Intelligibility Issues Faced by Smart Speaker Enthusiasts in Understanding What Their Devices Do and Why

Mirzel Avdic
Department of Computer Science, Aarhus University,
Aarhus, Denmark
miavd18@cs.au.dk

Jo Vermeulen
Department of Computer Science, Aarhus University,
Aarhus, Denmark
jo.vermeulen@cs.au.dk

ABSTRACT

Studies of smart speakers highlight issues people face with understanding why unexpected behaviour occurs and with recovering from mistakes due to uninformative responses. Yet, our understanding of such *intelligibility* issues in smart speakers — difficulties in understanding the device's behaviour — remains limited, in particular, for long-term and frequent smart speaker users who may encounter more complex situations than first-time users. We conducted an online survey and interviews with smart speaker enthusiasts to investigate how they form an understanding of the device's behaviour and what strategies they use to recover from breakdowns. We identified seven different breakdown recovery strategies and found that enthusiasts particularly struggled with breakdowns in their IoT infrastructure. Informed by our results, we propose three research directions: infrastructural breakdowns as learning opportunities for understanding the smart speaker's behaviour; leveraging aspects of non-verbal communication as opportunities for design; and considering passive users' intelligibility and control needs.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI; Human computer interaction (HCI); Interaction paradigms; Natural language interfaces.

KEYWORDS

Smart Speakers, Intelligent Personal Assistants, Voice User Interfaces, Intelligibility, Breakdowns, Shared Usage

ACM Reference Format:
Mirzel Avdic and Jo Vermeulen. 2020. Intelligibility Issues Faced by Smart Speaker Enthusiasts in Understanding What Their Devices Do and Why. In *32nd Australian Conference on Human-Computer Interaction (OzCHI '20), December 02–04, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3441000.3441068

1 INTRODUCTION

Smart speakers are increasingly gaining a foothold in people's homes. In the US alone, there are currently more than 66 million adult smart speaker owners [66]. The concept of smart speakers is threefold: users interact with (1) Intelligent Personal Assistants (IPAs) through (2) the physical artefact (i.e. the smart speaker itself) by using primarily (3) a Voice User Interface (VUI). Examples of such IPAs include Alexa, Siri, and Google Assistant. IPAs aim to assist in various tasks/activities like cooking, checking facts, playing music, and making calls. Smart speakers also increasingly play a role in home automation [7, 38], acting as a hub that interfaces with various smart home appliances such as smart light bulbs and thermostats.

Despite their widespread use and popularity, recent research has identified several issues with IPAs in smart speakers. For example, prior studies have shown that they do not have appropriately designed conversational skills [6, 55, 56, 58, 62], which is further exacerbated by the difficulty of processing natural language [28, 47, 48]. Additionally, smart speakers' generic cylindrical form and lack of a display provide little information for users to infer its state, capabilities [30], and behaviour; which may draw comparisons to "*notions of a* '*black box*"' as argued by Porcheron et al. [56]. Smart speakers are also part of a long history of context-aware technologies [20] or so-called 'sensing systems' [4]. Researchers have pointed out the need for sensing systems to be *intelligible* [5], namely, inform users of what they infer, how they infer this, and what they are doing with that information. The issues regarding black box behaviour that prior work touches upon (e.g. [56]), suggest that smart speakers' behaviour is not always intelligible to users from a conversational perspective. This includes whether the smart speaker is attending to the user's input, whether it has correctly recognized the users' spoken utterances, whether the user is using the right voice command or whether the smart speaker is capable of responding to a specific type of query [6, 15, 47, 48, 56]. Yet, it is unclear what intelligibility issues users encounter beyond these conversational issues. While most studies to date have focused on first-time users, frequent and longer-term users of smart speakers may have different needs, use their device in different ways, and may encounter breakdowns in more complex situations than first-time users. For instance, Bentley et al. observed that the use of automation increased over time in smart speaker users [7]. This suggests that a common trend with these users may be that they integrate their smart speakers into a larger smart home setup over time, where they can interconnect and integrate with a large number of other smart home devices, which may lead to additional intelligibility issues with respect to the device's behaviour within this larger smart home infrastructure [44]. However, we do not know yet what intelligibility issues frequent smart speaker users encounter.

To address this gap, we contribute a detailed investigation of intelligibility issues with IPAs experienced by a user group that we

OzCHI '20, December 02–04, 2020, Sydney, NSW, Australia

ACM ISBN 978-1-4503-8975-4/20/12…$15.00
https://doi.org/10.1145/3441000.3441068

define as *smart speaker enthusiasts*: people who use smart speakers frequently and have done so over a longer period of time (i.e., more than two months), are excited about the technology, and often share their experiences among peers. In particular, we present what enthusiasts technically understand about a wide range of smart speakers they own and how they conceptualize these devices' behaviour. Moreover, we look at how enthusiasts address their smart speakers, when they encounter unintelligible behaviour, how they recover from such breakdowns, and how these issues are handled in multi-user settings and in conjunction with other Internet of Things (IoT) devices. We conducted two studies with enthusiasts: an online survey (N=102) in which respondents were asked to perform six tasks with their smart speakers, and semi-structured interviews (N=12) to complement the survey. From our findings, we distill three future research directions: infrastructural breakdowns as learning opportunities for understanding the speaker's behaviour; leveraging non-verbal communication as opportunities for design; and considering the intelligibility and control needs of passive users.

