s); and
- Action – The voice assistant acts on the environment.

This framework indeed forms a loop in which each piece of information the voice assistant assimilated and acted upon will be internalised and will contribute to the voice assistant’s understanding of environment for subsequent comprehension. When users speak to the voice assistant in dialogues, it receives the information, processes it through its working mechanism, and incorporates that piece of information in its understanding of environment on which its process will run subsequently (refer to SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above). Even though the voice assistant may not always receive the introduced information accurately (e.g. missing words spoken by users or wrongly identifying homophones)

and/or correctly respond to the call of assistance (i.e. giving irrelevant responses to processed requests received), it still nourishes the environment model which the voice assistant conceptualises. Regarding information related to usage environments, it is currently not fully processed in the step of environment modelling (refer to SECTION 3.2.1 Definition of Usage Environment above)—even though the voice assistant manages to transcribe the heard voice command into relevant words, the information which these words carry is not fully taken into the understanding process (e.g. Alexa does not interpret ‘here’ or ‘nearby’ in users’ **Utterances** that generally refer users’ real-time location; Siri does not process users’ preferences on the means of transport even when that information is explicitly spoken in their request¹⁰), and is thus not used in modelling the environment or the understanding of environment. Succeeding steps of goal formation and action are, hence, carried on without this information. Therefore, assistance offered by the voice assistant is sometimes independent of information about the usage environment. To maintain correct environment modelling, raising the *sensitivity* of voice assistants to their usage environments could be a measure (as already explained in SECTION 3.2.3 Usage Environments Sensitivity above).

3.3.2 Interaction as in Voice User Interface Design

In voice user interface design, interaction between users and voice assistants are often pictured and/or related to interpersonal **conversation** interaction. To attain good interaction, there are a lot of reference models available; the core interaction aspects, though, can be summarised by Usability.gov’s (2014) guidelines on six aspects; adapted to voice interaction, these aspects are:

- Clear instructions on interaction – Users are given clear and enough knowledge on how to use their voice to interact with voice assistants;
- Clues about behaviour before actions – Users are advised of what functions voice assistants can do before they perform actions;
- Error anticipation and mitigation – Constrained by helpful error messages which contain solutions and context, the interaction force users to adjust the behaviour in order to move forward with their intended action;
- System feedback and response time – The interaction allows voice assistants to acknowledge users’ action and carries the correct latency in terms of responsiveness;
- Strategic arrangement of interaction elements – Interaction elements are congealed judiciously in the voice user interface to help users understand how they are supposed to function; and
- Simple learning – The interaction is simple and familiar to learn with a delicate information load for users.

While some of the above-stated aspects are more limited by the technical working mechanism of voice assistants (e.g. Natural Language Understanding and Intent mapping in Utterances), manoeuvrable aspects within the voice user interface can be found on the side of interface design. Regarding ‘clear instructions’, ‘strategic arrangement of elements’, and ‘simple learning’, voice assistants have been advertised as conversational agents to whom users can speak naturally as if they talked to real people (Earley, 2016; Chen and Wang, 2017). Despite the surprising ability of Natural Language Understanding (NLU), as soon as one has interacted with any voice assistant for a while, the reality is clear that there is a gulf between what voice assistants’ capability now allows themselves in maintaining a **dialogue** and a real interpersonal **conversation**. This *try-and-learn* approach has pushed users to find out the interaction with voice assistants through exploration and created a heavier load on ‘behaviour clues’, and more particularly on

¹⁰ For example, asking Siri ‘How could I get to work?’ and ‘I don’t want to drive today. How can I get to work?’ results in offering directions by car (default means of transport) and gives the same route suggestion. Asking Google Assistant the same second question even results in an ordinary web search, implying that the voice assistant cannot process fully and correctly the request.

‘system feedback’ and ‘error anticipation and mitigation’. While the bearable cognitive load by users on voice interaction is more limited when compared to graphical interaction (Gonzalez, 2017), the amount of transmitted information is greater. Pearl (2016) explains that voice—as the major, if not the only, medium for interaction and communication at the same time with voice assistants—does not simply carry information context, satisfies enquiries, or realises tasks; through dialogues, voice also conveys emotions and personalities. This further complicates the design for voice user interface as users—human beings in general—are extraordinarily sensitive to characters and identities in voices (Mary Parks, *Interaction 18* Education Summit, 4 February 2018). Parks (*Interaction 18* Education Summit, 4 February 2018) points out that we always tend to create an image of the voice to which we listen—we try to match the voice to our acquaintances within seconds; even if we do not know the voice, we start to construct a mental image of the voice; we attribute to the voice a gender and an age; a social class status and origin from the way of speaking; and eventually we deduce the personality and the character. For the reason that these calculations take place at the same time as users perceive information from the voice assistants which they interact with, in order to keep the interaction going steadily and fluidly, voice user interface design has to take these unique concepts into account.

3.3.3 *Limitations of Voice Interaction*

While considering the interaction with voice assistants, the current state of the working mechanism (see SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above) is one of the limiting factors. According to Reeves’s (2017) framework, where voice assistants fall short in dialogue performance, when compared to interpersonal conversations, can be summarised and generalised as follows:

- *Action anticipation* – Voice assistants cannot anticipate forthcoming, necessary actions which a dialogue could lead whilst interlocutors in an interpersonal conversation could predict and act upon them in expectation (e.g. cutting in one’s discourse in a debate);
- *Flexible speaker selection* – Voice assistants only respond when they are addressed with the wake word at the beginning of a voice command, whereas a speaker can be assigned in the middle or at the end of a spoken phrase in an interpersonal conversation (e.g. ‘*What’s the time now, Walter?*’);
- *Awareness of rhythm and context repair* – Voice assistants’ interaction mechanism forces users to participate in a fluid, straightforward, turn-by-turn dialogue, whilst an interpersonal conversation is much more forgiving in context adjustment and repair¹¹; and
- *Coherence in sequential contexts* – Voice assistants often miscalculate whether a subject matter or element in a dialogue is still ongoing or has yet ended, and thus can give irrelevant, abrupt responses¹².

While Reeves (2017) argues in his paper whether dialogues with voice assistants should be regarded correspondingly with interpersonal conversations, Chen and Wang (2017) spotlight the consequences of user frustration and ‘learnt helplessness’ which the disparity in users’ estimation and the reality of voice assistants’ capabilities can cause; these results could be detrimental to the conversational interaction with voice assistants and lead to the rejection of voice assistants.

¹¹ Pauses and rephrasing are often in human speeches. When in an interpersonal conversation, interlocutors have the capability to adjust their understanding of context and ignore intermittent pauses, at which voice assistants are, if not completely incapable, very weak at this stage.

¹² For example, as Reeves (2017) demonstrated, when a user enquired of Alexa weather forecast in Denver, USA, Alexa, Alexa first responded correctly, but failed to relevantly go on when the user asked a follow-up question about the forecast in the next week and reported weather in Nottingham, UK (according to the registered address in the device location setting).

Augmenting voice assistants' *sensitivity to usage environment context* is thus vital (as explained in SECTION 3.2 Usage Environments of Voice Assistants above).

Anders (2017) acknowledges that the progress in the achievement of Natural Language Understanding (NLU) is aspiring; basing on large swathes of data gathered from the conversational interaction with various users, taking account of previous mistakes, and making provisional guesses, machines can eventually conclude in patterns and adapt themselves to conform to us. Nonetheless, Anders (2017) points out that patterns of everyday life change at an even faster speed which outpaces what machines could learn and catch up in time. Besides, regarding the cultural background, personality, the environment, the state of mind, personal preferences, and so on, the way users stitch their thoughts, the way they express them out, and the way they speak can all differ. The combination of utterances is infinite and encompassing all of them would be simply beyond the possible framework of the current voice assistants' development and/or working mechanism. Users are, therefore, inevitably expected to adapt themselves to voice assistants' capabilities (Dulin, 2016), relatively adjust their voice commands and utterances in dialogues in the manner that voice assistants can understand and process, lest they be disappointed. This is contrary to how hard voice assistant companies try to package their assistants.

3.3.4 Potentials of Voice Interaction

In respect of voice assistants' current limitations, there are potentials that voice assistants could be tuned to resolve the existing problems and/or even bring the interaction forward and beyond the current state.

As the voice assistants' working mechanism is currently limited to turn-by-turn dialogues, and that users might have an overexpectation on how they can interact with the agent, special attention is needed in designing the interaction dialogue for assistance access (and failure of assistance access) in order to minimise the aspects where Alexa might fail users' anticipation (see the bullet points in SECTION 3.3.3 Limitations of Voice Interaction above).

Turning the voice user interface more transparent could be one way to help resolve the existing problems. Chen and Wang (2017) suggested that, when voice assistants fail to process users' utterances or realise users' intents, they can provide in-situation, knowledge-filling feedback so as to elicit from users an appropriate, realistic and usable mental model which allows them to apprehend what agents are capable of and how to access their assistance. For example, when the user is voicing out a command to the assistant in a manner that the latter could not process, instead of simply replying '*Sorry, I don't know that one*' or '*I am not sure how to help you with that*' (examples extracted from interaction with Amazon Alexa), the voice assistant could give instructive feedback to help the user learn how to reformulate their command for their desired assistance or services; when the user is enquiring about unserved information, it could give examples of what it could provide instead. Informative, explanatory, and knowledge-filling feedback could be a good leverage approach to, on the one hand, clear the doubt caused by obscure negative responses, and help users develop a more proper mental model of voice assistants on the other. Moreover, this could also raise the affordances of voice assistants and make their services more perceptible (Novak *et al.*, 2016), reduce the discrepancy between expectation and reality (Luger and Sellen, 2016), and prevent learnt helplessness over repetitive interaction failure (Maier and Seligman, 1976). In consequence, voice assistant systems could become more transparent for users while maintaining them at the helm of control. This would align with the Heuristics of Nielsen (1993), particularly on system visibility, user control, and helping users recognise, diagnose, and recover from errors. As a result, the potential threat of abandon of use could be reduced and the experience for users could be improved (refer to the following SECTION 3.4 Experience with Voice Assistants).

Further projecting the capabilities of voice assistants, the turn-by-turn dialogue interaction mechanism could also be overtaken. McMillan (2017) suggests that Automatic Speech Recognition (ASR) algorithms could be expanded to include implicit and ambient interaction with

voice assistants. ASR could be developed to analyse background sounds of an environment so as to augment voice assistants' capabilities in environment modelling and carrying out more adapted actions. Under their idea of 'continuous speech stream' (McMillan *et al.*, 2015), by constant, always-on audio sampling, voice assistants could identify non-speech audio such as emotion-expressing sounds (e.g. crying and laughing), environment sounds (e.g. footsteps, *transport* and *movements*), and activity sounds (e.g. eating). On that account, voice assistants could *anticipate* users' next actions. Chachada and Kuo (2014) suggest that, with the similar technology Environmental Sound Recognition, background sounds could be analysed by machines to enhance home security and to anticipate needed assistance in household environment (e.g. for elderly people who live alone). On a social aspect, Eagle and Pentland (2006) suggest that wearable gadgets, such as even just Bluetooth-enabled mobile phones without Internet connection, could already provide enough data to detect interpersonal relations in complex social systems. All these could raise the **usage environment sensitivity** of voice assistants, enrich the interaction with voice assistants, and enhance the user experience by *anticipating* unspoken context in the interaction (see SECTION 3.4.4 Anticipation below).

3.4 Experience with Voice Assistants

After explaining the interaction between users and voice assistants, it is time to look at the result of the interaction and comprehend how voice assistants contribute to the experience which users undergo through interacting with them.

User experience is a complex entity which can be affected by the context of experience. In general, it encompasses the usability, accessibility, and enjoyability offered through interaction with a product (Kujala *et al.*, 2011). To objectify user experience, Hassenzahl (2004) suggests a framework which can break experience down into four main categories of attributes which users construct through the interaction with a product; this can help render experience more tangible for evaluation. Basing on the framework of Hassenzahl (2004), regarding the use of voice assistants, user experience can be cloven as follows:

- Manipulation – Whether the voice assistant has relevant functionality (i.e. utility) in respect of users' expectation and whether it offers users the ways to access its functionality (i.e. usability);
- Stimulation – Whether the voice assistant has triggers for users to activate its features (to achieve their intent) and whether it can provoke them to explore and use these features;
- Identification – Whether the voice assistant can socially communicate its identity (a conversational agent which offers assistance) to users; and
- Evocation – Whether the voice assistant can create a link with users and elicit a position in their memory.

Focusing on the service of voice assistants, however, manipulation remains a predominant, prerequisite aspect. This is because all other three aspects depend on the existence of functionality (i.e. the assistance related to users' intent). If the functionality does not exist or only exists fragmentarily—in this study, for example, assistance related to usage environment changes and/or adaptation of the voice assistant's behaviour of offering the assistance—other factors simply cannot be wholly addressed.

Radziwill and Benton (2017) suggest in their extensive literature review another framework for quality attributes of conversational agents. This is conventionally inherited from the ISO 9241 concept of usability¹³:

'The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments.' (Abran *et al.*, 2003)

To further explicit Radziwill and Benton's (2017) concept in the use of voice assistants:

- Effectiveness indicates the accuracy and completeness with which users achieve their intents with the help of voice assistants—this is related to voice assistants' performance, whether voice assistants can robustly receive users' inputs, capture their manipulation, and appropriately respond to users without damaging their expectations;
- Efficiency denotes the quality of how resources attributed within voice assistants themselves to achieve users' intents—this is related to the functionality (e.g. whether the working mechanism of voice assistants is allowing the extraction of users' intents from their utterances aptly; whether the algorithms in Machine Learning is allowing the adaptation of voice assistants to user profiles);
- Satisfaction points to the facet of voice assistants in terms of fulfilling users in the experience engagement—this comprises the affective, behavioural, and accessibility traits which range from greeting users in a proper manner and contributing to a fun involvement with users, respecting users' dignity, safety and privacy, to meeting neurodiverse needs amongst users such as extra response time and the way of speaking.

Combining the frameworks of Hassenzahl (2004), and Radziwill and Benton (2017), experience with conversational agents can be sectioned and outlined as these three following criteria:

- Efficiency is subjected to the usability of voice assistants, which could be typified by the ease of access to voice assistants' applications and connected services;
- Effectiveness can be referred to the accuracy in processing users' utterances;
- Satisfaction can be represented by affectional influence.

3.4.1 Accessibility of Applications and Connected Services

To realise users' intents, applications of voice assistants corresponding to users' call of assistance need to be available for voice assistants to invoke. Basing on the current Working and Interaction Mechanism of Voice Assistants chain as explained in SECTION 3.1.3 above, by logical interpretation of the steps how users can interact with voice assistants, this signifies that an accessible application in question has to:

- i) First, exist in the voice assistants' application store;
- ii) Then, be findable by users and/or by voice assistants;
- iii) Next, be activated; and
- iv) Finally and usually, have a name remembered and called by users, and/or keywords recognisable by voice assistants.

Depending on the voice assistant companies' policy, voice assistants' application stores can be region-discriminative. For example, Amazon has distinct Alexa Skills Stores in different countries. Although Echo speakers (the major Alexa-enabled device product line) have been available to over 80 countries since 8 December 2017 (Blankenburg, 2017), Alexa Voice Service and Alexa Skills Store are still limited from language to language and from region to region (Amazon, 'Alexa Voice Service [AVS] International', 2017); as of 5 May 2018, supported regions

¹³ In Radziwill and Benton's (2017) literature review, 7,340 articles have been assessed and a summarised table detailing quality attribute and reference for each criterion is available. The relevant concepts are adapted to voice assistant context and summed up in the following bullet points.

include the USA, the UK, Ireland, Canada, Germany, Austria, India, Japan, Australia, and New Zealand. An Alexa Skill uploaded to a regional Store does not necessarily mean that it can be accessed in another. Skills which require region-specific functions such as pulling local restaurant locations can also be likewise limited. This can create a discrepancy in which identical hardware devices could be fully supported by the software of voice assistants in one country, but fragmentarily in another. The situation is further amplified by the fact that the hardware and software are not developed and/or launched in a synchronised way (see CHAPTER 2 RESEARCH PROBLEM and SECTIONS 3.1.2 Services of Voice Assistants & 3.2 Usage Environments of Voice Assistants above).

If the concerned voice assistant applications need to forward users' request to other third-party services and pull their data in the process of realising users' intents, the availability of these third-party services and the intercomputational communication and information exchange between voice assistants and them could also supplement the factor list and affect the voice assistants' performance (as explained in SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above and illustrated in APPENDIX C Alexa's Working Mechanism).

In this case, voice assistant application developers are mostly responsible for this 'accessibility of applications and connected services' criterion of user experience with voice assistants. We could imagine that the situation would gradually improve when the technical development matures and when voice assistants' services expand and propagate.

3.4.2 Accuracy in Processing Utterances

Understanding users' way of expression is the key factor of the interaction between voice assistants and users. How effective this process is brought through weighs hugely in the user experience with voice assistants. To have a functional interaction, voice assistants are supposed to fully and precisely capture users' utterances and take them into their environment model (as explained 'sensing and understanding' in SECTION 3.3.1 Interaction as in Human-Computer Interaction above). Accuracy in this step is crucial as it predetermines whether the interaction bridge can be set up between voice assistants and users before they can react upon input from users. All succeeding actions and reactions between the two sides also rely on this precondition. Any misinterpretation in users' request could lead to huge deviation for the following. Failure or inexactitude in this step will entail irrelevant succeeding actions performed by the voice assistants and user frustration (Chen and Wang, 2017).

At the current stage, the process works on Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), and Machine Learning (ML) (as explained in Section 3.1.3 Working and Interaction Mechanism of Voice Assistants above). How well the voice assistant companies have at their disposal these technologies, of course, play a role here. Voice assistant application developers, nonetheless, play the other. This is because during the process, extraction and recognition of users' intent in Utterances are encoded by them. The successful rate is delimited by the range and the combination of mappable Utterances, Slots and keywords (refer to SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above and APPENDIX C Alexa's Working Mechanism); so is the variety of Intent outcome, too. In a figurative sense, they create the links how the technologies on the Cloud are used in specific voice assistant applications to suit users' intents and needs.

Nowadays, the only source of decoding Intents is users' Utterances. We could picture, though, in the future that other improved or innovated auxiliary technologies could exist to help voice assistant process accurately what users desire and/or need. The concept of augmented Automatic Speech Recognition (ASR) algorithms (McMillan, 2017) (explained in SECTION 3.3.4 Potentials of Voice Interaction above) could be one environment context-enriching means for the voice assistants' sensing and understanding model. Better connected and shared data sets across services could be another (see FOOTNOTE 9 above for Johnson's ['Alexa and Google Assistant Should Tell You When the next Bus Is Coming', 2017] example). Understanding could

take place without users' voice commands (see SECTION 3.3.4 Potentials of Voice Interaction above for the ideas of 'continuous speech stream' [McMillan *et al.*, 2015], Environmental Sound Recognition [Chachada and Kuo, 2014], and complex social system detection of [Eagle and Pentland, 2006]). Users' intents could be *anticipated* (see SECTION 3.4.4 Anticipation below). Accuracy could hence be raised.

3.4.3 *Affectional Influence*

As voice can transmit emotions and personalities (see SECTION 3.3.2 Interaction as in Voice User Interface Design above), when users interact with voice assistants, the emotions and personalities expressed by the assistants' voice can affect the experience on users, too. In general, users are sensitive to personality traits implemented in technology (Lee *et al.*, 2005; Nass *et al.*, 1995); when users interact with voice assistants, the perceived personalities in their head determine the sociableness which users would take to voice assistants. Mary Parks (*Interaction 18* Education Summit, 4 February 2018) explains this effect of sociableness with voice assistants by relating it to similarity attraction in interpersonal interaction—like-minded characters would tend to stay for and appreciate interacting with each other, whereas opposite ones would tend to repel each and deprecate the interaction (Lydon *et al.*, 1988). Parks (*Interaction 18* Education Summit, 4 February 2018) also adds that, since even voice assistants themselves carry profiles through their voices, how specifically assistant profiles could address and match with user profiles would determine the user experience.

Explicit characters in voice assistants do not only help create (or prevent) the bond between voice assistants and their users, they can not only save some marks when the assistants fail users' expectation (e.g. rounding a dialogue with a sassy joke when being unable to deliver information which users ask for) (Kosner, 2013). Their characters also help fortify users' social, mental model of the assistants (Guzman, 2017), making the interaction immersive and the service memorable (Tapp, 2015). In this way, the affectional influence does not only contributes to user experience, but can also play a critical part in users' judgement (Norman, 2004)—for example, whether they would continue using a particular voice assistant or not.

Besides voice assistants' mimetic, human character which users have to find out over use, emotion components could also attract users. For instance, typical 'Easter eggs' in technology products use surprise to incite users to keep exploring the system for rewards and encourage them to stay within the system (Gkogka, 2017). In voice assistants, emotion plays an important role in building the bond with users and keeping users to invest in the system (Austin, 2017). It completes the Behaviour Hook (Eyal, 2014) to form a constant loop to stimulate users to interact with voice assistants. Surprise is a one of the key rewards which fortifies the Hook (Austin, 2017).

3.4.4 *Anticipation*

As stated earlier, effectiveness, efficiency and satisfaction compose the user experience with voice assistants. Besides the described criteria, this experience could further be elevated by the anticipation (i.e. anticipating users' needs and assistance offer). According to Moorjani (2018), anticipatory design can create delightful experience by understanding users' needs and eliminating needless choices; the psychology of anticipatory design could be applied to voice assistants as:

- i) Principle of least effort – Assistance could be delivered to users with lesser effort (e.g. assistance is made easily accessible);
- ii) Feeling of being understood – Assistance (and voice assistants' behaviour) could be tailor-made to users' liking (and disliking), making them feel familiar and understood;
- iii) Desire to feel important – Voice assistants could address attention to users personally, making them feel special through interaction; and
- iv) Surprise – Voice assistants could offer the unexpected and the interaction could turn out unpredicted.

The psychology of anticipatory design resonates with user experience criteria. It responds to the research objective of keeping users at the nucleus of the service. It simplifies the interaction whilst it further amplifies the experience quality (effectiveness, efficiency, and satisfaction—as explained at the beginning of this SECTION 3.4 Experience with Voice Assistants above). Anticipation goes along with the purpose of artificial intelligence (AI), aligns with the development of voice assistants (on Machine Learning) (Whitenton, 2017) and enhance the *role of voice assistants*. **Anticipatory assistance** could, hence, be considered as a factor to enrich users' experience with voice assistants. Further backed by findings of the usage analysis (refer to Section 5.1.2 Insights below), this is later chosen to be the exploration focus of voice assistant design in this study.

* * *

Through literature review, the current stage of voice assistants and their issue with the usage environment have been identified. Factors of the interaction between users and voice assistants have been recognised, alongside their limitations and potentials. How the experience, which results from the interaction, is formed and perceived by voice assistant users are found out. These varieties of knowledge have formed the foundation of the application of User Experience Design perspective in the following parts of my study.

4 METHODOLOGY

Supported by the knowledge obtained and established through literature review, in order to address the research questions, the applied study was divided into two distinct sequential phases. Within the available resources, the adopted methods in this applied study remained qualitative.

The first phase of applied study worked towards understanding current voice assistant users and looking into their experience with current voice assistants. Its objective was to understand how current voice assistant users use their voice assistants, discover the appreciable and improvable parts of voice assistants, and gain insights into which directions voice assistants and the experience are worth making better. To understand current users, I chose to establish a questionnaire to approach the user group, followed by user interviews to study more in depth relevant profiles. In this understanding phase, I focused on Amazon Alexa users because the voice assistant is reported to be the one which holds the most accountable market penetration (Kinsella and Mutchler, 2018), owns the most number of supported applications (Kinsella, 2018), and allows open, adaptable services (as explained in SECTION 3.1.2 Services of Voice Assistants and FOOTNOTES 6 & 7 above). Whilst trying to apprehend current usage behaviour, preferences, and user needs with Amazon Alexa, I tried to relate the scope to the recognised **extended usage environments** (as explained under SECTION 3.2 Usage Environments of Voice Assistants above)—on smartphones and/or tablets, on private and/or public transport, in outdoor and unsheltered space, and potentially even a mix of them—to find out associated possible improvements.

The second phase of applied study followed the findings of the first phase. Its goal was to depart from the gained insights, recognise the feature and/or sensitivity improvements needed for the most improvable extended usage environment (i.e. the **commute environment**; see SECTION 5.1.2 Insights below), and shape the interaction and maximise the experience around the voice assistant. **Anticipation for assistance** and/or services for users basing on an *augmented sensitivity* to usage environments was selected as the focus owing to the benefits which literature review and findings from the first phase of applied study pointed out (refer to SECTIONS 3.2 Usage Environments of Voice Assistants, 3.3 Interaction with Voice Assistants, 3.4 Experience with Voice Assistants, 4.3 Participatory Workshop / Co-Design Workshop & 5.1.2 Insights). The voice assistant capable of achieving so is dubbed hereinafter as ‘**augmented voice assistant**’.

Attention was paid to the assessment of user experience with voice assistants by drawing the focus on to the assistance and/or services of voice assistants, *not* the technical aspect (e.g. how well voice assistants function on hardware devices). When designing the interaction and experience with the augmented voice assistant, emphasis was also put on the assistance and/or services so as to avoid reducing potentials by fitting them into current interaction and technological limitations.

4.1 Questionnaire

4.1.1 Contents of the Questionnaire

Aiming at understanding current Alexa usage behaviour, user preferences, and potential needs regarding the recognised extended usage environments, the questionnaire was distributed on active online forums and Facebook groups which gather large numbers of Alexa users¹⁴, as well as on

¹⁴ The number of members these forums and Facebook groups ranges from roughly 3 000 to 45 000. The last activities observed on these forums and Facebook groups were within two weeks’ time. These conditions should allow a significant number of collected response.

professional communication channels¹⁵. The open window for collecting responses was set as four weeks (28 days). This questionnaire contains four principal sections:

- i) Identifying general information on how users use Alexa—through which Alexa-enabled devices, at what frequency, for what purposes, at what time, where, and with how many Skills activated;
- ii) Gathering data on how many users have adopted the use of Alexa in recognised extended usage environments—on smartphones and/or tablets, in private and/or public transport, and in outdoor and unsheltered space—and evaluating their corresponding experience;
- iii) Collecting qualitative responses on users' preferences and expectations in the services of Alexa; and
- iv) Understanding respondents' profile and obtaining contact information of interested, potential user interview candidates.

Caution was specially taken in designing this questionnaire to make sure that the focus is put on *services* and *functions* of Alexa, *not* the Alexa-enabled hardware device or other technical matters.

4.1.2 *Model of the Questionnaire*

To precisely estimate user experience in recognised extended usage environments in the section ii) of the questionnaire, three questionnaire models which are widely used in the User Experience Design industry were considered, namely, System Usability Scale (SUS), AttrakDiff2, and User Experience Questionnaire.

The SUS model was presented by Brooke in 1986 and allows system usability evaluation through 10 questions. Each question contains an affirmative statement and asks participants to rate their agreement on a Likert-type scale of 5. The answers are then transformed and added up into a global score, which is compared to a grading scale to deduce the rating of system usability (Sauro, 2011; Bangor *et al.*, 2009). Yet, this model evaluates only users' impression on system usability and does not encompass user experience completely.

Hassenzahl *et al.* (2003) provide the AttrakDiff2 model which measures a product or an interactive system on four aspects—pragmatic manipulation, hedonic stimulation, hedonic identification and attractiveness¹⁶. It contains 28 questions in total (equally 7 questions for each aspect). Each question contains a pair of predefined antonyms which describe the assessed product or system, and allows participants to indicate their affinity on an anchor scale of 7. Answers are then converted into statistical values and plotted on charts to show the general representation of qualities of the product or system from users' representation of their experience (Ísleifsdóttir and Lárusdóttir, 2008; Lallemand and Gronier, 2015, pp. 367–382). Hedonic qualities take a big part in this evaluation model and may not be suitable for evaluation of Alexa service as the agent's assistance functionality is also fairly influential on users' experience.

Laugwitz *et al.* (2008) provide the User Experience Questionnaire (UEQ) model where evaluation of experience is broken down into six aspects—attractiveness, perspicuity, efficiency,

¹⁵ Apart from active online forum and Facebook groups, the questionnaire was distributed through the mailing list of *Ergonomie des Interfaces Homme-Machine* of the *Réseau National de télécommunications pour la Technologie, l'Enseignement et la Recherche* and the Slack of Interaction Design Association. It was also spread during the *Interaction 18* conference week in which I participated personally.

¹⁶ The first three aspects are identical as the corresponding ones described in the beginning of (see SECTION 3.4 Experience with Voice Assistants above for 'manipulation', 'stimulation', and 'identification'). The only difference is the last aspect which measures the global judgement on the product's appeal and the effects of other attributes on this judgement.

dependability, stimulation and novelty¹⁷. UEQ balances the dominance of hedonic aspects in AttrakDiff2 evaluation with pragmatic qualities (see FOOTNOTE 17 below) and is said to be more adapted to ‘professional systems’ (Lallemand and Gronier, 2015, p. 377). The model includes 26 questions which are presented in an antonymous anchor scale of 7 similar to AttrakDiff2. It covers rather thoroughly evaluation of the pragmatic, ergonomic and hedonic qualities of a system (see FOOTNOTE 17 above). This model was adopted as the basis in drafting the second section of the questionnaire for evaluating user experience of Alexa’s assistance in recognised extended usage environments.

However, considering the overall length of the questionnaire could become too long if a respondent declares his/her experience in all the three listed recognised extended usage environments (he/she would have to go through the UEQ thrice), the short version of User Experience Questionnaire (UEQ-S) was eventually used in my questionnaire to evaluate the experience of extended usage environments. UEQ-S can evaluate the pragmatic qualities and hedonic qualities¹⁸ of Alexa through 8 questions (instead of 26 questions in its extensive original version) (Schrepp *et al.*, 2017). This shortening meant abandoning detailed insights in the six aspects covered in UEQ, but could still offer hints on user preferences. Together with the third section of the questionnaire (i.e. qualitative responses on users’ preferences and expectations in the services of Alexa), this would allow clues on good qualities in the current Alexa and improvable directions for new feature introduction when designing the augmented voice assistant.

4.2 User Interview

User interviews were planned after analysing the responses from the questionnaire to further acquire qualitative data for realising the objective of the first phase of applied study (Lallemand and Gronier, 2015, pp. 47–64). These interviews would be prioritised for Alexa users who have used the voice assistant in the recognised extended usage environments in order to deepen insights into these environments; the discussion would surround their usage behaviour and habit, satisfaction and dissatisfaction of the current Alexa’s performance, and expectation of how the voice assistant could be improved. Otherwise and complementarily, interviews could also be conducted with active Alexa users to further understand generic user preferences and expectations in the voice assistant services.

In the end, user interview was omitted from the first phase of applied study because the use of Alexa in extended usage environment was uncommon amongst the respondent pool; collected opinion from responses also showed that respondents’ concerns were fairly limited by the current functions of Alexa, especially technical issues (see SECTION 5.1 Questionnaire below). This provoked the question of properness of addressing active users of current, major voice assistants on imagining how future, extended usage environment–capable, augmented voice assistants could be. This reflection has also given rise to some modifications in the anticipated methods for the second phase of applied study, in particular the recruitment of workshop participants (see Section 4.3.1 Profiles of Workshop Participants below).

¹⁷ Attractiveness measures the global appeal of the system. Perspicuity measures the effort of learning and understanding the system. Efficiency measures the ease of manipulating the system in achieving tasks. Dependability refers to the transparency and feedback of system, whether it functions as predicted and in a secure manner. Stimulation evaluates whether users find the system incite users’ interest and excitement. Novelty rates the ability to attract users’ attention through its originality. Amongst these six factors, perspicuity, efficiency and dependability represents the pragmatic, ergonomic qualities, whilst stimulation and novelty stand for hedonic qualities.

¹⁸ Pragmatic qualities describe the interaction that relates to the tasks or goals which the user aims to reach when using Alexa (e.g. the effort of learning how to use it, the efficiency of solving their tasks, and their ability to control the interaction with Alexa at each point); whereas hedonic qualities are linked to the interest-arousing and/or stimulating aspects in the interaction with Alexa (Schrepp *et al.*, 2017).

4.3 Participatory Workshop / Co-Design Workshop

By analysing collected information from the questionnaire, user concerns regarding the voice assistant example Amazon Alexa and insights into how voice assistants should be shaped for extended usage environments were obtained (see SECTION 5.1 Questionnaire below). These were brought into the second phase of applied study for designing the interaction and experience with the augmented voice assistant. Through a design workshop, the possibilities of voice user interface were explored; the interaction between target users and the augmented voice assistant in one particular usage environment. Initiated by literature review (see ‘transport’, ‘getting-around’, and ‘outdoor environments’ in SECTION 3.2 Usage Environments of Voice Assistants above; see also the possibility of *anticipating* users’ next action from *environment* sounds such as *transport* and *movement* by McMillan *et al.* (2015) in SECTION 3.3.4 Potentials of Voice Interaction above), basing on the findings of the first stage of applied study, this usage environment was set as the **commute environment**¹⁹ (see SECTION 5.1.2 Insights below), which relates the environments of *moving between two places* and *public transport*. The objectives of the workshop were to discover opportunities for **anticipatory assistance** on the basis of *augmented sensitivity* to usage environments so as to maximise user experience (as demonstrated in SECTIONS 3.2 Usage Environments of Voice Assistants, 3.3.4 Potentials of Voice Interaction, 3.4.2 Accuracy in Processing Utterances & 3.4.4 Anticipation above and SECTION 5.1.2 Insights below). This was done by first departing from existing elements in the commute environment and finding out where the augmented voice assistant could be helpful (as demonstrated in SECTION 3.2 Usage Environments of Voice Assistants above); and then creating basic prototypes of the interaction and experience of such (refer to SECTION 4.3.2 Structure and Activities of Workshop below).

4.3.1 Profiles of Workshop Participants

To amplify user aspect in building the interaction and experience, a participatory workshop was planned to begin this second phase of applied study. Desired participants would include users of experience in extended usage environments, a majority of User Experience and/or Service Designers (to keep the workshop efficient), and supplementarily interested voice assistant application developers (to maximise different perspectives in the design process). Participant selection was expected to be based on their availability, requirements of the workshop (influenced by the results of the first phase of applied study), and their experience in design practices, whilst keeping the number of participants at around six for organisational efficiency.

Apropos of the reflection—initiated by the results of the questionnaire—on the aptitude of active users of current voice assistants for designing interaction and experience of the augmented voice assistant, it was concluded that current users might not be the most desirable profile to include in this workshop as they might be too restricted to their mental model of the existing voice assistants. Users of scarce experience, on the contrary, could be more suitable as they could still hold a mental model which is more liberated from voice assistants in the contemporary reality. They could thus be less conformist to current limitations and more capacious in their imagination. Noteworthy user concerns obtained from the first phase of applied study can still be included in the workshop by highlighting them during the activities.

Since the focus of workshop participant recruitment shifted from including current voice assistant users to gathering a group of inventive people, the workshop is labelled hereinafter as *co-design workshop* instead.

Co-design workshop could allow different perspectives in realising the second phase of applied study, diversification in ideas, and a comprehensive combination of design approaches

¹⁹ The commute environment encompasses the whole period of time in which people move from one place to another with the help of public transport. It does not only cover the part when people are physically present in public transport, but also before and after the use of public transport (e.g. walking from home to a bus station and waiting for a bus there). The entire movement from point A to point B is covered.

(Anić, 2015; Mothu, 2017). Ideally, the workshop was to be conducted with the same number (six) and individuals of participants. This was practically subjected to their availability, though, and was unfortunately not the case in the real situation of this applied study—albeit the majority of the individuals were retained throughout the workshop, with a handful joining in amidst the agenda and withdrawing before the end. Hence, some adaptations in the workshop structure and the planned activities were carried out in respect of the presence alterations in the group.

4.3.2 *Structure and Activities of Workshop*

The co-design workshop was planned on a two-day time span (8 hours each day). Following a Double Diamond design approach (Design Council, 2007) of *'first designing the right thing and then designing things right'* (Nessler, 2016) and the objectives described at the beginning of this Section 4.3 Participatory Workshop / Co-Design Workshop above, the workshop structure could be concluded in five stages (see Co-Design Workshop Agenda under COMPLEMENTARY MATERIALS for full workshop agenda):

First day—Explore where to put anticipatory voice assistance in the commute environment—linearly:

- i) Introduction to voice assistant interaction and **anticipatory assistance**;
- ii) Exploration of the **commute environment** and identification of problems;
- iii) Ideation on solutions—anticipatory assistance implementations in augmented voice assistants;

Second day—Shape the experience and interaction of the anticipatory voice assistance experience in the commute environment—cyclically:

- iv) Introduction to voice user interface and shaping the experience of anticipatory voice assistance; and
- v) Prototyping the voice user interface and interaction with the augmented voice assistant.

Activities were based on the 5-day Design Sprint framework by Google Ventures (Knapp *et al.*, 2006), Gamestorming methods (Gray *et al.*, 2010), Courtney's (2017) decision-making method, Brand of Innovation's (2018) ideation workshop framework, and Stacey Seronick's framework for her *How to (Build and) test conversational / spoken UIs* workshop which I participated personally in *Interaction 18* conference week organised by Interaction Design Association (IXDA) on 5 February 2018 in Lyon, France. Relevant activities were adapted to the theme and the requirements of the co-design workshop.

INTRODUCTION (ASSISTANT INTERACTION AND ANTICIPATION)

At the beginning, participants were offered an overview of the working mechanism of interaction with current voice assistants. They were also introduced to what anticipatory assistance could be and how it could function with augmented voice assistants. These concepts were presented in an interactive manner so as to elicit memory in participants' mind as they would be of use throughout the entire workshop. Caution was taken when demonstrating these concepts, especially with the use of examples which are unrelated to commute environment, lest the imagination of participants could be limited in the following activities.