2 RELATED WORK

2.1 Intelligibility

The notion of computers retreating into the background and the use of natural ways to interact with those computers were part of early visions of ubiquitous computing and calm computing [69]. Bellotti and Edwards [5] argued for such context-aware systems [20] to be *intelligible*, i.e. to provide users with an understanding of how the system is "*interpreting the state of the world*". Researchers have provided several frameworks for intelligibility [4, 5, 17, 37], investigated intelligibility issues in commercial products and concepts [17, 73], and demonstrated possible ways to provide intelligibility such as graphical interfaces [21, 34, 67], explanations [33, 36, 37, 68] or shape-change [54]. Despite the benefit of increased transparency, intelligibility can also be harmful if a system behaves appropriately yet shows high uncertainty [35].

2.2 Intelligibility of IPAs in Smart Speakers

Designing more intelligible voice-based IPAs is challenging and the recent advances in artificial intelligence (AI) and autonomous systems have again moved intelligibility issues to the forefront of the HCI community [1]. While there are several VUI guidelines [30, 45, 74], some argue that we lack guidelines for voice-based human-AI interaction [2]. Amershi et al. [2] present 18 guidelines for human-AI interaction and observe that AI systems such as voice assistants without any Graphical User Interfaces (GUIs) are the least compatible with these guidelines, indicating a need for further research into efficient human-AI interaction for products such as smart speakers that primarily use VUIs. While manufacturers have started producing complementary smart displays for additional feedback (e.g., Google Nest Hub), Sciuto et al. [60] found that some users prefer to hide their smart speakers away from direct view for aesthetic reasons. This indicates that IPAs would not necessarily benefit from displays as a medium to provide intelligibility [5].

2.2.1 When Smart Speakers Break Down. Breakdowns as described by Winograd and Flores [71] are common phenomena where seamless interactions with the world through artefacts get disrupted by a shift of focus toward artefacts since they stop working. Breakdowns offer opportunities for users to learn about how artefacts work by inspecting them closely, which is a core element of intelligibility [5]. Porcheron et al. [56] also suggest that designers of smart speakers consider the given responses by the system as "*the design of interactional resources for users*" to understand and overcome breakdowns and misconceptions, pointing out the unhelpful responses.

Similarly, Beneteau et al. [6] conducted a study on how families collaboratively repaired communication breakdowns with Amazon Alexa. The authors observed how a mother *directly instructed* her son on how to get the desired outcome by telling him what and how to say it – teaching her son about the cause of the breakdown. This shows that people do attempt to learn from the breakdowns. Beneteau et al. [6] also propose that IPAs become more adaptive to the breakdown and provide users with useful responses about what to do to overcome communication breakdowns.

These studies [6, 56] show that inexperienced users often experience conversational breakdowns with IPAs, i.e. the speakers suffer from *conversational* intelligibility issues. What remains insufficiently investigated with smart speakers are potential *infrastructural* intelligibility issues [23], which may occur with users who have IoT ecosystems in place that are connected to the smart speaker such as our smart speaker enthusiasts. Infrastructural breakdowns could remain hidden to the users due to the focus on seamlessness and minimalistic design of the devices [56], potentially hindering users from learning about types of system errors occurring in IoT ecosystems. Users may suspect a natural language processing (NLP) error as it is the most frequent type of error [47, 48], yet it is unclear how users deal with non-NLP errors in IoT ecosystems.

2.2.2 Users' Understanding of IPAs in Smart Speakers. Cho [15] examined first-time users' mental models of IPAs in smart speakers, showing that they use *push* and *pull* strategies for error handling. When participants use the push strategy, they provide the IPA with more contextual information to ensure that the IPA interprets the request correctly. When participants fail to get a desired answer, they employ the pull strategy, in which they use broader terms to test the boundaries of the IPA's comprehension. Similarly, Myers et al. [47] found that using an unfamiliar voice-based calendar on a smart speaker display (e.g. Echo Show) presents obstacles for users. They showed that hyperarticulation is a popular tactic against the most common type of error: NLP errors. These studies [15, 47] show promising results about what strategies users employ to overcome breakdowns. However, it remains unclear whether such strategies are common among enthusiasts and to what extent these and other strategies are used in contexts that go beyond NLP errors, like controlling smart home devices.

2.3 Multi-User Experience with Smart Speakers

Several studies on smart speakers mention a multi-user aspect. Household members share smart speakers [38] and in some cases use them simultaneously in a single session [6, 22, 56]. A few studies [6, 56, 59] investigated how couples and families interacted with smart speakers, indicating that smart speakers are becoming a part

of households with multiple inhabitants. However, it is unclear if and how smart speaker owners share their devices with others outside of their households (e.g., friends or guests), and how owners handle potential interaction challenges and breakdowns in such scenarios, in particular with interconnected IoT devices. Lau et al. [32] identified scenarios in which primary users of smart speakers would (un)intentionally exclude secondary or incidental users such as partners, children, and guests from participating and interacting with the smart speaker. The idea of making users more aware and accountable in interactions with technologies in social settings is not new [51], yet it remains an issue in the context of smart speakers.