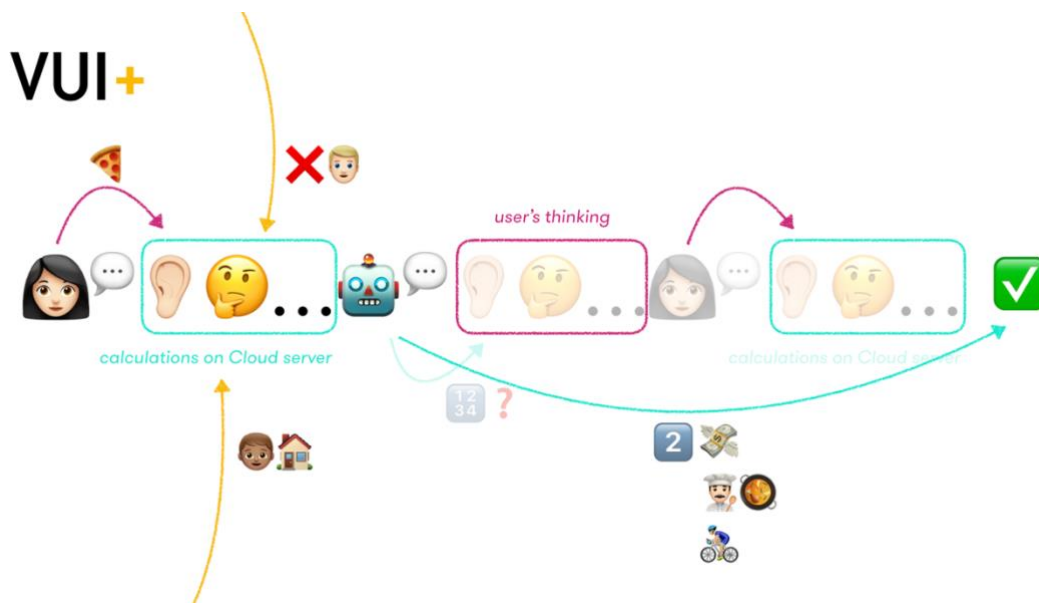


Figure 2 Example of anticipatory assistance by augmented voice assistant (ordering pizza delivery)

Through this interactive introduction, the theme of the workshop, the objective, goal factors, common understanding, and reminder questions were set²⁰. They were summarised by some reference slides which were displayed on a visible surface, called the Project Board, throughout the entire workshop, so that participants could always refresh their memory when necessary.

Participants were also advised that they could fully assume the technology needed for the functioning of the augmented voice assistant in the workshop as long as it was necessary (e.g. assuming the voice assistant is ubiquitous and users could talk to it through wireless earbuds with no Internet and battery worries, assuming the voice assistant can make use of any third-party service data sets and accurately anticipate what they wanted the assistant to). This was to make sure that the workshop focus stayed on designing the assistance interaction and experience, not diverted to ruminating on technical issues—assumed solvable presumed to be a matter of time.

EXPLORATION OF THE COMMUTE ENVIRONMENT

After the first stage of introduction, six rounds of brainstorming were conducted on a modified basis of Gray *et al.*'s (2010, pp. 78–79) 3-12-3 Brainstorm framework so as to break down the commute environment for holistic consideration (see Section 3.2.1 Definition of Usage Environment above). Each round of brainstorming included one question corresponding to an element aspect related to the commute environment, respectively:

- i) Occasions on which people might be **on commute**²¹;
- ii) Things which people might do when they are on commute;
- iii) Subjects of which people might think on commute;
- iv) Surrounding subjects and objects which might be present on commute;

²⁰ Over the entire co-design workshop, everything on the Project Board remained valid. More contents were added to the Board following the progress of the workshop. Objective here was to explore how to bring anticipatory assistance through voice assistant interaction in commute environment. Goal factors here included ensuring that users would find the assistance anticipatory, useful; and that they would love the overall experience. Goal factors also covered criteria such as making sure the conversation is maintained understood, the tasks set for the assistance are done, and the assistance offers surprises. Common understanding denoted the presented knowledge on the workshop theme, such as voice assistant interaction mechanism. Reminder questions carried the same ideas as goal factors, but were reformulated in questions, so as to allow participants to self-check and self-verify throughout the workshop when necessary.

²¹ 'On commute' refers to being in the commute environment as described in the FOOTNOTE 19 above.

- v) Needs which people might have on commute; and
- vi) Subjects and objects which people might want to avoid on commute.

This opened up to a huge pool of element ideas written by participants on individual sticky notes. When the brainstorming was done, participants were asked to present their element ideas aloud for mutualising the knowledge. Similar and related sticky notes were put in clusters question by question so as to create an Element Board. This Element Board was used to identify problems which are worth the intervention of augmented voice—especially from aspects iii), v), and vi). It could also facilitate spotting inspiration in the following activities.

Six worthwhile problems (see SECTION 5.2.1 Elements below) were identified from the pool of element ideas. They were reformulated in suggestive *how-might-we* questions on sticky notes (Courtney, 2017). A round of Dot Voting (Gray *et al.*, 2010, pp. 63–64; Courtney, 2017)—where participants were allowed 3 big sticky dots to favour however the ideas they like—was conducted to help prioritise these problems. The one with the highest priority, *‘How might we use anticipatory assistance through voice interaction to maximise safety and comfort in commute environment?’*, was brought to the next stage of the co-design workshop.



Figure 3 Dot-voting for identifying the *how-might-we* problem-to-solve

IDEATION ON SOLUTIONS

The third stage featured the Tuesday and Wednesday framework set by Google Ventures’ Design Sprint (Knapp *et al.*, 2006). With some framework adaptations, this third stage contained four timed, individual yet cooperative activities which aimed at efficiently forming ideas and gradually narrowing down to one solution which could solve the voted *how-might-we* problem in the previous stage; and a fifth collaborative discussion activity. These activities are, namely, Note-Taking, Exquisite Corpse, Crazy 8s, Storyboarding, and Speed Critique.

Participants were first given 20 minutes and asked to look around in the room at the Project Board and the Element Board to take personal notes of any important information they could find useful in the way which could facilitate their creativity. These notes should include information such as project objectives and goal factors, and also any elements in the commute environment which are worth consideration. Apart from working on already-created materials, participants were also encouraged to bring in to their notes any unconsidered, possible matter which they think noteworthy regarding voice assistant interaction, anticipatory assistance, and the commute environment. They were also advised that they might look up any reference materials online whereof they might want to make use. When the time was up, participants were expected to have lists of important information for their upcoming solution design. They were allowed 3 minutes for quickly reviewing their personal notes, tidying them up if necessary, and highlighting outstanding parts before moving on to the next activity.

In the second activity, as the name Exquisite Corpse suggested, participants were asked to mix and match across their lists of personal notes in all the different ways in order to discover opportunities for anticipatory assistance to solve the voted *how-might-we* problem while ensuring fulfilment of goal factors. This is similar to carrying out Alex Osborn's Forced Association for brainstorming across their lists. They were asked to jot down any matches personally without worrying the feasibility on voice assistants. The principle was to obtain as many variations of anticipatory assistance service as possible. Technical concerns were not in the consideration here. Participants were allowed 20 minutes for this activity, followed by another 3 minutes of reviewing their combinations to identify the one they liked the most, only which would be brought forward.

In the third activity Crazy 8s, focus was pinned on how to offer this assistance combination to users. Participants were asked to look at their favourite combination idea in different perspectives, ask themselves how another good and practical way to bring the assistance up could be (Knapp *et al.*, 2006), and to swiftly draft and/or sketch it out personally within 1 minute. Eight rounds were conducted—hence the name Crazy 8s—to obtain alternative viable approaches to realise the assistance and warm their mind up for the next activity. When participants were unable to push further alternative approaches on their single favourite assistance combination, they were advised to pick their second favourite combination from the second activity Exquisite Corpse to continue exploring alternative approaches until all the eight rounds passed.

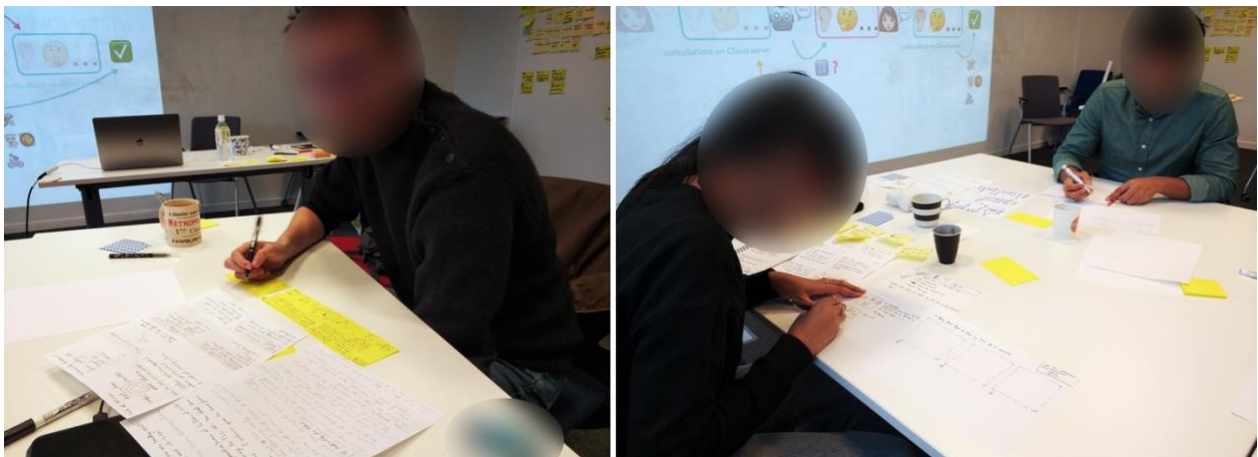


Figure 4 Alternatives created in Crazy 8s on the only Exquisite Corpse combination (left) and three-frame Storyboarding (right)

In the fourth activity Storyboarding, participants were given 30 minutes and asked to present their assistance solution on a storyboard. The storyboard was limited to three frames—so as to keep it concise—and should narrate objectively how users are going to interact with the augmented voice assistant in obtaining the anticipatory assistance in commute environment. They were asked to present their idea *in detail* on the storyboard, so that people could understand the proposed service solution in their use case by simply reading silently alone. Participants were asked to keep their storyboard anonymous and give their use case a catchy title. All these were to keep the next activity objective and efficient. In the end I collected the storyboards and stuck them on the Gallery Board, on which they were spread out spaciouly in one horizontal row for the next activity, just like paintings hung against the wall in art museums or galleries (Knapp *et al.*, 2006).

The last activity of this stage, Speed Critique, consisted first of a 15-minute independent review of all the established storyboards. Participants were asked to look into every storyboard, given unlimited number of sticky dots which they could put next to interesting and/or outstanding parts of the use cases to create a visual Heat Map, and advised to write down any concerns or questions on sticky notes below the corresponding storyboard. The second part, the critique, was

animated by me—the workshop facilitator—and timed to be 6 minutes for each storyboard. During the critique, the same procedures went through use cases one by one:

- i) I objectively narrated the story while highlighting appreciated, outstanding parts represented by the Heat Map;
- ii) I mentioned concerns and questions raised on sticky notes below the storyboard;
- iii) Meanwhile during the narration, I quickly jotted features and drawbacks of the use case on sticky notes and stuck them above the storyboards side by side;
- iv) I then asked the participants what the potential advantages and reproaches of the solution proposal could be and also jotted their ideas on sticky notes, which joined the ones already located above the storyboard in step iii), creating roughly a Pain-Gain Map (Gray *et al.*, 2010, pp. 190–191);
- v) I finally asked the solution creator to reveal his/her identity and whether we missed any ideas out;
- vi) The solution creator added any missed ideas;
- vii) We moved on to the next storyboard for a new round of critique.

This approach allowed an identical, structured review on all use cases, avoided lengthy discussions, and kept visible and objective track of expressed comments (Knapp *et al.*, 2006). This drew the focus to ideas and created a recorded analysis, which would be particularly useful for the third part of this activity—also the last one for this third stage of ideation in the entire workshop—straw poll on the only use case to be brought forward to the following prototyping activities.

Before the straw poll, participants were reminded once again of the workshop objective, the identified *how-might-we* problem, goal factors, and reminder questions. They were then allowed 10 minutes to go through all the use cases and associated sticky notes on the Gallery Board to decide silently and personally on their favourite use case, or their favourite part amongst all proposals. After that, they were given one big sticky dot each to stick next to their own decision at the same time. Each participant was then given 1 minute to explain their choice. All these votes and comments acted as consulting advice, which was meant to help the product owner—me—to decide on the use case for prototyping in the next stage of the workshop.

Following the framework of Design Sprint (Knapp *et al.*, 2006), the product owner was supposed to choose amongst the proposed use cases without making modifications to them; the chosen use case will be the scenario for the following prototyping activities. Nonetheless, looking into the use cases and comparing them to the identified *how-might-we* problem, I realised that the use cases created by workshop participants did not fully address it. Therefore, I worked on extracting outstanding parts and favourable potentials in use cases (see Figure 5 below), brought out their background purposes and/or tasks which the anticipatory assistance could realise, evaluated the possible gains and prioritised the potentials, pulled them altogether with *safety* and *comfort* identified in the *how-might-we* problem-to-solve, and fabricated two scenarios for two distinct user profiles. These two user profiles could allow better situating the dialogue interaction when offering assistance (i.e. how the augmented voice assistant should talk to its users). They were based on the same background and motive—‘*The local authority wants to introduce an anticipatory voice assistant to newcomers in the city for helping them better integrate to the new environment*’. These two user profiles were set as:

- i) Underage teenagers, whom the assistant help prepare for their new school; and
- ii) Night-time travellers, whom the assistant help explore entertainment, reach the destination, and/or go home.

These two scenarios were expected to cover the following stages of the co-design workshop. This was meant to expand the interaction for anticipatory assistance with the augmented voice assistant from a smaller, more specific target group to a bigger, more heterogeneous one in the subsequent prototyping activities.

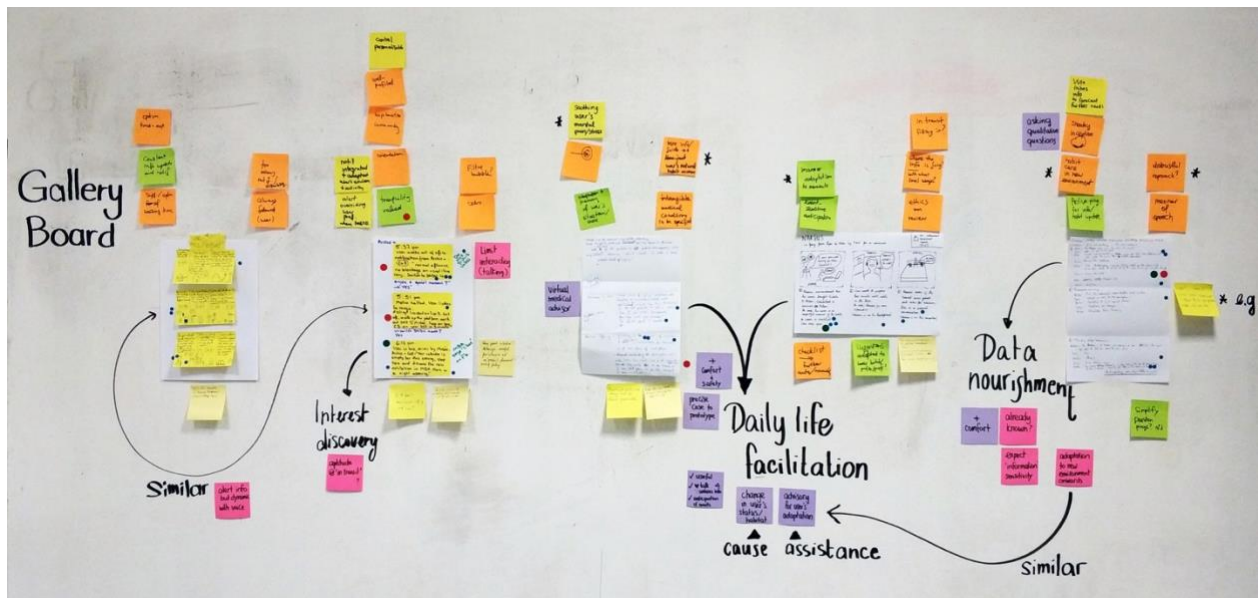


Figure 5 Gallery Board—Extraction of Snippets from Use Cases to form Scenarios and User Profiles

INTRODUCTION (VOICE USER INTERFACE)

The fourth stage of the co-design workshop was a short lecture on voice interaction and voice user interface concerns. The purpose was to prepare workshop participants on the prototyping activities in the fifth stage. Here, specific aspects in voice interaction and voice user interface design, which are uncommon in graphical interaction and graphical user interface design, were presented; principles of anticipatory design, too. The contents covered highlights from literature review, conference meetings, and personal experience in design workshops (see SECTIONS 3.3 Interaction with Voice Assistants & 3.4 Experience with Voice Assistants above). Topics included information perception and cognitive load, conversational structure and evolution, tones and image, and utterances and intents. They were not presented all at once, but alternatingly with the prototyping activities in the fifth stage, matching preceding topics with relevant activities part by part. Additional information and clarification regarding voice interaction and voice user interface design was provided to workshop participants, often upon questions, when necessary.

PROTOTYPING

The fifth stage of the workshop—prototyping—is based on Seronick's *How to (Build and) test conversational / spoken UIs* workshop framework (*Interaction 18* conference, 2018). Participants were divided into two groups. They were given the first scenario (underage teenagers; chosen here to be the first one because the assistance objective is more specific than the other scenario) which I concluded at the end of the third workshop stage (see IDEATION ON SOLUTIONS under the same SECTION 4.3.2 above) and were asked to design the voice interaction experience. They were asked to first divide the experience into blocks of actions. A brief session of presenting ideas from each group was then conducted so as to mutually inspire each other. Next, they were asked to pin these actions on concrete turn-by-turn dialogues—detailing only the parts which the augmented voice assistant would say (i.e. assistant's responses to **mappable Intents**). They were reminded that the number of dialogue turns could be reduced by *anticipation*—using external information which the assistant could gather; the Element Board served here as an inspiration source. After that, a round of cross-group user test was conducted on the produced prototypes in the form of Wizard of Oz Experiment (Gray *et al.*, 2010, p. 227)—alongside passive observation during the test—to find out improvable parts in the prototypes and gather feedback on the experience and interaction. Iteration was carried on the prototypes for refining and progressing them.



Figure 6 Participants deciding the blocks of actions (left) and filling these blocks of actions with dialogues (right)

Whilst the second concluded scenario from the end of the third workshop stage was meant to be brought up and trigger another set of prototyping activities, it was removed from the workshop agenda so as to accommodate the prototyping progress and participants' availability. Instead, participants were led to further bring deployable elements from the Element Board to enhance functions of their augmented voice assistant prototypes.

4.4 Experience Blueprinting

The prototypes from the co-design workshop were reviewed. On the basis of keeping the same intentional experience being offered to users whilst including as much potential anticipatory assistance as possible, dialogue interaction with the augmented voice assistant was further completed. These prototypes were transformed into digital, interactive prototypes on the open-source platform named *Twine*. Dialogue parts of the augmented voice assistant were written on cards, on which mappable Intents were left as options for possible users' responses. According to the response, it leads to other cards as the dialogue continues to evolve. Thus, networks of cards were constructed to picture the whole dialogue flow (see SECTION 5.3 Experience Blueprinting for the results). Background elements which are necessary for the augmented voice assistant to drive the anticipatory assistance were also marked in tags. These tags were colour-coded for clarity.

The platform *Twine* also allows a demonstration mode (called '*Play*') in which interaction through cards can be done by clicks. This created a solidier prototype for testing the entire designed dialogue. When Wizard of Oz Experiment (Gray *et al.*, 2010, p. 227) pursues, for example, the augmented voice assistant player could read verbatim what has been written and shown on the cards and select the mappable Intents according to testers' responses—this could help guarantee the objectivity of playing the role of the augmented voice assistant. Meanwhile, passive observation could be carried at the same time to mark down any irregularities. Improvement could then be made to the dialogue interaction design.

* * *

The chosen qualitative methods—together with literature review—have allowed the gain in user insights into how to improve usage environment sensitivity of voice assistants. They have helped discover the possible directions for anticipatory assistance and assisted in recognising the particular commute environment for this study. They have also led to the design of voice interaction and experience with the augmented voice assistant in the commute environment for raising qualities in user experience.

5 RESULTS AND DISCUSSION

Under the implementation of the chosen methods, the obtained findings not only respond to the research questions, but also allowed the evaluation of some subsequent methods alongside the applied study and led to some issues for discussion.

5.1 Questionnaire

5.1.1 Responses

Over the four weeks' time, 20 validated responses were collected. 14 reported that they have used Alexa in smartphone/tablet environment; 4 in private/public transport environment; 3 in outdoor-unsheltered environment. With this number of collected responses, no generalised representation of Alexa usage within the community can be concluded for section i) and ii) of the questionnaire (i.e. background information on Alexa usage and Alexa service adoption rate in extended usage environments) (see SECTION 4.1.1 Contents of the Questionnaire above). Nevertheless, they still provided information to better understand the responses for section iii) (collection of qualitative remarks on Alexa's performance in extended usage environments and user expectations).

5.1.2 Insights

While these remarks often related to current technical and performance constraints of Alexa (e.g. *'Not understanding some basic questions'*, *'Needs better integration with search engines'*, *'No support for multiple commands in a single statement yet'*, and *'Real[-]life one[-]on[-]one conversation'*), some pointed to users' desire of some non-existing functions. By interpretation, these functions could be listed as:

- Enhanced sensitivity to user profiles
 - *'If [I] ask to define a scientific term, she hears it wrong and gives me the wrong answer, or doesn't understand at all.'*
- Augmented sensitivity to usage environment context
 - *'It can't schedule actions [...] It doesn't feel like a true AI assistant because there is no sense of time. It's always in the present';*
 - *'[For hotel concierge] Room X needs another towel';*
 - *'[Making use of the microphones on Alexa-enabled devices,] when you leave your home[, Alexa could alert] you when it detects any noise or human voice.'*
- Anticipate users' needs
 - *'Reminders to take medications';*
 - *'Pausing loud music quickly when someone is trying to talk to me';*
 - *'Setting a timer as [soon as] I put food in the oven';*
 - *'Automation [in] [t]riggering [c]ommand[s] to perform over 20 [a]ction[s] from one simple request.'*
- Interconnected, cross-usage environment services
 - *'Reminders only come to the device you spoke [to]. Horrible[—]as if you ask for a reminder [for] later [or] the next day from your bedroom [device], it will not remind you at you[r] home office Echo.'*
 - *[For patients recovering from chronic illnesses] [Alexa can ask routine diagnostic questions and send the information to the doctor. This] enables the person to remain home and live their regular life without calling the doctor or traveling to report basic updates of recovery.'*
- Localised services
 - *'[I would hope to use Alexa] with local information (outside US)';*
 - *'Asking translations for [some] random languages';*
 - *'When/if Alexa is officially released in my country, I hope to be able to use her for local services like ordering food, groceries, etc.'*

Relating these insights to my research question and highlights from the literature review, expanding personal assistance *around* an extended usage environment in an anticipatory manner was set as the objective for the following second stage of applied study of voice assistants. This could respond to the proliferation of voice assistant-enabled devices: while voice assistants are becoming more ubiquitous, this objective could help coordinate different provided services and instantiate the useful role of voice assistants. Anticipating these connections on a personal scale could also improve the user experience with voice assistants (see SECTION 3.4 Experience with Voice Assistants above for the outlined criteria).

Amongst the recognised extended usage environments (smartphones/tablets, private/public transport, and outdoor-unsheltered space) (see SECTION 3.2 Usage Environments of Voice Assistants above), *public transport environment* was selected to be the centre *around* which personal, anticipatory assistance by voice assistants would be explored. As the insights from the questionnaire show that cross-usage environment services could be beneficial to users, the public transport environment has been extended to cover beyond the physical space of public transport. This does *not* limit the implementation of voice assistants to only inside public transport, but also cover the interconnecting services which voice assistants could expand to and/or integrate around the use of public transport. For clear distinction, this particular territory of usage environment is named as **commute environment** in this research. This environment was chosen on the assumption that it could lead to more yielding **anticipation** (owing to its diverse range of available elements which exist in the environment); it also crosses with a broader fan of user profiles. This would allow a greater opening for ideas in the co-design workshop and better guarantee its success. In addition, less research has been done on voice assistants in public transport environment, too.

5.2 Co-Design Workshop

5.2.1 Elements

‘Elements’ refers to the brainstorming ideas on the six aspects of the commute environment (see EXPLORATION OF THE COMMUTE ENVIRONMENT under SECTION 4.3.2 above). These elements provide inspiration for anticipatory assistance in the commute environment (see Element Ideas (Co-Design Workshop) under COMPLEMENTARY MATERIALS). While they should be kept raw and untreated lest it restrained their inspiration, they can be roughly classified as situations, triggers, and needs. Originally, as for inspiration, only the written words matter on these sticky notes. Small sticky dots were later added to them by workshop participants for highlighting the more worthwhile ideas to deploy for further enhancing anticipatory assistance (refer to the end of PROTOTYPING under SECTION 4.3.2 above).

Six *how-might-we* problems (see *How-Might-We* Questions under COMPLEMENTARY MATERIALS), which question and discuss the implementation of anticipatory assistance through voice interaction, were identified from the original element ideas. The number of big sticky dots attributed to them here represents the recognised priorities for consideration by the workshop participants (see EXPLORATION OF THE COMMUTE ENVIRONMENT under SECTION 4.3.2 above).

These problems, all starting with ‘*How might we use anticipatory assistance through voice interaction to?*’ are as follows:

- Maximise safety and comfort on commute (7 dots);
- Ensure users have [brought along] what they need (4 dots);
- Support and/or boost personal efficiency (3 dots);
- Facilitate communication with transport service providers (1 dot);
- Get transport information (0 dots); and
- Get news information (0 dots).

5.2.2 Use Cases

Use cases are the end products of the first day of the co-design workshop (i.e. at the end of the third stage). They address the *how-might-we* problem defined as ‘*How might we use anticipatory assistance through voice interaction to maximise safety and comfort in commute environment?*’ (see EXPLORATION OF THE COMMUTE ENVIRONMENT under SECTION 4.3.2 above). They were created through four rounds of selective activities, aiming to departing widely to a variety of ideas and then narrowing down one single storyboard (see IDEATION ON SOLUTIONS under SECTION 4.3.2 above). There are altogether five use cases produced, each by a workshop participant present on the first day. These use cases offer different anticipatory assistance and can be categorised as follows, sorted from more ordinary to more astounding:

- Transport Information/Alert
 - With calendar events and location information, the augmented voice assistant could anticipate the best means of transport in terms of time utility and suggest it to users.
 - With real-time traffic and transport service information, the augmented voice assistant could notify users of any irregularity and suggest alternative means of transport so as to get users to their destination on time.
 - Similarly, when users are not in a hurry, the augmented voice assistant could prioritise comfort in the transport experience over time. This could help not only bring better experience to individual users, but also contribute to crowd control in public transport services and improve experience to the passenger community.
- Interest Discovery in Neighbourhoods
 - With information about users’ liking and calendar events (users’ availability), the augmented voice assistant could suggest events happening in venues which users pass by on commute.
- User Situation Adaptation
 - With information about any changes in users’ living situation (e.g. injury), the augmented voice assistant could provide information adapted to their unprecedented needs (e.g. step-free access for railway stations, and/or step-free access from home to office).
- Habit Adaptation
 - With information about users’ habits, the augmented voice assistant could provide adapted information in new neighbourhoods where users frequent (e.g. suggesting Japanese restaurants around a new office as the user used to have Japanese food for lunch, proposing tram rides instead of the underground railway after work because the user used to find watching the streets on the journey home relaxing).

In regard to the following second day of the co-design workshop, snippets of these created use cases were extracted to reorient the focus back to *safety* and *comfort* as decided in the *how-might-we* problem-to-solve (see IDEATION ON SOLUTIONS under SECTION 4.3.2 above). This has

led to the creation of two scenarios for two user profiles, i.e. helping underage teenagers prepare for their new school, and helping night-time travellers explore entertainment, reach their destinations and/or go home.

Some use cases provide insights into voice interaction with the augmented voice assistant, too:

- Do-Not-Disturb-Unless Mode
 - Imagining that the augmented voice assistant could trigger interaction with its users, a do-not-disturb-unless mode could be included in the assistant, so that users could choose to solely receive the crucial information and interact with the assistant only when necessary.
- Data Nourishment
 - The augmented voice assistant could detect and proactively solicit information from its users (e.g. changes in office address, diet preferences)—similar to getting updates from a friend—and anticipate correspondingly the assistance which it could offer. This provides an alternative method to setting up voice assistants and/or updating information with them—which is often done through associated graphical interfaces at the present time—and maintain the interaction habit within voice.

5.2.3 Prototypes

Prototypes are the end products of the second day of the co-design workshop (i.e. at the end of the fifth stage). They are products of the given scenario (i.e. helping underage teenagers prepare for their new school; as explained in IDEATION ON SOLUTIONS under SECTION 4.3.2 above), which I concluded from the advantageous parts of use cases proposals. They are produced by expecting the maximum number of possible needs, anticipating corresponding assistance, and shaping the user experience on interacting by voice dialogues with the augmented voice assistant. There are two prototypes produced, each by a group of workshop participants present on the second day with my lead. As they are created within the limited time of the co-design workshop, the interaction and experience are only in parts; nevertheless, they still carry experience clusters which form scenes of the experience (see Figures 7 & 8 below).



Figure 7 Prototype B covers the whole school day and tends to help teenage users integrate to their new school



Figure 8 Prototype A focuses mainly on the period before school and on preparing the teenage users for their school day

Aiming at the underage teenager profile whom the assistant help prepare for their new school, the two prototypes originate from different approaches in assistance offer. While they remain two distinct augmented voice assistants and hold dissimilar clusters of experience scenes,

some parts share similar ideas. In general, these ideas of augmented voice assistants can be summarised as follows:

- Educational Voice Assistant
 - The augmented voice assistant aims to help develop independence of the teenager users. It helps users learn the sense of time and grow their own responsibility.
 - The assistant physically prepares users for school by helping them develop the proper habits (e.g. check their items needed for class, have balanced breakfast before going to school). This is similar to having a virtual nanny accompanying the teenagers.
 - The assistant mentally elicits the sense of time in users by often relating its suggestions to the amount of time needed for completing these actions. It also emphasises any effects by these actions. By presenting choices with consequences and letting users weigh and decide the action, this helps teenagers to become self-responsible and autonomous.
 - The assistant also gathers data on users' behaviour to adapt its succeeding assistance. For example, if the user was eventually late for school that day, it would be registered in the assistant and thus it could suggest setting the alarm clock and leaving home earlier for the next day.
- Discovery Voice Assistant
 - The augmented voice assistant aims to smooth the transition and help users integrate into their new environment.
 - The assistant makes use of analysed information from previously collected data (e.g. users' hobbies based on location footprints) and suggests means which users' habits could be pursued in the new neighbourhoods (e.g. suggesting wind bands contact and judo club addresses for classical music players/learners and martial arts lovers respectively). This helps users to blend into their new environment.
 - Similarly, the assistant makes use of analysed information and cross the information with other users, so as to encourage social interactions with new people around (e.g. suggesting getting to know a choir group member at school for users who like doing so). This helps users to develop their social life in their new environment.
- Infotainment Voice Assistant
 - The augmented voice assistant aims to enrich the commute experience by providing information around the new neighbourhood along the route. This is similar to having a guided tour on the way of the journey. The assistant provides location knowledge and helps users to get to know their environment. Hence, it could help users settle in their new environment personally.

Some ideas concerning the interaction with the augmented voice assistant derived from the prototypes and the activities, too:

- Automatic Do-Not-Disturb Mode
 - The augmented voice assistant could develop the sensitivity to usage environment and automatically switch to do-not-disturb mode when it is necessary (e.g. during class at school).
- Efficiency-Enjoyment Toggle
 - The augmented voice assistant could include a toggle switch which allows users to favour efficiency or enjoyment in its assistance (e.g. letting users choose

whether they would like to arrive at the destination more quickly with the underground railway or enjoy the sun with bike hire).

- Tonal Adaptation
 - The augmented voice assistant could include the sensitivity to understand users' personality and adapt to its evolution across time. This sensitivity could also ensure a flexible coverage of teenager profiles across age and characters.

5.3 Experience Blueprinting

The 'education', 'discovery' and 'infotainment' voice assistant ideas obtained from the prototyping activities in the co-design workshop were further intertwined to realise the experience for the scenario of preparing underage teenagers for their new school. As presented in SECTION 4.4 Experience Blueprinting above, on the basis of the basic prototypes from the workshop, dialogue parts of the augmented voice assistant were written on interconnected cards with the aid of the open-source platform *Twine*. They were interlinked by the mappable Intents. Further reflection was carried on the implementation of the designed anticipatory assistance, especially on the necessary information which could drive the anticipation in the background. The types of necessary information were tagged and colour-coded for distinction. This created two networks of dialogues which allow walking the experience through, providing blueprints for developing the voice assistant application once the supporting technology of the interaction mechanism could allow.

The major, unique use of anticipation in the first experience blueprint—Educational Voice Assistant—is to develop teenager users' own independence and responsibility. This can particularly be seen at utilising the time before the commute to prepare themselves for the school day and instructing them to bear responsibility when they are late for school (see Figures 9 & 10 below).

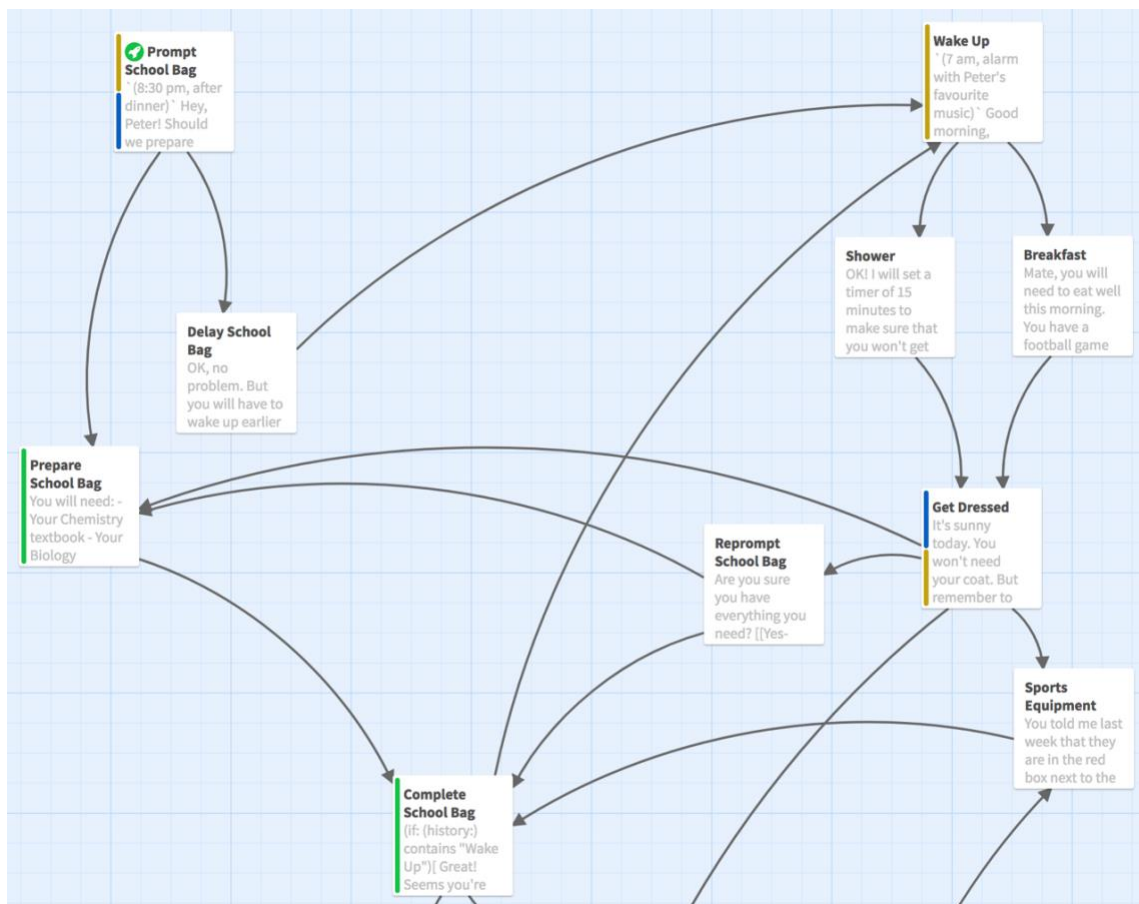


Figure 9 Educational Voice Assistant (partial) utilising the time before the commute to prepare teenage users for their school day

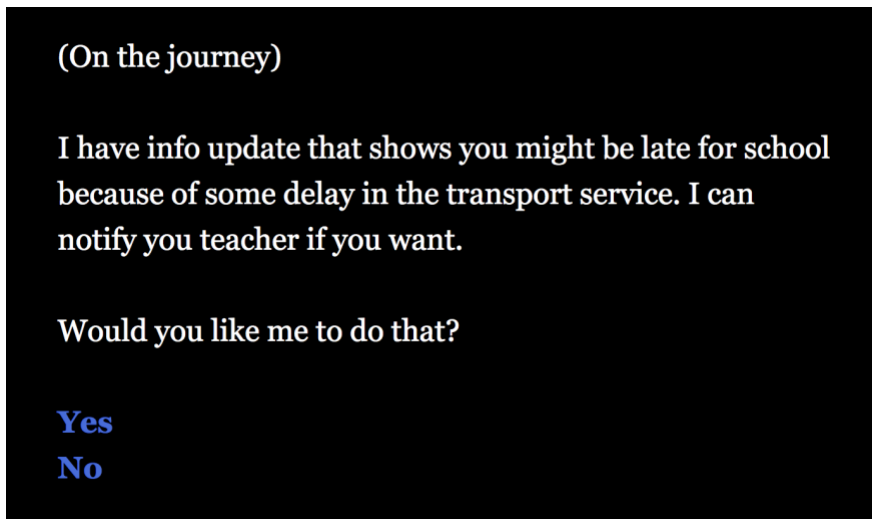


Figure 10 Educational Voice Assistant inviting teenage users to bear responsibility with transport service delay notice

For the second experience blueprint Discovery Voice Assistant, the remarkable use of anticipation is adapting teenage users to their new neighbourhood (especially with their habits and hobbies). This can be highlighted by the suggestion of after-school activities (see Figure 11).