2.4 Summary

Related work reveals that when studies touch upon intelligibility issues of IPAs in smart speakers, they mostly discuss conversational breakdowns that first-time users experience. It remains unclear to which extent and in which situations smart speakers lack intelligibility for enthusiasts, how these enthusiasts deal with, and help others (e.g. friends or guests) deal with breakdowns, including *infrastructure* issues [23] that involve other IoT devices, and how enthusiasts conceptualize their smart speakers' behaviour.

3 APPROACH

3.1 Research Questions

To address the gaps pointed out in prior research, we deployed an online survey and conducted semi-structured interviews with smart speaker enthusiasts to investigate the following research questions:

- **RQ1**: How do enthusiasts conceptualize their smart speaker's behaviour?
- **RQ2**: How do enthusiasts address their smart speaker? Do they approach the device and face it, and what other modalities do they use besides speech?
- **RQ3**: What strategies do enthusiasts employ to recover from mistakes and system breakdowns?
- **RQ4**: How do enthusiasts use their smart speaker with others in their households and/or when having visitors?

Due to the lack of visual feedback, unclear state, and issues with discoverability of available commands, smart speakers suffer from potential issues with each of Bellotti et al.'s five questions for designers of sensing systems [4]. We briefly summarize these below:

- Address: How do I address one (or more) of many possible devices?
- Accident: How do I avoid mistakes?
- Attention: How do I know the system is ready and attending to my actions?
- Action: How do I effect a meaningful action, control its extent and possibly specify a target or targets for my action?
- Alignment: How do I know the system is doing (has done) the right thing?

Research questions **RQ2** and **RQ3** are directly inspired by Bellotti et al.'s [4] design concerns *address* (in our case, how can users direct communication or avoid directing communication to the smart speaker?), and *accident* (how can users avoid or recover from errors?). The reason for these two design concerns as our outset is due to their strong relations to breakdowns. We hypothesise that breakdowns might disrupt the voice interaction flow with the smart speaker, which raised questions such as whether breakdowns made enthusiasts approach or orient themselves differently around the device. We were also interested in potential strategies enthusiasts used to recover from breakdowns. Enthusiasts' understanding determines how they handle errors, the way they trust and address their smart speaker, and might reveal their conceptualizations of smart speakers (**RQ1**). The way enthusiasts communicate with smart speakers (**RQ2**) can provide new insights into strategies to provide intelligibility. In our results, we will also touch upon Bellotti et al.'s other questions [4]: *attention* (how do users establish that the smart speaker is attending?), *action* (how do users discover the available commands? [30]), and *alignment* (how do users know the smart speaker is doing the right thing?). **RQ4** is motivated by prior research [32] suggesting that primary smart speaker users (un)intentionally exclude secondary or incidental users from interacting with the smart speaker. In this study, we investigated whether this was the case for enthusiasts and how enthusiasts feel about sharing their smart speakers with secondary and incidental users in their smart homes.

3.2 Methodology

To answer the four research questions, we combined an online survey with smart speaker enthusiasts (N=102) and semi-structured interviews with 12 smart speaker enthusiasts. The online survey (Section 4) allowed us to gain an understanding of smart speaker enthusiasts' general experiences with and their usage of smart speakers. The majority of the survey questions revolved around six tasks that the respondents had to carry out with their own smart speaker. They were asked about their experiences with breakdowns and potential intelligibility issues, how they perceived their smart speakers during interactions, and their thoughts about sharing the device with members and non-members of their household.

The semi-structured interviews (Section 5), on the other hand, allowed us to gain deeper insights into the issues smart speaker enthusiasts faced, beyond what could be gathered from the online survey. While two interviews were conducted at participants' homes, the majority of interviews (10/12) were conducted remotely through video calls due to the large geographical distance. The interviewees shared information about themselves and their households, general smart speaker experience and usage, and they reflected on their level of confidence in using and understanding their device. Participants also shared how they understood their smart speaker alone and in relation to other IoT devices, and how they felt about sharing and using the smart speaker with others in the household (e.g. family and guests). We chose to conduct semi-structured interviews instead of on-site observations with participants because we were interested in participants' reflections on their overall experiences with their smart speaker, rather than specific instances of breakdowns that happened to occur during our site visit, if at all. The online survey and interview questions both cover **RQ2–4**. To answer **RQ1**, we only relied on the interviews due to the survey not giving us meaningful and sufficient data. For each of the six tasks in the survey, respondents were asked to rate their level of

understanding of what the smart speaker was doing on a 7-point Likert-scale, and optionally, to provide additional information to describe their lack of understanding. Most respondents provided little to no additional information about their understanding and the quantitative results showed no clear trends. The online survey and interview questions are available in supplementary material A and B respectively.

4 ONLINE SURVEY

4.1 Respondents

We sought out respondents through online communities dedicated to smart speakers (e.g. on Reddit and Facebook) as well as through snowball sampling, asking colleagues, students, and acquaintances. We chose this recruitment strategy as it allowed us to specifically target smart speaker enthusiasts. Initially, 119 respondents completed the survey; however, we excluded 17 participants: One due to uncertainty about the seriousness of their answers, another for failing to follow the instructions regarding anonymization, and 15 for using their smart speaker infrequently or owning it only for a short period of time. We base our analysis on the remaining 102 respondents.