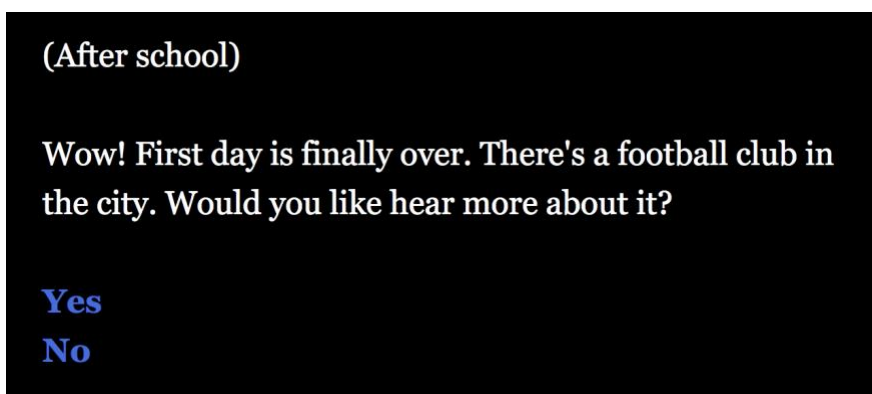


Figure 11 Discovery Voice Assistant suggesting after school activities in the neighbourhood (alternatively proposing museum visit, a different route to go home, and grocery shopping on the way)

Since *Twine* runs on either local browser memory or its dedicated offline applications, these networks are exported and are uploaded on to *GitHub*, so that everyone could simply import them to view and/or edit the contents. Prototype demonstrations, though, are directly consultable online, thanks to the free *Twine* hosting service *philome.la* offered by Colin Marc and Trucy Phan. Previews of the dialogue networks, owing to their sizes, are also uploaded and available on *GitHub*.

- Educational Voice Assistant
 - [Experience Blueprint](#) – *Twine* file downloadable on *GitHub*
 - [Demonstration](#) – click-through prototype hosted on *philome.la*
 - [Preview](#) – full-scale image of the dialogue network viewable on *GitHub*
- Discovery Voice Assistant
 - [Experience Blueprint](#) – *Twine* file downloadable on *GitHub*
 - [Demonstration](#) – click-through prototype hosted on *philome.la*
 - [Preview](#) – full-scale image of the dialogue network viewable on *GitHub*

* * *

The objective of the applied study was to start from current voice assistant users' concerns, gather insights into usage environment sensitivity, explore the possible improvements, and design the voice interaction and experience of the augmented voice assistant. In spite of the evolution and

adjustments of methods during the execution, the results have successfully shown a wide range of possibilities, with prototypes exploring more in-depth the specific commute environment.

5.4 Concerns and Highlights

Besides findings aligning with the principal objective, there are other auxiliary discoveries which are worth note-taking during the implementation of the applied study. They are explained here.

5.4.1 Concerns of Applied Study Partakers

The following issues are revealed in questionnaire responses and the co-design workshop.

USER CONTROL

While the augmented voice assistant prototype is designed with the assumption of the capability to trigger interaction with users any time, it is important to leave the options to users so that they can choose how they want to have voice interaction with the assistant. Although the augmented voice assistant could anticipate how and when it should and/or should not trigger interaction with users (e.g. during a class and/or meeting), this anticipation is still an estimation of what could suit users. Their preferences, which could be calculated on data pools, may not always be static and could vary upon time, situations, and their own state of mind. In the end, only users themselves are absolutely sure of what they want to have (and not to have). Leaving the control to users allows a back door to the augmented voice assistant. This significance is particularly discussed under ‘user control and freedom’ of Nielsen’s (1993) Heuristics.

CONNECTIVITY AND ACCESSIBILITY

In the co-design workshop, technical issues were deliberately taken out of the focus. Prototypes of the augmented voice assistant were built on the assumption of omnipresence—constant accessibility. While wireless earbuds were borrowed as an easy trampoline for imagination, in real life and with the technology nowadays, Internet connection and battery could still be the limiting factors. In spite of the necessity of the former for major voice assistants at the present time, voice assistants can be built to work offline (see SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above).

The method of how the augmented voice assistant can be accessed also weighs on the ways of interaction. The necessity and the choice of wake words, action buttons, gestures (e.g. for proximity sensors), or any other actions are still open for discussion. Since voice interaction can take place in a larger space and can be more obvious when compared to graphical user interaction (i.e. talking aloud contrasted with clicking and/or tapping graphical buttons), the way how users access the augmented voice assistant could determine how personal the assistant could be. Socio-emotional pressure in use would be a factor to take care of.

PRIVACY, ARTIFICIAL INTELLIGENCE (AI), AND DATA

Current voice assistants work on wake word detection (see SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants above). This requires always-on audio sampling for its detection. Despite the fact that this sampling is usually done locally on the voice assistant-enabled device, privacy concerns are still raised regarding wake word miscapture, triggering audio streaming by accident. Amazon Alexa’s ‘creepy laugh’ incident²² was an example (Liao, 2018).

²² It was reported that Amazon Echo speakers miscaptured ‘*Alexa, laugh*’ (as a programmed voice command) accidentally from background environment sounds and streamed a female (Alexa’s) laughing voice in response. The only thing was that users did not ask for this odd laugh and were, reported hence, horrified. Amazon’s explanation was that microphones on Echo speakers are so sensitive that they captured the programmed voice command involuntarily (Liao, 2018). They modified the command to ‘*Alexa, can you laugh?*’ and added ‘*Sure, I can laugh*’ in the assistant’s response before proceeding to laughing.

Anticipatory assistance is pictured to be supported by data and Machine Learning algorithms. The higher the amount and the more diverse the types of data which the augmented voice assistant is nourished with, the more accurate and the more personal its assistance could be. Some questionnaire respondents and co-design workshop participants voiced their ‘Internet paparazzi concern’²³ and ‘Skynet worry’²⁴. These remarks highlight that users can be sceptical about the Machine Learning and the use of data. This also implies that users still hold a certain distance from being ready to receive an AI companion. A UX study for building trust between voice assistants and users would be in particular worth while.

Besides, it is vital that the interconnectivity and mutualisation of data be well explained to users. Optimally, a platform precisely inventorying shared data, on which users could check and alter permission grant, understand the use of each source of data, and erase their shared data, should be offered. The augmented voice assistant company has to engage in data protection, too. If the use of data is not well explained, or if the shared data is not well protected, users could be surprised when they learn it later. This could engender detrimental consequences on users’ experience with the augmented voice assistant. Facebook’s scandal with Cambridge Analytica and Internet users’ responses around the *#DeleteFacebook* movement²⁵ could be taken as an example.

5.4.2 Anticipatory Assistance and Voice User Interface

In this applied study, only the commute environment was explored for anticipatory assistance in voice user interface. Nevertheless, it is worth mentioning once again that voice user interface can be used for anticipatory assistance in other usage environments which are not necessarily linked to transport or city journeys. Commuting between two points is just one of the possibilities.

Reversibly, the commute environment has a lot of room for anticipatory assistance. While the prototypes only cover a handful of suggested functions for the augmented voice assistant, there are still a lot of possibilities which have not been carried forward in the co-design workshop (see SECTION 5.2.1 Elements above) owing to the limitations of resources. That being said, these ideas could also be adopted in anticipatory assistance which does not require voice user interface. Voice user interface is just one of the means for offering assistance.

5.4.3 Assistance and User Profile

Assistance, especially anticipatory ones, is hugely dependent on user profiles. Whilst user profiles have been narrowed down to underage teenagers who have just moved to a new neighbourhood in this applied study, it remains as a niche of the entire targetable users for the augmented voice assistant. The interaction designed in the prototypes would be unfit if they were applied to other user profiles, e.g. an adult who has just moved to a new city. Nonetheless, some touchpoints might still be relatable; adjustments in the interaction would be necessary. The original method in the fifth stage of the co-design workshop (i.e. prototyping) expected designing for a second scenario (i.e. night-time travellers). This was meant to expand the interaction for anticipatory assistance with the augmented voice assistant from a smaller, more specific target group to a bigger, more heterogeneous one. Unfortunately, it has not been carried out due to organisational limitations.

²³ The concern of having one’s Internet footprints—history and cookies from digital services—followed and collected everywhere, and eventually used against one’s will.

²⁴ *Skynet* is the representation of the superintelligence in the science-fiction series *Terminator*, in which the agent gains artificial consciousness through artificial intelligence and possesses intelligence superior to mankind. Eventually it defends itself against humanity and destroys the latter.

²⁵ In March 2018, it was reported that Facebook had leaked user data to the political consulting firm Cambridge Analytica, which was believed to have taken part in influencing the presidential election of the USA in 2016 and the Brexit universal suffrage in the UK in 2016 (Cadwalladr and Graham-Harrison, 2018). As a consequence, Facebook users initiated the *#DeleteFacebook* to vent their discontent against the company’s dreadful data policy and protection engagement (Beres, 2018).

Another concern with anticipatory assistance is that it could reduce actions and/or initiatives driven by users themselves. The presence of a ubiquitous, augmented voice assistants could make information become so available that it could lessen self-motivation in taking actions or in solving problems independently. It could also diminish serendipity in daily life encounters and discoveries.

5.4.4 (Anticipatory) Assistance and Cultural Environment

Cultural aspect is a highly important factor in communication, so is it in assistance. Farai Madzima has summarised in the *Interaction 18* conference week (2018) ideas of Hofstede's (1983) and Meyer's (2014) ideas of cultural aspect in collaboration. People from different cultural backgrounds have different ways of being; particularly related to assistance, aspects including how they prefer when persuading ideas, making decisions, evaluating options, scheduling and using their time, and building trust. While these ideas are originally for collaborative work in multicultural environments, they could also be applied to address assistance offer calculations and personalisation, and furthermore, in the interaction with the voice assistant.

5.4.5 Voice Assistants and Anticipation

In the prototypes, any technology needed to achieve the desired interaction and experience was assumed. These technologies mainly require data sets and Machine Learning algorithms which nourish the augmented assistant's capability of anticipation. While this imagined mechanism is theoretically feasible (see SECTIONS 3.3.4 Potentials of Voice Interaction & 3.4 Experience with Voice Assistants for interaction and experience potentials), technology advancements are rapid and their direction could change. At the same time, users' behaviour and needs may also change over the time while technologies are being developed. The voice interaction and experience designed in this applied study might not be fully suitable by then. This is a limitation for this research and the prototyping of the voice interaction and experience design.

5.4.6 Voice Interaction and Experience Design

Users remaining as the nucleus of User-Centred Design, it is indispensable to make sure that they are thoroughly included in the design process. In this study, although they were not physically present in the design section, their concerns were still condensed from the research section and taken care of in the following steps by interpretation and representation. If time allowed a longer window for the collection of questionnaire responses, if the access to voice assistant users was more specifically introduced, and if the respondent list enabled a more comparable candidate selection for the co-design workshop, it could have beneficial to include current voice assistant users into the design section. Nevertheless, from the results of this applied study, prototypes could still be used and tested with targeted user profiles (i.e. teenagers who have just moved to a new neighbourhood) to guarantee the users' satisfaction in their experience (see SECTION 3.4 Experience with Voice Assistants above). This could be a following step for further studies and bring users closer back to the design process.

6 CONCLUSION

FULFILMENT OF OBJECTIVES

Users' experience with voice assistants is closely related to their interaction with the assistants. Voice assistants' capability of understanding users' voice input remains as a dominant factor of the interaction. It predetermines the engagement of users. This understanding can be uplifted, on the one hand, by instructing users on the comprehensible ways of interaction; on the other hand, and by enhancing voice assistants' sensitivity to usage environments' context and their processing ability on the other. Both ways can align users and voice assistants closer to each party. In this thesis, enhancing the sensitivity of voice assistants to usage environment has been explored. Notable extended environments—in which the use of voice assistants either is emerging within the user community or could benefit to their users by elevating qualities of their experience—have been first recognised as smartphones/tablets environment, private/public transport environment, and outdoor-unsheltered space environment; the focus of this study has been further narrowed down to the commute environment. By enlarging the amount and the types of usage environment information which voice assistants could process, the assistants could enrich the interaction by anticipating unspoken context. The experience, as the perceived result of the interaction by users, can be represented by the effectiveness and the efficiency voice assistants, and the satisfaction obtained through the interaction. Anticipation, as regarded as one of the means to magnify qualities of the experience, has been brought along and experimented with the services of voice assistants in the particular commute environment. This has led to the creation of in-situation voice assistant services which restores the role of the assistant as what they can serve to target users holistically and enriches their experience by providing more holistic assistance.

RESEARCH CONTRIBUTION

By investigating the correlation of the tripartite bond of voice assistants' services, i.e. users, usage environments, and the role of voice assistants, a holistic approach has been applied to designing the augmented voice assistant. It has allowed a closer association of the services which the assistant could offer, resolved the gap between hardware support and the software AI with the proposal of augmented sensitivity and anticipatory assistance, and brought back users to the centre of the product. This has enabled users to be engaged in the development, rendered voice assistants more human, and favoured its sustainability.

Besides, this study has also inventoried methods for voice interaction and experience design, in particular relating to the co-design workshop. Despite the fact that the workshop was held within a limited period of time, its skeleton could be flexibly adapted to longer incubation. The pools of ideas obtained in the applied study have provided trails ready for further studies.

FURTHER STUDIES

This study has also pointed to some areas which are worth further exploration concerning voice assistants and User Experience Design. Revealed by study partakers, trust building between users and voice assistants on privacy issues is a pivotal subject. Profile adaptation of (anticipatory) assistance can be considered as another rewarding topic.

APPENDICES

Appendix A Alexa-Enabled Devices

There are many Alexa-enabled devices, made by Amazon, third parties or both in collaboration. As of April 2018, available devices and upcoming products with built-on Alexa²⁷ could be grouped in three main categories shown below:

	Indoor, static environment					Portable environment			Driving environment
Base form	Static speaker	Tabletop accessories	Camera	Smart home control	Home entertainment	Portable speaker	Smartphone and tablet	Wearables	Car voice-control
Functionality ²⁸	Basic Alexa assistance ²⁹	Basic Alexa assistance (enhanced with visual feedback), video call	Basic Alexa assistance, algorithm-driven fashion consultation	Basic Alexa assistance, corresponding smart home control	Basic Alexa assistance, television control	Basic Alexa assistance	Basic Alexa assistance, file search on device	Basic Alexa assistance	Basic Alexa assistance, navigation, car control
Product examples	Amazon Echo, Echo Dot, Echo Plus ³⁰ ; Anker Eufy, Sonos One	Amazon Echo Show, Echo Spot	Amazon Echo Look	Ecobee4 Smart Thermostat, Ecobee Smart Light Switch	Amazon Fire TV	Amazon Tap; Invoxia Tribu	Amazon Fire Tablet, Amazon Shopping App (iOS and Android OS); Huawei Mate 9	Martian mVoice smartwatch, iMCO CoWatch, onVocal wireless headphone	Garmin Speak, BMW and MINI private cars
Start-on mechanism	On-device wake word detection, action button	On-device wake word detection, motion detection	On-device wake word detection	On-device wake word detection, action button	Action button	On-device wake word detection possible (hands-free mode), action button	Action button, on-device wake word detection possible (hands-free mode)	Action button	On-device wake word detection

Table 1 Categories of Alexa-Enabled Devices (Gordon and Wroclawski, 2018; McGrath and Hastings, 2018)

²⁷ Devices with Alexa the conversational agent installed and incorporated in the products itself—Alexa service is, therefore, accessed not through a third-party application but directly through the device.

²⁸ Functionality refers to the functionality available on the device at its out-of-the-box, factory state—without third-party configurations. Other third-party Alexa Skills and device pairing can always be added later manually by users.

²⁹ Basic Alexa assistance refers to the Alexa assistance service which can be accessed universally on all Alexa-enabled devices without particular hardware or software enhancement. Skills (applications). For details, see APPENDIX B Alexa Assistance Service.

³⁰ Echo Plus can also connect with Smart Home appliances and work as the Smart Home hub which emits control signals to supported household items.

Appendix B Alexa Assistance Service

Alexa’s assistance is based on the enabled Alexa Skills linked to the registered Amazon account (see SECTION 3.1.3 Working and Interaction Mechanism of Voice Assistants). Listed below, non-exhaustively, are some categories of functions which Alexa can execute to offer assistance universally on all types of Alexa-enabled devices:

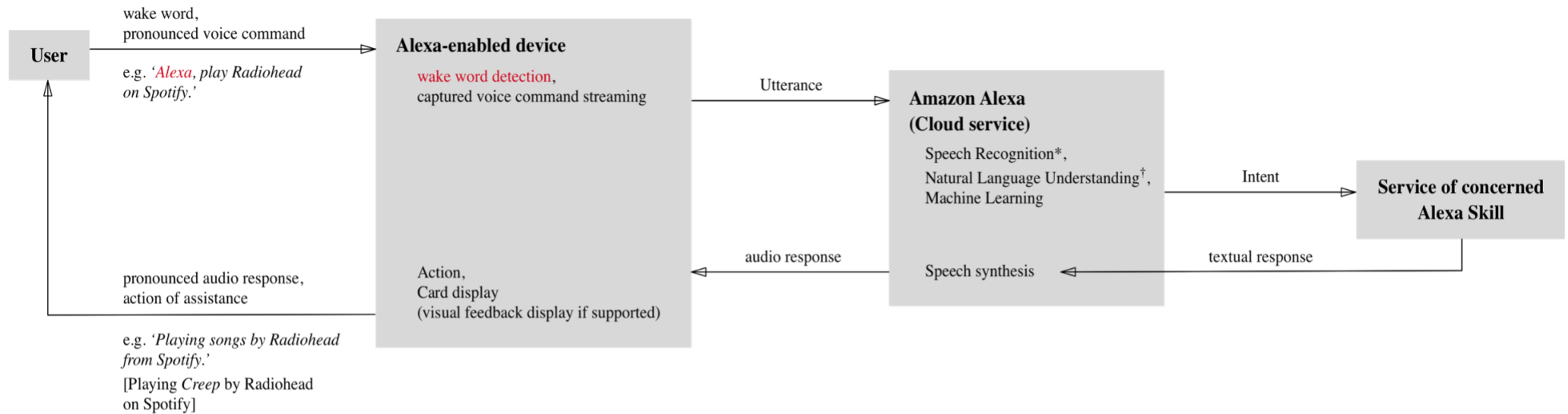
	Personal organisation	Information	Ordering	Entertainment	Home automation	Communication
Functions	Coordinate miscellaneous common organisational tasks.	Get information from the Internet	Make orders on online-supported services	Trigger audio-based entertainment	Control smart home devices connected with Alexa	Contact people and services
Examples of services	Setting up alarms, reminders, timers, making to-do lists, calendar management	Wikipedia, news, weather, sports scores, traffic information, restaurant location, dictionary lookup, fun facts	Online shopping (on Amazon), food delivery, taxi pickup	Music streaming, Internet radio streaming, podcast playback	Indoor temperature control, electricity control on sockets, diverse household appliances control	Call and send voice messages Alexa-enabled devices of people in users’ synchronised contact list, call phone numbers (US only)
Dialogue example (U: User, A: Alexa)	U: <i>Alexa, remind me to pick up my parcel.</i> A: <i>When should I remind you?</i> U: <i>At 5 pm.</i> A: <i>Okay, I’ll remind you at 5 pm.</i>	U: <i>Alexa, how’s the weather tomorrow in Lisbon?</i> A: <i>Tomorrow in Lisbon, Portugal, there will be partly sunny weather, with a high of 16 and a lot of 8.</i> U: <i>Alexa, is it going to rain?</i> A: <i>No rain is expected in Westlake today.</i> ³¹	U: <i>Alexa, ask Uber for a ride.</i> A: <i>There is an UberX 3-minute away from your home address. Should I order it?</i> U: <i>Yes.</i> A: <i>Okay, requesting your ride.</i>	U: <i>Alexa, play some jazz music.</i> A: <i>‘Coffee Table Jazz’ from Spotify.</i>	U: <i>Alexa, dim the lights to 50 per cent.</i> [Connected lights dimmed.]	U ₁ : <i>Alexa, call [name of U₂ registered in U₁’s synchronised contact list].</i> A ₁ : <i>Okay, [name of U₂].</i> A ₂ : <i>[Name of U₁] would like to talk.</i> U ₂ : <i>Alexa, answer the call.</i> [Call connection is set up.]