We found that the majority (71/102) of respondents were between 24 and 42 years old and the majority (86/102) shared their household with either family, a partner, or another person. While the participants' native languages spread across 11 different languages, most survey respondents were native English speakers and used English with their smart speaker (87/102), of which most were US-based (61/87). Nearly all (99/102) of respondents configured their smart speakers to English, while the rest used German, Spanish, or French. In addition, almost all (90/102) used English as their daily spoken language. The respondents' smart speakers ranged over 15 unique models. Some participants had several smart speakers; one participant even owned nine. Nearly half of the respondents (50/102) owned Google Home products exclusively, 16 respondents owned Apple HomePods exclusively, 13 Amazon Echo products exclusively, and 20 reported a mixture of different products. In terms of the rarer types, one respondent reported having a Harmon Kardon Invoke, another had an Insigna NS-CSPGASP-B. One respondent did not specify the type.

4.2 Questions and Tasks

The survey consisted of 143 items, which are a mixture of questions, rating statements (7-point Likert Scale), and open-ended text fields to elaborate their answers (62/143 items were optional open-ended text fields). The full survey is available in supplementary material A. The median time for completing the online survey was 14.2 minutes and it ran for 97 days. The majority of the questions (114/143) revolved around six tasks that the respondents had to carry out with their smart speaker; the use of tasks was inspired by Cowan et al.'s approach [19]. Our six tasks were as follows:

1. How will the weather be in your location today?
2. Find an Italian restaurant within 5 km. If possible, ask a follow-up question referring to one of the restaurants mentioned in the list of results.
3. Play "Thunderstruck" by AC/DC. When it starts playing, tell your smart speaker to play a song of your choice instead.
4. Set a 5-minute timer.
5. Translate 'Hello, how are you doing today?' into German.
   - Answer: Hallo, wie geht es dir heute?
6. Who is the Prime Minister of United Kingdom and where is she born?
   - Answer: Theresa May, born in Eastbourne, Sussex.

For each task, the respondents were asked the same set of 19 questions to evaluate their overall experience and how they experienced their smart speaker's responses. Respondents rated their experiences on a 7-point Likert scale with an option to elaborate. We made tasks 2, 5, and 6 slightly more difficult than the rest to see whether respondents would report having difficulties completing some tasks and why. Task 2 consisted of two requests (one asking about the previous one), task 5 included a different language, and task 6 was a two-step question. Note that the survey tasks did not include instructions on controlling other IoT devices due to uncertainty about how many respondents would use other IoT devices, and which types of devices they would own. Use of smart speakers with IoT devices was covered in the semi-structured interviews. We also asked the survey respondents to rate survey items **I1**–**I5** (see Figure 1) on a 7-point Likert Scale, which reflects their level of confidence in using and understanding their device, whether they faced it during interactions, and to which extent they liked using smart speakers with others.

4.3 Analysis

The quantitative data was analyzed using descriptive statistics. One of the authors analyzed the qualitative data in the survey using comparative analysis, open coding, and conceptual saturation [18]. Subsequently both authors discussed using an *interpretivist semi-structured approach* [8], which is commonly applied in HCI research. The findings were then grouped according to the research questions (Section 3.1).

4.4 Results

In the following sections, we discuss the findings from the online survey along the following main themes: respondents and their smart speakers (4.4.1); how respondents face and address their smart speakers (4.4.2); the smart speakers' responses (4.4.3); and multi-user experiences with smart speakers (4.4.4). Section 4.4.1 covers general findings about the survey participants and their smart speakers, while the following sections cover **RQ2**–**4**, as indicated in the section titles.

4.4.1 Respondents and Their Smart Speakers. The majority of respondents (95/102) reported using their smart speaker daily while the rest reported using it 3–5 times per week. The respondents had owned their smart speaker(s) between 3 months and 4 years – the majority (64/102) owned their device for a year or more. We looked into usage domains, and we were particularly interested in home automation, which was the most common group (68/102) (together with music), and consisted of, e.g. controlling lights, thermostats, smart plugs, and unspecified "*home automation*." This is in line with Bentley et al. [7] who also found home automation among the top domains.

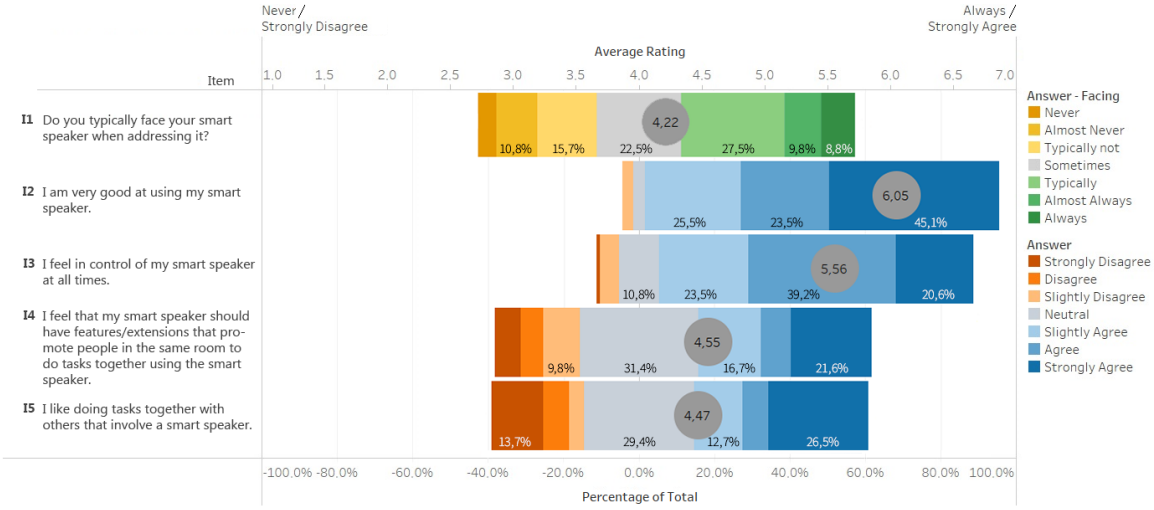

Figure 1: Survey items (I1–I5) and results. Top axis represents average Likert scale rating while bottom axis represents the percentage of negative and positive ratings of the total number of participants; 'Neutral' and 'Sometimes' equally distributed on both sides of the scales. The rectangles represent percentages while circles represent averages.

4.4.2 RQ2-3: Facing and Addressing the Smart Speaker. Respondents rated the statement on whether they faced their smart speakers during interactions (**I1**, see Figure 1) close to neutral with an average of 4.22 (with 1 = Never, 7 = Always), indicating that they sometimes face their smart speakers. This is corroborated in two optional follow-up questions where 85/102 of respondents gave examples of when they would face their smart speaker, and 83/102 gave examples of when they would not face their smart speaker. Respondents frequently (55/85) reported that proximity to their device, having the device already in their view, high background noise/music, breakdowns, and locations in different rooms caused them to face their smart speaker. Of those 55, one respondent who always faced their smart speaker during interactions explained that, "[. . .] *it doesn't understand me otherwise, even if I'm right next to it.*" However, respondents predominantly (60/83) reported that they would not face their smart speaker when they were sitting in the same room as the device or were located in a different room with the device out of sight. One respondent who never faced their smart speaker said: "*If I fail to get a response at my first or second try, I give up due to the nature of asking the same question of an inanimate object coming across as strange, even to family.*" Despite the breakdowns, this respondent did not mention moments in which they would face their smart speaker, as was common with other respondents. Other reasons for not facing the smart speaker included other activities (e.g. household tasks or conversing), and a general preference to not face the device due to confidence in it being able to understand the respondents' requests.

The above shows that addressing smart speakers (**RQ2**) depends on the context in which the device and person find themselves. The data also suggests that some people's first instinct if the smart speaker does not behave as expected, i.e. a breakdown (**RQ3**), is to face or walk up to the device, or, use an available app on their phone to control appliances (e.g. smart light bulbs) or stream music to the speaker.

4.4.3 RQ3: Response from Smart Speakers. The tasks that the participants completed with their smart speakers revealed interesting breakdowns; especially tasks 2, 5, and 6 due to their inherent complexity. The respondents' answers showed how the same smart speakers replied with different error messages. For instance, 26 respondents had to split the question from task 6 into two, and 8 HomePod owners mentioned that their device replied, *"I can't find the answer to that on HomePod"* while splitting the question gave them the right answer. One HomePod requested the owner to use their iPhone. It is unclear to what extent such uninformative responses mislabel available features as unavailable due to smart speakers not providing sufficient information to recover. In most of the difficult tasks, respondents managed to complete the tasks using between one and four tries while some (21/102, of which 18 native English speakers) reported unsuccessful completions in tasks 2, 5, and/or 6. Notably, in task 2, a HomePod owner reported that their device replied "*I didn't find any matching restaurants.*", while performing a Google search on another device showed "[...] *there's one 3.8km from me. It's called* [Italian name] *with 'northern Italian cuisine' as the tagline, so no excuse to* [having] *missed that.*" Some Amazon Echo owners were unable to find restaurants due to unavailable features. One was asked to install a 'skill' (a plugin that enables Alexa to perform requests), while others could complete the task.

Interestingly, there are examples of respondents who feel confident about their smart speaker usage and yet experience breakdowns beyond their comprehension. One respondent slightly agreed to being very good at using the smart speaker (**I2**) and slightly agreed to being in control (**I3**) while also noting that "*Commands suddenly don't work anymore or give strange outputs.*" Similarly, another respondent who slightly agreed with both **I2** and **I3** said that, "*Sometimes I wonder if I just don't know how to word the question or command in a way that the speaker would understand.*"

4.4.4 RQ4: Multi-User Experience. We asked the respondents to share their experience of using smart speakers with other people (**RQ4**). 72/102 respondents reported using their smart speakers together with others. One respondent reported, "*I would love it if it were more intuitive to use devices like this. Everything seems geared toward people who are more gadget-oriented or willing to research things they can do.*" This might be why responses such as "*An intercom feature would be helpful in this scenario*" occur from a U.S. Google Home owner, even though that feature has been available since November 2017 [75]. This respondent had owned a smart speaker for 12–23 months and used the device daily with a partner and had rated **I2** and **I3** at 7 and 6 respectively. This shows that even experienced and confident enthusiasts experience discoverability problems as is common with VUIs [30, 74].