Table 2 Categories of Alexa’s assistance functions (Johnson, ‘Everything Amazon’s Alexa Learned to Do in 2017’, 2017; Activate, 2017)

³¹ This example illustrates exactly sequential context coherence problem highlighted in SECTION 3.3.3 Limitations of Voice Interaction. Alexa reverts to the registered device address (Westlake, Seattle, USA) when the user asks a follow-up question about the chance of rain for tomorrow in Lisbon, Portugal.

Appendix C Alexa's Working Mechanism

Alexa works on voice command dialogues. Users can trigger interactions with Alexa by pronouncing the registered wake word ('Alexa' by default). The succeeding treatment procedures of the mechanism is illustrated below (Doring and Srivastava, 2017):



* Speech recognition here can be understood as the process of speech-to-text, in which the audio of user's voice command is transcribed to texts for algorithmic computations.

† In the Utterance example, Natural Language Understanding process could be briefly presented as:

'Alexa, play Radiohead on Spotify.'

where colours represent correspondingly:

wake word detection (done previously locally on the Alexa-enabled device),

Intent of playing music,

Slot of artist name, and

Slot of required service.

Figure 12 Alexa's Working Mechanism (Doring and Srivastava, 2017)

GLOSSARY

Agency

Agency is the ability and need to effectively act, with the objective of realising goals.

Agent

Agent is one which acts with a goal-oriented purpose or purposes.

Anticipatory assistance

Anticipatory assistance is the assistance offered without specific, designated **request**.

(Automatic) Speech Recognition (ASR)

Automatic Speech Recognition is the process which transformed streamed audio into written text.

(Voice assistant) application

A voice assistant application is often an additional software program which can enhance voice assistants' services. It is often activated manually by the users on the application store with the users' personal account.

Artificial intelligence (AI)

Any device which perceives its environment and takes actions that maximise its chance of success at some goal is said to possess artificial intelligence.

Augmented voice assistant

An augmented voice assistant is a voice assistant which possesses augmented sensitivity (when compared to the current state) to **usage environments**. It also can make use of the information about usage environments to anticipate assistance and/or services for users.

Cloud (service)

Cloud service refers to the information technology which transfers data from a device to a remote server and vice versa, and which enables data treatment and storage.

Commute environment, on commute

The commute environment encompasses the **environment** along the entire period of time when people move from one place to another with the help of public transport.

Conversation

Conversation refers to the exchange of words between humans.

Dialogue

Dialogue refers to the exchange of words between human(s) and machine(s).

(Usage) Environment

Usage environment refers to the space in which voice assistants are being accessed. It comprises i) tangible data and metadata of the environment which the voice assistant is being accessed (e.g. time, location, and users' profile) and ii) more intangible information of the background context which is associated to the moment when the voice assistant is being accessed (e.g. context of the surrounding space, users' emotion, and users' preferences).

Extended usage environment

An extended usage environment of a voice assistant is a **usage environment** which has not yet been officially supported the voice assistant company. In an extended usage environment, the service performance of the voice assistant is often only partial.

Intent

An Intent represents a programmed action that can be recognised from the processed **Utterance** through **Natural Language Understanding (NLU)** (see APPENDIX C Alexa's Working Mechanism). Intents can optionally have arguments called **Slots**.

Mappable Intents

Mappable Intents refer to the range of multiple **Utterances** included in a (voice assistant) application for the same **Intent**. The more mappable Utterances that an Intent has, the higher the accuracy that the machine can have in **Natural Language Understanding (NLU)** for the Intent identification; hence, the higher the flexibility that the application allows users to voice their requests.

Natural language

Language evolved naturally through human-to-human communication by repetition of use and implicit imitation, spoken by human in the habitual manner of speech without particular adaptation to interactions with machines.

Natural Language Understanding (NLU)

The process done by the machine to interpret **natural language** spoken by humans and received by the machine (i.e. **Utterance**). In Amazon's **Cloud service**, this process is part of the processing together with **Automatic Speech Recognition (ASR)** (recognising spoken words in context) and Machine Learning (seeking patterns in pronounced words in a spoken phrase) (see APPENDIX C Alexa's Working Mechanism). This process recognises **Intent(s)** and **Slot(s)**.

Request

A request is a spoken phrase pronounced by the user in the purpose of interacting with a machine (synonymous with **voice command**).

(Usage environment) sensitivity

Voice assistants' capability to receive information existing in the **usage environment** in which they are being accessed by users.

(Alexa) Skill

Alexa Skills are coded **voice assistant applications** dedicated for the voice assistant Amazon Alexa.

Slot

A Slot is a perimeter variable which can be necessary to perform certain **Intents**. Slots carry information to specify Intents. For example, an Intent for fetching weather forecast may require the Slot dedicated for geographical reference point.

Speech Synthesis

The process to generate speech audio by machines; an artificial production of spoken speech based on phonetic compositions and recordings to form pronounced words in the way of mimicking human speech.

Utterance

An utterance is a spoken text phrase by the user (to the voice assistant; synonymous with **voice command** and **request**). An Utterance is a recognised text phrase transformed from the streamed voice command or request. It is the product of **Automatic Speech Recognition (ASR)**. It is used in **Natural Language Understanding (NLU)** to identify any **Intent** (a programmed action to be done by the machine).

Voice assistant-enabled devices

Hardware devices which connect to voice assistant services (often run by **Cloud service** in remote servers).

Voice command

A voice command is a spoken word or phrase pronounced by the user in the purpose of interacting with a machine (synonymous with **request**).

Voice (user) interface (VUI)

Voice user interface is the interaction mode based on the use of voice and speech in the communication between a machine and a user with the objective to access an automated process/service.

Wake word

Wake word is the word which the user can trigger interactions with the voice assistant. For example, 'Alexa' is the wake word by default for Amazon Alexa. Users can choose and change the wake word to 'Amazon', 'Echo', or 'Computer'.

BIBLIOGRAPHY

- Abran, Alain, et al. 'Consolidating the ISO Usability Models'. In *Submitted to the 11th International Software Quality Management Conference and the 8th Annual INSPIRE Conference*, vol. 2003, 2003.
- Activate. *Activate Tech & Media Outlook 2018*. www.slideshare.net/ActivateInc/activate-tech-media-outlook-2018.
- Amazon. 'Alexa and Alexa Device FAQs'. *Amazon.Com Help & Customer Service*, 9 Nov. 2014, www.amazon.com/gp/help/customer/display.html?nodeId=201602230.
- . 'Alexa Voice Service (AVS) International'. *Amazon Developer - Amazon Alexa*, 10 Feb. 2017, developer.amazon.com/alexa-voice-service/international/.
- . 'Amazon Echo Now Available to All Customers'. *Business Wire*, 23 June 2015, www.businesswire.com/news/home/20150623005861/en/Amazon-Echo-Customers.
- Amyx, Scott. 'Wearing Your Intelligence: How to Apply Artificial Intelligence in Wearables and IoT | WIRED'. *Wired*, Dec. 2014, www.wired.com/insights/2014/12/wearing-your-intelligence/.
- Anders, George. "'Alexa, Understand Me'". *MIT Technology Review*, 9 Aug. 2017, www.technologyreview.com/s/608571/alexa-understand-me/.
- Anić, Ines. 'Participatory Design: What It Is, and What Makes It so Great'. *Medium*, 4 Nov. 2015, medium.com/@i_anic/participatory-design-what-it-is-and-what-makes-it-so-great-6bad619e8735#.s9c6ee19i.
- Austin, Darren. 'This Is How Amazon's Alexa Hooks You'. *InVision Blog*, 18 July 2017, www.invisionapp.com/blog/amazon-alexa-hook/.
- Baig, Edward C. 'Amazon Alexa Is Coming to \$1,000 Smart Glasses — and PCs'. *USA TODAY*, 8 Jan. 2018, www.usatoday.com/story/tech/columnist/baig/2018/01/07/ces-2018-alexa-what-im-looking/1011288001/.
- Bangor, Aaron, et al. 'Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale'. *Journal of Usability Studies*, vol. 4, no. 3, May 2009, pp. 114–123.
- Beres, Damon. "'Delete Facebook" Searches Hit 5-Year High on Google after Cambridge Analytica Crisis'. *Mashable*, 22 Mar. 2018, mashable.com/2018/03/22/delete-facebook-searches-cambridge-analytica/.
- Berg, Markus M. *Modelling of Natural Dialogues in the Context of Speech-Based Information and Control Systems*. Christian-Albrechts-Universität zu Kiel, 24 July 2014. *AKA Verlag*, macau.uni-kiel.de/receive/dissertation_diss_00016739.
- Blankenburg, Jeff. 'Amazon Echo Devices Rolling Out to Customers in Over 80 Countries'. *Alexa Blogs*, 8 Dec. 2017, developer.amazon.com/blogs/alexa/post/3d8b2e1e-d8c5-41f6-a193-5456a590ef9a/amazon-echo-now-available-to-customers-in-over-60-countries.
- Bouchard, Anthony. 'Apple Opens Siri to Developers with iOS 10's New SiriKit API'. *IDownloadBlog*, 13 June 2016, www.idownloadblog.com/2016/06/13/apple-opens-siri-to-developers-with-ios-10/.
- Brand of Innovation. 'Ideation Workshop: Tools in PDF and 1-Day Agenda Template'. *Board of Innovation*, Jan. 2018, www.boardofinnovation.com/ideation-workshop-agenda-template-examples/.
- Burgess, Brian. 'How to Stop Google Assistant, Siri, Cortana, and Alexa Active Listening'. *GroovyPost*, 29 Aug. 2017, www.groovypost.com/howto/stop-google-assistant-siri-cortana-alexa-active-listening/.
- Bushnell, Mona, and B2B Tech Writer. 'AI Faceoff: Siri vs. Cortana vs. Google Assistant vs. Alexa'. *Business News Daily*, 8 Jan. 2018, www.businessnewsdaily.com/10315-siri-cortana-google-assistant-amazon-alexa-face-off.html.
- Cadwalladr, Carole, and Emma Graham-Harrison. 'Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach'. *The Guardian*, 17 Mar. 2018, www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.
- Canbek, Nil Goksel, and Mehmet Emin Mutlu. 'On the Track of Artificial Intelligence: Learning with Intelligent Personal Assistants'. *Journal of Human Sciences*, vol. 13, no. 1, Jan. 2016, pp. 592–601. NA, www.j-humansciences.com, doi:10.14687/ijhs.v13i1.3549.
- Chachada, Sachin, and C. C. Jay Kuo. 'Environmental Sound Recognition: A Survey'. *APSIPA Transactions on Signal and Information Processing*, vol. 3, Jan. 2014. *Cambridge Core*, doi:10.1017/ATSIP.2014.12.
- Chen, Mei-Ling, and Hao-Chuan Wang. 'Bridging the Communication Gap: Conceptualizing Issues of Learnability in Using Intelligent Personal Assistants'. *Talking with Conversational Agents in Collaborative Action*, 2017, talkingwithagents.wordpress.com/position-papers/.
- Coucke, Alice, et al. 'Benchmarking Natural Language Understanding Systems: Google, Facebook, Microsoft, Amazon, and Snips'. *Medium*, 2 June 2017, medium.com/snips-ai/benchmarking-natural-language-understanding-systems-google-facebook-microsoft-and-snips-2b8ddcf9fb19.
- Courtney, Jonathan. 'Use This Exercise to Solve Any Product Design Challenge'. *Muzli*, 29 Mar. 2017, [medium.muz.li/a-super-simple-exercise-for-solving-almost-any-product-design-challenge-f9e6c0019d7d](https://medium.com/muzli/a-super-simple-exercise-for-solving-almost-any-product-design-challenge-f9e6c0019d7d).
- Crook, Jordan. 'Kohler Wants to Make Your Bathroom Smarter with Konnect'. *TechCrunch*, 8 Jan. 2018, social.techcrunch.com/2018/01/08/kohler-wants-to-make-your-bathroom-smarter-with-konnect/.
- Design Council. *11 Lessons: Managing Design in Global Brands*. Design Council, 20 Jan. 2007, www.designcouncil.org.uk/resources/report/11-lessons-managing-design-global-brands.
- Doring, Memo, and Akersh Srivastava. *Voice Design 101: An Introduction*. go.amazonappservices.com/E0i0J0ISM1w0j0PM0W0fs0j. Amazon Alexa Webinar - Voice Design 101: An Introduction.
- Dulin, Oleg. 'What I Learned from Using Amazon Alexa for a Month'. *InfoWorld*, 19 July 2016, www.infoworld.com/article/3097278/personal-technology/what-i-learned-from-using-amazon-alexa-for-a-month.html.

- Eagle, Nathan, and Alex (Sandy) Pentland. 'Reality Mining: Sensing Complex Social Systems'. *Personal and Ubiquitous Computing*, vol. 10, no. 4, Mar. 2006, pp. 255–268. *ACM Digital Library*, doi:10.1007/s00779-005-0046-3.
- Earley, Seth. 'Lessons from Alexa: Artificial Intelligence and Machine Learning Use Cases'. *Earley Information Science*, Mar. 2016, www.earley.com/blog/lessons-alexa-artificial-intelligence-and-machine-learning-use-cases.
- Eyal, Nir. *Hooked: How to Build Habit-Forming Products*. 1st ed., Portfolio, 2014.
- Falkner, Stephanie. 'An Introduction To Intelligent Personal Assistants'. *Chatbots Magazine*, 28 Mar. 2018, chatbotsmagazine.com/an-introduction-to-intelligent-personal-assistants-37c07aa7c5ab.
- Franklin, Stan, and Art Graesser. 'Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents'. *Proceedings of the Workshop on Intelligent Agents III, Agent Theories, Architectures, and Languages*, Springer-Verlag, 1997, pp. 21–35. *ACM Digital Library*, dl.acm.org/citation.cfm?id=648203.749270.
- Gaines, Cork. 'I Installed an Amazon Echo Dot in a Car and It Was the Best Infotainment System I've Ever Used'. *Business Insider Deutschland*, 13 July 2017, www.businessinsider.com/using-amazon-echo-dot-in-a-car-2017-7.
- Gitlin, Jonathan M. 'Dragon Drive Is the Best Car Voice Activation System We've Spoken to'. *Ars Technica*, 12 Jan. 2016, arstechnica.com/cars/2016/01/dragon-drive-is-the-best-car-voice-activation-system-weve-spoken-to/.
- . 'Infotainment Bad, Blind Spot and Lane-Departure Warnings Good, Studies Say'. *Ars Technica*, 6 Oct. 2017, arstechnica.com/cars/2017/10/infotainment-bad-blind-spot-and-lane-departure-warnings-good-studies-say/.
- Gkogka, Eleana. 'Easter Eggs, Little Delights in UX Design'. *Muzli - Design Inspiration*, 10 Apr. 2017, medium.muz.li/easter-eggs-little-delights-in-ux-design-fa26911cd8a3.
- Gonzalez, Robbie. 'The Real Problem With Voice Assistants Like Siri Is Your Brain'. *WIRED*, 28 Sept. 2017, www.wired.com/story/multitasking-problem-with-virtual-assistants/.
- Goode, Lauren. 'Alexa, When Will You Work Well Outside the Home?' *The Verge*, 6 Jan. 2017, www.theverge.com/ces/2017/1/6/14186174/amazon-alexa-ces-2017-smart-home-ford-wearables.
- Gordon, Samantha, and Daniel Wroclawski. 'Everything That Works with Amazon Echo and Alexa'. *Reviewed.Com Smarthome*, 30 Jan. 2018, smarthome.reviewed.com/features/everything-that-works-with-amazon-echo-alexa.
- Gray, Dave, et al. *Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers*. 1 edition, O'Reilly Media, 2010.
- Guzman, Andrea. 'Making AI Safe for Humans: A Conversation With Siri'. *Socialbots and Their Friends: Digital Media and the Automation of Sociality*, Routledge, 2017, pp. 69–85.
- Harris, Randy Allen. *Voice Interaction Design: Crafting the New Conversational Speech Systems*. Morgan Kaufmann Publishers Inc., 2005.
- Hassenzahl, Marc, et al. 'AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. [AttrakDiff: A questionnaire for the measurement of perceived hedonic and pragmatic quality]'. *Mensch & Computer 2003*, Vieweg+Teubner Verlag, 2003, pp. 187–96. *link.springer.com*, doi:10.1007/978-3-322-80058-9_19.
- . 'The Thing and I: Understanding the Relationship Between User and Product'. *Funology: From Usability to Enjoyment*, Springer Science & Business Media, 2004, pp. 31–42.
- Headrick, Daniel. 'Is Amazon's Alexa Really Always Listening?' *Reviewed.Com Smarthome*, 5 Feb. 2018, smarthome.reviewed.com/features/is-amazons-alexa-really-always-listening.
- Hetzner, Christiaan. 'VW Joins Ford, BMW to Offer Amazon's "Alexa" Voice Service'. *Automotive News*, 20 Jan. 2017, www.autonews.com/article/20170120/copy01/301209965/vw-joins-ford-bmw-to-offer-amazons-alexa-voice-service.
- Hofstede, Geert. 'The Cultural Relativity of Organizational Practices and Theories'. *Journal of International Business Studies*, vol. 14, no. 2, June 1983, pp. 75–89. *link.springer.com*, doi:10.1057/palgrave.jibs.8490867.
- Isbitski, David. *Introducing the Alexa Skills Kit, Enabling Developers to Create Entirely New Voice Driven Capabilities: Alexa Blogs*. 25 June 2015, developer.amazon.com/blogs/alexa/post/Tx205N9U1UD338H/introducing-the-alexa-skills-kit-enabling-developers-to-create-entirely-new-voice-driven-capabilities.
- Ísleifsdóttir, Jónheiður, and Marta Kristín Lárusdóttir. 'Measuring the User Experience of a Task Oriented Software'. *Valid Useful User Experience Measurement*, Institute of Research in Informatics of Toulouse (IRIT), 2008, pp. 97–102. *ResearchGate*, www.researchgate.net/publication/259823102_Valid_Useful_User_Experience_Measurement.
- Johnson, Khari. 'Alexa and Google Assistant Should Tell You When the next Bus Is Coming'. *VentureBeat*, 22 Nov. 2017, venturebeat.com/2017/11/22/alexa-and-google-assistant-should-tell-you-when-the-next-bus-is-coming/.
- . 'Amazon Alexa Skills Can Now Access User Location for Services Such as Food Delivery'. *VentureBeat*, 5 Apr. 2017, venturebeat.com/2017/04/05/amazon-alexa-skills-can-now-access-user-location-for-services-such-as-food-delivery/.
- . 'Everything Amazon's Alexa Learned to Do in 2017'. *VentureBeat*, 29 Dec. 2017, venturebeat.com/2017/12/29/everything-amazons-alexa-learned-to-do-in-2017/.
- . 'Google Assistant SDK Launches for Developers and Device Makers'. *VentureBeat*, 27 Apr. 2017, venturebeat.com/2017/04/27/google-assistant-sdk-launches-for-developers-and-device-makers/.
- Kapitall. 'Why Cortana Assistant Can Help Microsoft in the Smartphone Market'. *TheStreet*, 18 Mar. 2014, www.thestreet.com/story/12534433/1/why-cortana-assistant-can-help-microsoft-in-the-smartphone-market.html.
- Kelly, Heather. 'Voice Assistants Are Coming to Some Very Weird Places'. *CNNMoney*, 17 Jan. 2018, money.cnn.com/2018/01/17/technology/ces-voice-alexa-google/index.html.
- Kinsella, Bret. 'Amazon Alexa Skill Count Surpasses 30,000 in the U.S. - Voicebot'. *Voicebot*, 22 Mar. 2018, www.voicebot.ai/2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/.
- Kinsella, Bret, and Ava Mutchler. *Voicebot Smart Speaker Consumer Adoption Report*. Voicebot, Mar. 2018, www.voicebot.ai/wp-content/uploads/2018/03/the_smart_speaker_consumer_adoption_report_2018_voicebot.pdf.