53/72 of the respondents used their smart speaker with others by playing music at gatherings, entertaining children with music, video chatting through their smart speaker displays (e.g. Amazon Show) among other activities. And 13 out of those 53 reported non-collaborative activities as well. These were things like family calendar updates, shopping lists, adjusting thermostats, or as a respondent said, "*I sometimes set reminders for my husband.*"

We analyzed the respondents' open-ended answers to **I4**-**I5** since the neutral average ratings made the Likert scale difficult to assess. Those who disagreed and strongly disagreed with **I4**, did not elaborate their answer, while two who slightly disagreed mentioned that they would have difficulties imagining how features and extensions to the smart speaker would work with respect to the device encouraging multi-user interactions. One participant who slightly agreed said, "[. . .] *asking my smart speaker for facts feels more like part of a conversation, while pulling out my phone seems rude in the midst of a discussion.*" Finally, a respondent who strongly agreed with **I4** thought, "*It would be fantastic to include some type of presence detection based off of Wi-Fi that greeted new users and ask*[ed] *them their name to learn any preferences they have.*" These examples show how a more inclusive smart speaker could work.

Regarding the preferences of using smart speakers with others, the optional open-ended question: *In which context would you do tasks with others?* showed that even those that (slightly) disagreed with **I5** had moments in which they would use their smart speakers with others. Those that were neutral had a mixture of experiences such as "*when we are cooking together, one of us will set a timer, another one will check on or clear it.*", to "*generally, we don't use Alexa 'together'*". Meanwhile, one who strongly disagreed with **I5** said, "*I find that people who don't use Alexa (or similar) often get confused when speaking to the smart speaker, which means it takes much more time to perform tasks, so I prefer to not use it with others.*"

The above findings shed light on **RQ4**, showing that owners of smart speakers who are more acquainted with the IPA do not always desire asking new users to partake in unclear interactions through a VUI. On the other hand, simpler and playful commands such as turning on kitchen lights or playing a game seem more common.

5 INTERVIEWS

5.1 Participants

For the interviews, we again recruited smart speaker enthusiasts who owned and used smart speakers regularly. We announced a different call for interviews through the same online communities as for the survey, supplemented with local advertising at our university to widen our range of participants. 14 people responded to our call for interviews, one was rejected due to only using their smart speaker sporadically, while another one was excluded due to audio issues. Of the remaining 12 interviews, two were conducted at the participants' homes and ten via video calls due to the large geographical distance. Interviews lasted on average 75 minutes. Three interviewees also participated in the survey. The majority (7/12) of participants were between 24 and 42 years old. While none of the interviewees were native English speakers, the interview was conducted in English (which all interviewees were comfortable with) and most interviewees used English with their smart speakers (9/12), while three used it in their native language (two in German and one in French). 7/12 participants owned Google Home products, three owned Amazon Echo products, and two an Apple HomePod. Interviewees reported four types of households: 7/12 lived as a family, two lived alone, two shared their home with roommates, and one was living with their romantic partner. While we did not interview other household members of the interviewees, discussions in the interviews also reflected on these other household members' experiences. Participants' duration of owning a smart speaker varied from 2 months to 42 months. Note that the participant who owned their smart speakers for 2 months already had an initial IoT setup and plans for expansion. All participants used their smart speaker(s) daily. Four considered themselves smart speaker developers (working on features for smart speakers or home automation), four worked in the IT industry, while four did not specify.

5.2 Procedure

First, we introduced the participants to the topic of the study, followed by a short survey to collect initial data (e.g. on language use, types of smart speakers, household information, confidence in using and sense of being in control of their smart speaker). We divided the rest of the interview into two themes: understanding the smart speaker and multi-user scenarios. The full interview procedure is available in supplementary material B. The participants were furthermore encouraged to share anything else they found relevant to the topic being discussed. The interviews were audio recorded and transcribed. The two in-person interviews were also video recorded.

We also asked the participants a leading question: *Do you sometimes perceive your smart speaker as a 'black box'?* We asked this because Porcheron et al. [56] suggested that the Amazon Echo's lack of transparency regarding its state, was akin to a 'black box' system. Leading questions are commonly used to verify and check the reliability of interviewees' answers, and interviewer's interpretation [10]. We then followed up by asking the participants to explain how the smart speaker works, to get a nuanced understanding. We also ensured that the participants knew what we meant by 'black box'.

5.3 Analysis

As in the survey, the interview transcripts were analyzed using comparative analysis, open coding, and conceptual saturation [18]. Moreover, one author coded the data to familiarize himself with the data and then further discussed together with the other author to develop themes as in an *interpretivist semi-structured approach* [8].

5.4 Results

In the following sections, we discuss the results from the interviews along the following main themes: confidence in using the smart speaker(s) (5.4.1); understanding of the IPA's behaviour (5.4.2); trust in the smart speaker (5.4.3); addressing the smart speaker and the use of feedback (5.4.4); strategies to recover from mistakes and breakdowns (5.4.5); and use of smart speakers with multiple people (5.4.6).

Lastly, to distinguish the participants who considered themselves to be smart speaker developers, participants will be referred to as **P1**–**P12** with the suffix **D** (developers) or suffix **N** (non-developers): e.g. **PN1**/**PD7**.

5.4.1 Confidence in Using and Feeling in Control. On average, participants rated themselves as relatively good at using their smart speaker (5.25 on a 7-point Likert Scale) while feeling slightly in control of their device (4.58 on a 7-point Likert Scale). These numbers are in line with the results of the online survey, shown in Figure 1.

5.4.2 RQ1: Participants' Understanding of IPAs in Smart Speakers. Five participants said that they perceived their smart speaker as a 'black box'. **PD7**, whose occupation was within smart home automation, perceived their Google devices as a 'black box' *to a certain extent*. **PD7** found it difficult to identify the locus of an error, yet, they said, "[Google] *does a good job of describing exactly what happened and why it did* [what it did]." **PD7** acknowledged this contradiction while stressing that they would not prefer the IPA to explain how things worked. This is in contrast to **PN1** and **PN11** who perceived their Google Home and HomePod respectively to not be a 'black box' due to the lack of artificial intelligence. **PN1** said that they believed their Google Home consisted of preprogrammed if-then statements, similar to a trigger-action approach [65]. Conversely, **PN11** did believe that the Google Home was more intelligent than their HomePod since it used the collected data to improve its services. Contrary to our expectations, that enthusiasts who had a less complete understanding of their smart speaker would be more likely to identify the device as a 'black box', they seemed to fill in the gaps in their understanding with what they thought must be true. The interviews also showed, unsurprisingly, a contrast between the developer participants (4/12) and the non-developers, in terms of describing how a smart speaker worked. The four developers had a much clearer model of the smart speaker's process, while the other eight had a vague idea that the device connected to 'cloud' servers and processed requests there. **PD12** said the following:

> *"I basically know what's in there, like you have the hot word at the beginning, after that you have ASR [Automatic Speech Recognition] to recognize what you say, and after you have the NLU [Natural Language Understanding], that understands what you say. And the request is sent and is done."*

On the other hand, **PN3** used available information about the smart speaker's interpretation as a cue for how the device worked and explained the following:

> *"I guess it records and translate[s] my voice and translates it into something written and then it reads that and tries to answer the best with sort of machine learning or something like that."*

Participants with little to no technical background had little knowledge about the way their devices communicated with other smart home devices. Six participants (of which one developer) indicated that the current mobile applications were missing clearer information about how the users' various smart devices were interconnected, while 3/6 participants suggested that some type of visualization would be great. **PN4** came up with the idea that their smart speaker could dim light bulbs to indicate the light bulb's connectivity while Bluetooth speakers could make sounds. In contrast, **PD10** mentioned that they were working on their own desktop view of the different connections:

> *"I think it would be helpful if you could see more than just these devices connected. It would be helpful if it would give me options there as well. So, sort of like a discovery feature for things you can do with your own stuff."*

On the other hand, five participants (of which three developers) found it mostly sufficient to use the available app for the smart speaker to check connections between devices, while two of them said they were open to further improvements. It is also important to emphasize that five participants (both with IT and non-IT backgrounds) mentioned that the increased number of interconnected devices with the smart speaker raised the level of complexity making it difficult to keep an overview of commands and devices. On a similar note, **PD7** said:

> *"[...] it's incredibly unclear precisely how often [smart speakers] are being updated, if they work at all, who's responsible for them working. [...] from a [...] developer perspective, I understand that Philips is contacting Google APIs... and I know how technically this all works together. Once it stops working it's incredibly frustrating that you don't really know who to talk to... to get things to work..."*

Despite **PD7**'s better understanding (**RQ1**), this shows how difficult it can be to identify where mistakes happen if the speaker is connected to other devices and services.

5.4.3 RQ1: Trusting the Smart Speaker. We examine trust both from a perspective of privacy as has been investigated by, e.g., Lau et al.[32], and Bellotti et al.'s *alignment* (How do I know the system is doing the right thing?) [4]. This contributes to **RQ1** as the depth of enthusiasts' understanding of the smart speaker's behaviour and its impact to what extent it is trusted to handle potentially harmful actions. We asked the following question: *Do you trust your smart speaker?* From a privacy perspective, only **PD12** was unsure about how to answer this question while all other participants trusted their devices despite acknowledging their lack of knowledge

about what happens to their data. **PD12** designs smart speakers and explained that they only feel comfortable having their HomePod in their living room. Our participants seemed to trust the speaker companies with relatively few privacy concerns, which is in line with findings by Lau et al. [32]. Lau et al. suggest that it might be due to an incomplete understanding of privacy risks, or, as we observed with **PN2** and **PN4**: convenience outweighed privacy concerns, for instance, over the device listening all the time.

From an *alignment* perspective [4], five of the participants (all non-developers) mentioned that they did not feel confident enough about the smart speaker handling potentially harmful things, due to the technology's early stage of development – yet, three developers suggested two-factor authentication as a possible solution. Only in rare occasions would three participants consider turning off their devices. For instance, **PN2** mentioned that they would unplug the device in case they and their family rented their house to others while they were gone to avoid any uncomfortable experiences for the guests. **PN4** recalled an experience when the lights in their apartment, connected to home automation, were turned on for a whole week while they were on vacation. While notified on their phone, they were unable to switch the lights off, and referred to the danger of connecting smart speakers to appliances such as an oven, which can become dangerous if unattended. This shows that some enthusiasts are cautious of the novelty of the technology and the device's unpredictability, which relates to *alignment* [4], i.e. not doing something unwanted when the user is not present, or not being triggered accidentally.