- Knapp, Jake, et al. *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days*. Simon & Schuster, 2016.
- Knight, Will. 'AI's Language Problem'. *MIT Technology Review*, 9 Aug. 2016, www.technologyreview.com/s/602094/ai-language-problem/.
- Kosner, Anthony Wing. 'Can Siri Make Up For With Outright Personality What She Lacks In Data?' *Forbes*, 28 Jan. 2013, www.forbes.com/sites/anthonykosner/2013/01/28/can-siri-make-up-for-with-outright-personality-what-she-lacks-in-data/.
- Kujala, Sari, et al. 'UX Curve: A Method for Evaluating Long-Term User Experience'. *Interacting with Computers*, vol. 23, no. 5, Sept. 2011, pp. 473–83. *academic.oup.com*, doi:10.1016/j.intcom.2011.06.005.
- Lacoma, Tyler. 'Virtual Assistant Comparison: Cortana, Google Assistant, Siri, Alexa, Bixby'. *Digital Trends*, 29 Aug. 2017, www.digitaltrends.com/computing/cortana-vs-siri-vs-google-now/.
- Lallemand, Carine, and Guillaume Gronier. *Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Eyrolles, 2015.
- Laugwitz, Bettina, et al. 'Construction and Evaluation of a User Experience Questionnaire'. *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work*, Springer-Verlag, 2008, pp. 63–76. *ACM Digital Library*, doi:10.1007/978-3-540-89350-9_6.
- Lee, Kwan, et al. *Social-Psychological Origins of Feelings of Presence: Creating Social Presence With Machine-Generated Voices*. Vol. 7, 2005. *ResearchGate*, doi:10.1207/S1532785XMEP0701_2.
- Liao, Shannon. 'Amazon Has a Fix for Alexa's Creepy Laughs'. *The Verge*, 7 Mar. 2018, www.theverge.com/circuitbreaker/2018/3/7/17092334/amazon-alexa-devices-strange-laughter.
- Liszewski, Andrew. 'Forget BMW, the Garmin Speak Is a Cheaper Way to Put Alexa in Your Car'. *Gizmodo*, 17 Oct. 2017, gizmodo.com/forget-bmw-the-garmin-speak-is-a-cheaper-way-to-put-al-1819510039.
- Luger, Ewa, and Abigail Sellen. "'Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents'. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2016, pp. 5286–5297. *ACM Digital Library*, doi:10.1145/2858036.2858288.
- Lydon, John E., et al. 'Interpersonal Similarity and the Social and Intellectual Dimensions of First Impressions'. *Social Cognition*, vol. 6, no. 4, Dec. 1988, pp. 269–86. *guilfordjournals.com (Atypon)*, doi:10.1521/soco.1988.6.4.269.
- Lyonnais, Sheena. 'Ask An UXpert: What Does Holistic UX Design Mean To You?' *Adobe Blog*, 27 June 2017, theblog.adobe.com/ask-an-uxpert-what-does-holistic-ux-design-mean-to-you/.
- Maier, Steven F., and Martin E. Seligman. *Learned Helplessness: Theory and Evidence*. Vol. 105, 1976. *ResearchGate*, doi:10.1037/0096-3445.105.1.3.
- Maring, Joe. 'Google Assistant Is Coming Soon to Chromebooks That Aren't the Pixelbook'. *Android Central*, 8 Feb. 2018, www.androidcentral.com/google-assistant-coming-soon-chromebooks-arent-pixelbook.
- McGrath, Jenny, and Nick Hastings. 'The Best Alexa-Enabled Devices You Can Utilize with Amazon's Echo Lineup'. *Digital Trends*, 26 Feb. 2018, www.digitaltrends.com/home/everything-that-works-with-amazon-echo-smart-home/.
- McMillan, Donald. 'Conversational Agents and Implicit Interaction'. *Talking with Conversational Agents in Collaborative Action*, 2017, talkingwithagents.wordpress.com/position-papers/.
- . 'Repurposing Conversation: Experiments with the Continuous Speech Stream'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 3953–3962. *ACM Digital Library*, doi:10.1145/2702123.2702532.
- Meyer, Erin. *The Culture Map: Breaking Through the Invisible Boundaries of Global Business*. 1 edition, PublicAffairs, 2014, www.safaribooksonline.com/library/view/the-culture-map/9781610392594/.
- Moorjani, Yogesh. 'The Psychology of Anticipatory Design'. *UX Collective*, 10 Jan. 2018, uxdesign.cc/design-experiences-that-are-one-step-ahead-2254e98d07b2.
- Mothu, Audrey. 'Participatory Design Tools and Methods'. *The UX Blog*, 3 Feb. 2017, medium.theuxblog.com/participatory-design-tools-and-methods-741543b1ff5b.
- Moynihan, Tim. 'Alexa and Google Home Record What You Say. But What Happens to That Data?' *WIRED*, 5 Dec. 2016, www.wired.com/2016/12/alexa-and-google-record-your-voice/.
- Nass, Clifford, et al. *Can Computer Personalities Be Human Personalities?* Vol. 43, 1995. *ResearchGate*, doi:10.1145/223355.223538.
- Needleman, Rafe. 'Google I/O Day One: Google Continues Attacks on Apple, Amazon'. *CNET*, 27 June 2012, www.cnet.com/news/google-io-day-one-google-continues-attacks-on-apple-amazon/.
- Nessler, Dan. 'How to Apply a Design Thinking, HCD, UX or Any Creative Process from Scratch'. *Medium*, 19 May 2016, medium.com/digital-experience-design/how-to-apply-a-design-thinking-hcd-ux-or-any-creative-process-from-scratch-b8786efbf812.
- Nielsen, Jakob. *Usability Engineering*. Morgan Kaufmann Publishers Inc., 1993.
- Norman, Donald A. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, 2004.
- Novak, John, et al. 'Communication, Machines & Human Augmentics'. *Communication +1*, vol. 5, no. 1, Sept. 2016, pp. 1–33. *ResearchGate*, doi:10.7275/R5QR4V2D.
- O'Donnelly, Charlotte. 'The Artificial Intelligence Arms Race: Alexa, Cortana, Google Assistant, Siri, Viv... and What's Next'. *Blue Hill Research*, 12 Oct. 2016, bluehillresearch.com/the-artificial-intelligence-arms-race-alexa-cortana-google-assistant-siri-viv-and-whats-next/.

- Ong, Thuy. 'Garmin's Speak Puts Alexa-Powered Navigation in Your Car'. *The Verge*, 17 Oct. 2017, www.theverge.com/circuitbreaker/2017/10/17/16481284/garmin-speak-amazon-alexa-car.
- Osborne, Joe. 'Windows 10 Redstone 4 Will Turn Cortana into an Interactive User Manual'. *TechRadar*, 23 Mar. 2018, www.techradar.com/news/windows-10-redstone-4-will-turn-cortana-into-an-interactive-user-manual.
- Pall, Gurpreet Singh. 'Cortana Skills Kit Empowers Developers to Build Intelligent Experiences for Millions of Users'. *Windows Developer Blog*, 10 May 2017, blogs.windows.com/buildingapps/2017/05/10/cortana-skills-kit-empowers-developers-build-intelligent-experiences-millions-users/.
- Pearl, Cathy. *Designing Voice User Interfaces: Principles of Conversational Experiences*. 1st ed., O'Reilly Media, Inc., 2016.
- Perez, Sarah. '42 Percent of Smart Speaker Owners Have Bought a Second Device (or More)'. *TechCrunch*, 23 June 2017, social.techcrunch.com/2017/06/23/42-percent-of-amazon-echo-owners-have-bought-a-second-device-or-more/.
- . 'Alexa Is Coming to the Amazon App on Android, Starting This Week'. *TechCrunch*, 20 July 2017, social.techcrunch.com/2017/07/20/alexa-is-coming-to-the-amazon-app-on-android-starting-this-week/.
- . 'Alexa Is Coming to Wearable Devices, Including Headphones, Smartwatches and Fitness Trackers'. *TechCrunch*, 5 Jan. 2018, social.techcrunch.com/2018/01/05/alexa-is-coming-to-wearable-devices-including-headphones-smartwatches-and-fitness-trackers/.
- . 'Amazon Adds Alexa to Its Main Shopping App'. *TechCrunch*, 16 Mar. 2017, social.techcrunch.com/2017/03/16/amazon-adds-alexa-to-its-main-shopping-app/.
- . 'Amazon's Alexa Passes 15,000 Skills, up from 10,000 in February'. *TechCrunch*, 4 July 2017, social.techcrunch.com/2017/07/03/amazons-alexa-passes-15000-skills-up-from-10000-in-february/.
- Porcheron, Martin, et al. 'Talking with Conversational Agents in Collaborative Action'. *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ACM, 2017, pp. 431–436. *ACM Digital Library*, doi:10.1145/3022198.3022666.
- Radziwill, Nicole M., and Morgan C. Benton. 'Evaluating Quality of Chatbots and Intelligent Conversational Agents'. *ArXiv:1704.04579 [Cs]*, Apr. 2017. *arXiv.org*, arxiv.org/abs/1704.04579.
- Reeves, Stuart. 'Some Conversational Challenges of Talking with Machines'. *Talking with Conversational Agents in Collaborative Action*, 2017. *University of Nottingham*, eprints.nottingham.ac.uk/40510/.
- Rowinski, Dan. 'Virtual Personal Assistants & The Future Of Your Smartphone [Infographic]'. *ReadWrite*, 15 Jan. 2013, readwrite.com/2013/01/15/virtual-personal-assistants-the-future-of-your-smartphone-infographic/.
- Sauro, Jeff. *MeasuringU: Measuring Usability with the System Usability Scale (SUS)*. 2 Feb. 2011, measuringu.com/sus/.
- Schrepp, Martin, et al. 'Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)'. *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. Regular Issue, 2017. *international journal of interactive Multimedia and Artificial Intelligence*, doi:10.9781/ijimai.2017.09.001.
- Shakeel, Bisma, et al. 'Siri – Apple's Personal Assistant: A Review'. *International Journal of Computer Science and Mobile Computing*, vol. 6, no. 7, July 2017, pp. 44–48.
- Sharma, Gaurav. 'The Rise of Intelligent "Voice" Assistants (I.A.)'. *Chatbots Magazine*, 29 May 2016, chatbotsmagazine.com/intelligent-assistants-i-a-85c21f9d3b8e.
- Shenoy, Gourav G., et al. 'Leveling the Playing Field for Visually Impaired Using Transport Assistant'. *ArXiv:1703.02103 [Cs]*, Mar. 2017. *arXiv.org*, arxiv.org/abs/1703.02103.
- Statt, Nick. 'Apple to Release MacOS Sierra on September 20th'. *The Verge*, 7 Sept. 2016, www.theverge.com/2016/9/7/12758218/apple-mac-os-sierra-update-release-date-announced.
- Tapp, Meg. *Sassy Siri: Does Voice Search Need a Personality?* 27 Oct. 2015, www.branded3.com/blog/sassy-siri-does-voice-search-need-a-personality/.
- The Todoist Team. 'Keep Track of Your Tasks Hands-Free with the New Alexa Integration'. *Todoist Blog*, 13 Oct. 2016, blog.todoist.com/2016/10/13/todoist-alexa-integration/.
- Turunen, Markku. 'Adaptive Interaction Methods in Speech User Interfaces'. *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2001, pp. 91–92. *ACM Digital Library*, doi:10.1145/634067.634123.
- Turunen, Markku, and Jaakko Hakulinen. 'Jaspis - A Framework for Multilingual Adaptive Speech Applications'. *ICSLP 2000 / INTERSPEECH 2000*, 2000, www.researchgate.net/publication/221487328_Jaspis_-_A_framework_for_multilingual_adaptive_speech_applications. ResearchGate.
- Usability.gov. 'Interaction Design Basics'. *Usability.Gov*, 19 Feb. 2014, www.usability.gov/what-and-why/interaction-design.html.
- Velazco, Chris. 'Apple Reveals Siri Voice Interface: The "Intelligent Assistant" Only For iPhone 4S'. *TechCrunch*, 4 Oct. 2011, social.techcrunch.com/2011/10/04/apple-reveals-siri-voice-interface-the-intelligent-assistant/.
- VoiceLabs. 'The 2017 Voice Report'. *VoiceLabs*, 15 Jan. 2017, voicelabs.co/2017/01/15/the-2017-voice-report/.
- Whitenton, Kathryn. 'Voice First: The Future of Interaction?' *Nielsen Norman Group*, 12 Nov. 2017, www.nngroup.com/articles/voice-first/.
- Wollerton, Megan. 'Taking Alexa for a Ride'. *CNET*, 7 Apr. 2016, www.cnet.com/news/taking-amazons-alexa-for-a-ride/.

COMPLEMENTARY MATERIALS

TABLE OF CONTENTS

Co-Design Workshop Agenda	55
Day 1	56
Day 2	57
Element Ideas (Co-Design Workshop)	58
Overview of Element Board.....	59
Occasions on which people might be on commute	60
Things which people might do when they are on commute.....	61
Subjects of which people might think on commute	62
Surrounding subjects and objects which might be present on commute.....	63
Needs which people might have on commute.....	64
Subjects and objects which people might want to avoid on commute	65
<i>How-Might-We</i> Questions	65

LIST OF TABLES

Table 3 Co-Design Workshop Agenda—Day 156
Table 4 Co-Design Workshop Agenda—Day 257

LIST OF FIGURES

Figure 13 Overview of Element Board (generated from brainstorming activities) 59
Figure 14 Occasions on which people might be on commute—environment context and time ... 60
Figure 15 Things which people might do when they are on commute—environment context and activities 61
Figure 16 Subjects of which people might think on commute—possible needs62
Figure 17 Surrounding subjects and objects which might be present on commute—environment, space, and possible triggers.....63
Figure 18 Needs which people might have on commute 64
Figure 19 Subjects and objects which people might want to avoid on commute 65
Figure 20 How-might-we questions and the voted how-might-we problem 65

Co-Design Workshop Agenda

Day 1

Time: 9 am – 5 pm

Objective: Explore where to put anticipatory voice assistance in the commute environment
(short commutes up until travelling between cities—no long trips)

Activity	Duration	Reference Time
[Setting the scene] Introduction	55 mins	9:20 ►
Explain workshop	5 mins	
Introduce people	15 mins	
Set goal, map VUI & name fear	40 mins	
[Start with the problem] Exploration	75 mins	
(Demo & questions)	10 mins	
Brainstorming	20 mins (3–5 mins per question)	
Present and cluster ideas	30 mins	
How might we	10 mins	
Dot voting	5 mins	
Lunch break	90 mins	11:50 ►
[Ideation]	90 mins	13:50 ►
Setting the afternoon scene	15 mins	
Look around and take notes	20 mins	
Note-Taking & Exquisite Corpse	20 mins	
Crazy 8s	8 mins	
Storyboarding	20 mins	
Break	15 mins	► 15:15
[Decide the use case] Decision	80 mins	15:15 ►
Silent exploration & heat map	10 mins	
Speed Critique & Pain 'n' Gain	30 mins	
Straw poll	10 mins	
Explain poll	10 mins	
Sharing	15 mins	

Table 3 Co-Design Workshop Agenda—Day 1

Day 2

Time: 9 am – 5 pm

Objective: Shape the experience and interaction of the anticipatory voice assistance experience in the commute environment

Activity	Duration	Reference Time
Introduction	45 mins	9:10 ►
(Introduce people)	15 mins	
Concerns of VUI	20 mins	
Refresh memory of use case	10 mins	
Prototyping	100 mins	10:00 ►
Scenario 1: Shape dialogue elements and settings	30 mins	
Break	10 mins	► 10:40
Group sharing	10 mins	
Shape dialogue turns (assistants' utterances)	60 mins	
Lunch break	90 mins	12:00 ►
Test and Iterate	60 mins	13:30 ►
Explaining Test	10 mins	
Wizard of Oz Experiment	20 mins	
Iteration	20 mins	► 14:20
Group Sharing	10 mins	
Expansion	100 mins	14:30 ►
Explaining Scenario 2	10 mins	
Dialogue blocks/topics	30 mins	
Break	15 mins	15:15 ► 15:30
Shape utterances	60 mins	► 16:40
Final Sharing	15 mins	

Table 4 Co-Design Workshop Agenda—Day 2

Element Ideas (Co-Design Workshop)

Overview of Element Board



Figure 13 Overview of Element Board (generated from brainstorming activities)

Occasions on which people might be on commute



Figure 14 Occasions on which people might be on commute—environment context and time

Things which people might do when they are on commute

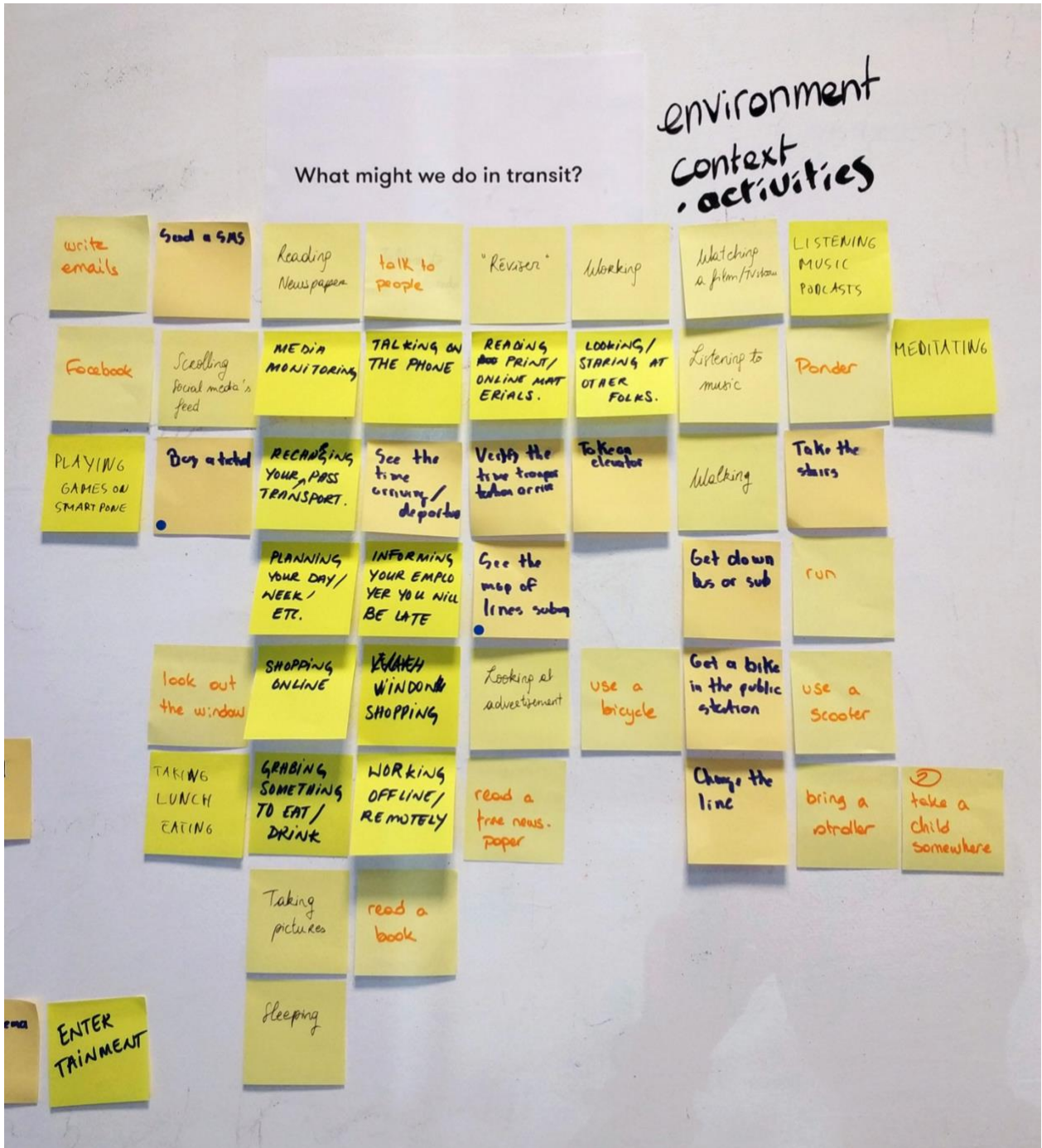


Figure 15 Things which people might do when they are on commute—environment context and activities

Subjects of which people might think on commute

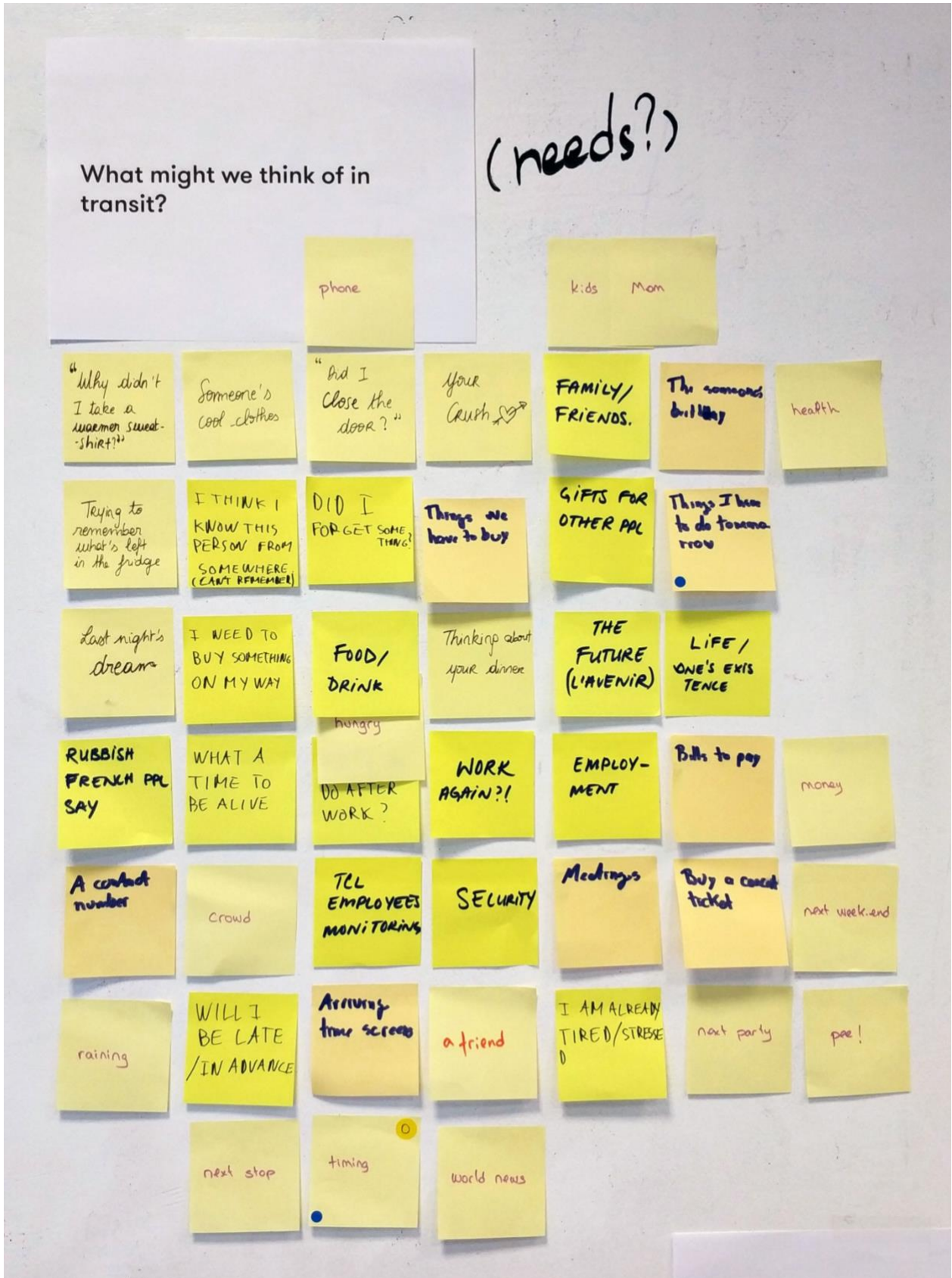


Figure 16 Subjects of which people might think on commute—possible needs

Surrounding subjects and objects which might be present on commute

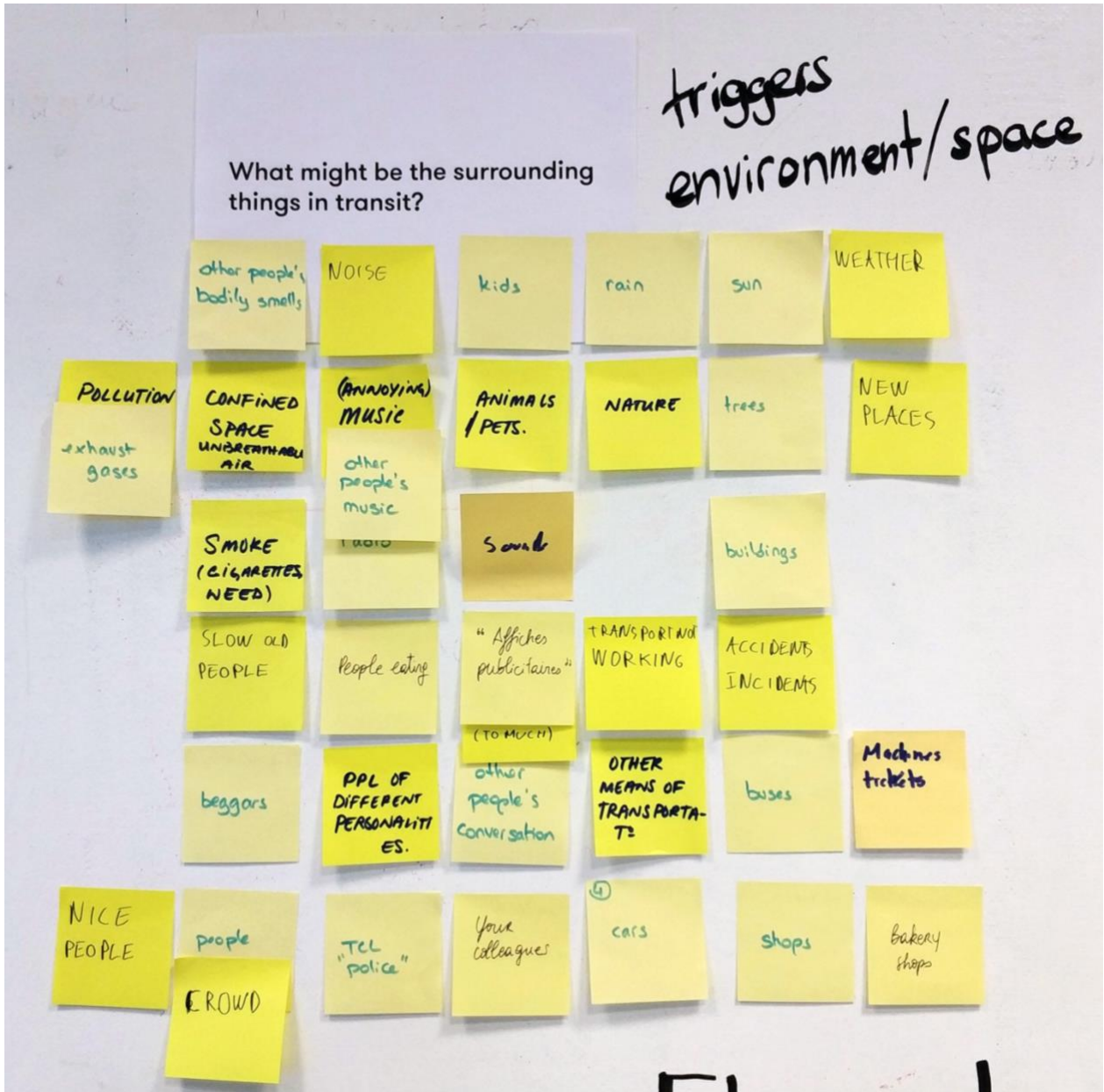


Figure 17 Surrounding subjects and objects which might be present on commute—environment, space, and possible triggers

Needs which people might have on commute

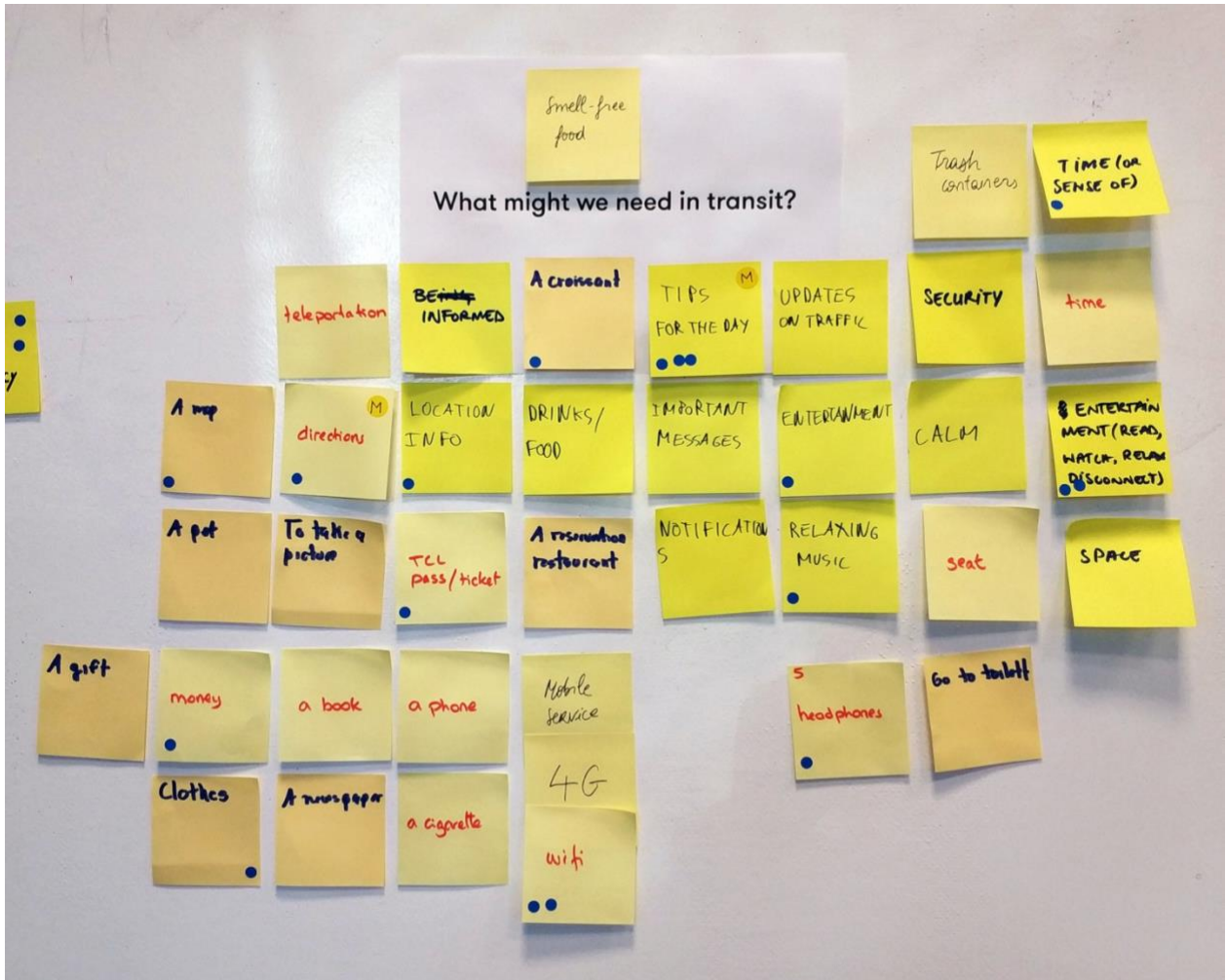


Figure 18 Needs which people might have on commute

Subjects and objects which people might want to avoid on commute

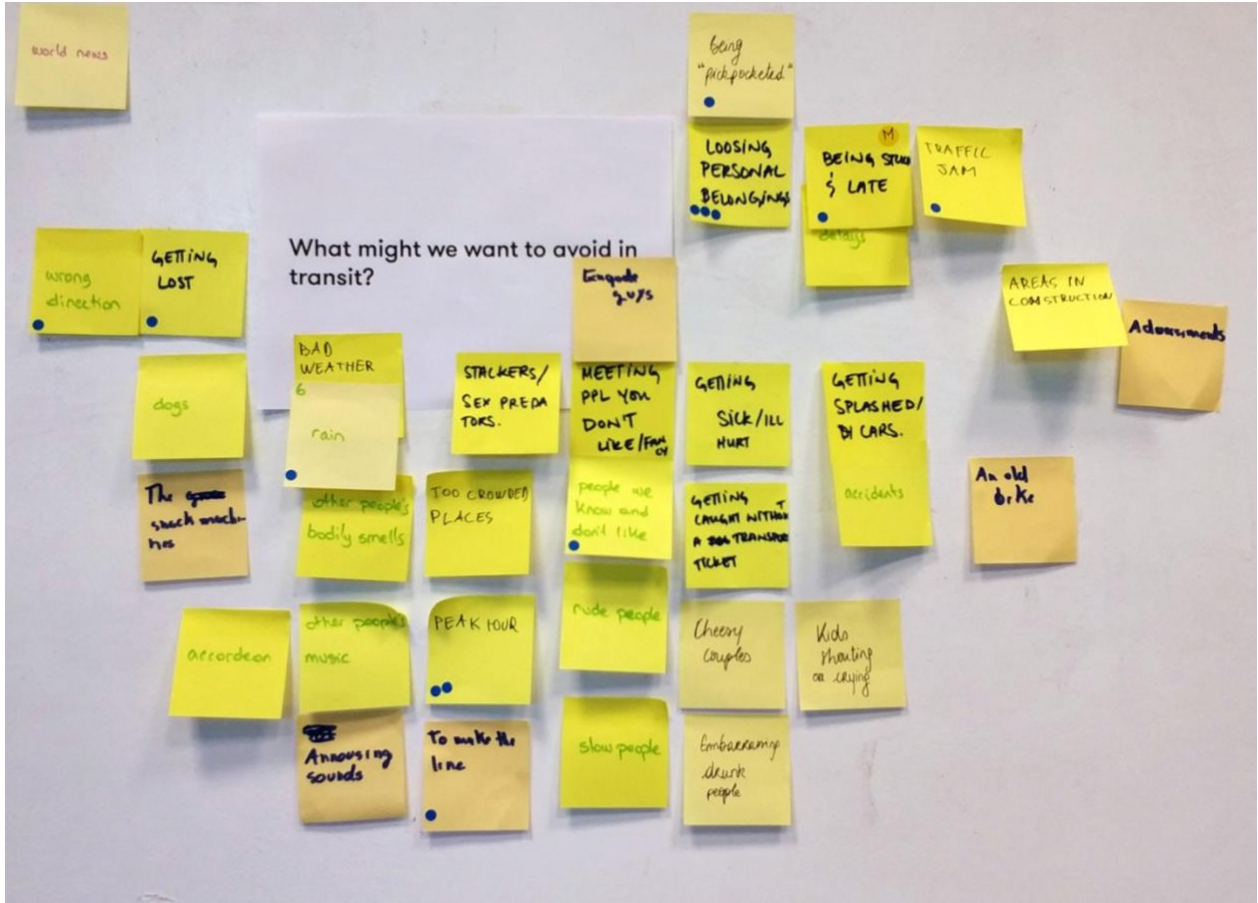


Figure 19 Subjects and objects which people might want to avoid on commute

How-Might-We Questions

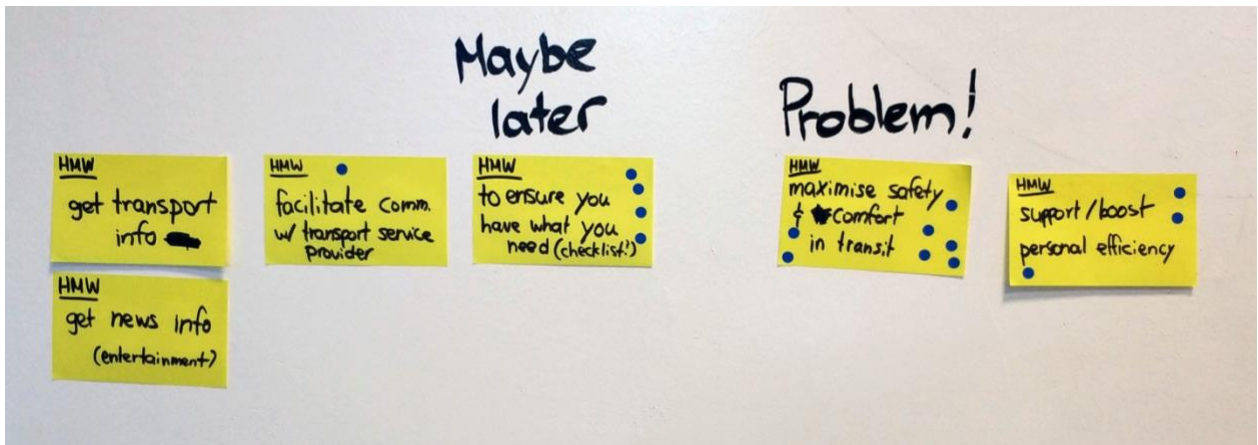


Figure 20 How-might-we questions and the voted how-might-we problem