5.4.4 RQ2: Addressing the Smart Speaker and Feedback. All 12 participants preferred not to face their smart speakers due to the nature of issuing voice commands and the provided audible feedback from smart speakers. The audible feedback the device provides when it triggers, frees up the user's visual attention. However, **PN2** and **PN5** both have two small children and have experienced that they interact with the smart speakers playfully by walking up to the devices and speaking *to* the devices. In addition to these two participants, some of the other participants explained that when addressing the smart speaker, it is normal to look at the device for the first period of owning it. Children have been observed facing and interacting playfully with smart speakers by moving their hands in front of the devices and touching them [22]. In contrast, some adults place their Amazon Echo out of direct view [60]. Hence, we investigated if physical interaction was preferred over voice sometimes. Three participants mentioned that their choice of input mechanisms depended on the situation. **PN11** mentioned that they usually trigger their HomePod by tapping on it when they lie in bed and know that other family members are asleep. **PN6** said, similarly, that their Google Home is within hand's reach on their desk and they find it more convenient to physically control the *trigger*, *stop*, *pause*/*resume*, and *volume up*/*down* commands. The context thus determines whether physical controls are used.

The participants occasionally used their companion apps on their phones to check what the smart speaker understood. Still, they deemed the physicality of the device to be important for them to have a point of reference in case of breakdowns that go beyond rephrasing requests. As **PN5** pointed out, "*Always turn off, turn on. That's rule number 1 within IT*" referring to the importance of being able to go back to the smart speaker and restart it. This refers to *accidents* [4], where knowing how to recover from breakdowns is important for users, through familiar interfaces that allow users to *address* [4] the system correctly. The majority of the participants also acknowledged that breakdowns do make them face their device, depending on the number of failed requests. **PN4** noted:

> *"Actually, when it's doing something wrong [. . .]. So, there I address it directly and I go to it or come near to it and then I repeat it slowly and loud again. So, that is kind of interesting because I could just also do it by my smartphone or something like that [. . .]"*

This echoes responses from the survey. Building on this close communication pattern, **PD7** who thought that the reason they were being conversational with the IPA on their smart speaker and not their IPA on their phone was exactly because "*smart speakers respond in a very personal sort of way, has a very personal sort of voice*". **PD7** further explained that the smart speaker does not tether them to a device like the phone does, allowing them to move freely about, while also pointing out that they find the IPA on the phone intrusive and unpleasant to interact with. This suggests that the type of voice and even the device through which the IPA is invoked can influence anthropomorphic characteristics.

PN4, **PN8**, and **PD12** mentioned that they would walk up to the smart speaker at times to improve its interpretation if the device would not execute the right action. Notably, **PN1**, who is a non-native English speaker, expressed that changing the temperature in the kitchen with Google Home works well "[...] *but only when I am facing the assistant and speaking loudly and clearly* [...]". In addition, many participants mentioned that if there was too much background noise, the smart speaker would have difficulties picking up trigger words and would require participants to either shout or get closer to the device. **PN5** had an interesting experience when they were cooking in the kitchen:

> *"[...] if the volume is turned down to [. . .] 10% in the kitchen and it talks to me and I have the exhaust hood turned on, then I can't hear the speaker. [. . .] Then I just turned around and [. . .] I said 'Hey Google, set volume to 80%'... ok, and I said, 'Please repeat command', because I wanted to know what it said, and it said 'Last command is, turn volume to 80%' [. . .]."*

This example shows how issues with *attention* [4] highlight problems such as unintended actions and wasted input efforts due to the system's lack of contextual awareness.

The above scenarios show how facing the smart speaker can depend on breakdowns (**RQ2**). This is in line with the online survey responses as they also suggested that addressing and recovering are intertwined.

5.4.5 RQ3: Recovering from Mistakes and Breakdowns. Inspired by Porcheron et al.'s [56] findings that showed how first-time users were unable to act upon default responses, we asked if there was any immediate action that enthusiasts would take to intervene in a mistake. Nine participants agreed that the current 'stop' command was sufficient in most cases. Yet, a few mentioned that a more effective way of intervening would be preferred due to sometimes having difficulties triggering the device, while **PN1**, **PN4** and **PD7** pointed

out the potential benefit of smart speakers giving options to users, where the device would ask them something along the lines of "*Did you mean...?*". When asked, all 12 participants agreed that alternatives would be a great feature though in moderation or as needed, since spoken lists of alternatives can become overwhelming.

Furthermore, **PN4** mentioned that smart speakers need to be able to handle different tasks appropriately with different procedures, which corresponds with Mennicken et al.'s suggestion of making voice assistants domain-specific [42]. **PN5** suggested, when they say "*'Hey Google, turn off light in kitchen' and what I actually meant was the living room* [...]" that the smart speaker would then reply, "*The light in the kitchen is already turned off, do you want me to turn it off in the living room?*" **PN6** viewed alternative interpretations as guidance on how to utter a command, *"If I ask a command* [that] *I don't know the exact phraseology or syntax of, then I would think it's very useful, I would definitely use this."* This suggests that alternatives do not necessarily have to relate to the interpreted input, but can also reveal available phrases. This could help to address discoverability issues [30, 74].

As for discoverability and Bellotti et al.'s *action* design consideration (i.e. how do users discover and select an available command) [4], **PN1**, **PN2**, and **PN5** said that they sometimes forget the exact phrasing of a request required by the smart speaker. During the interview with **PN1**, they attempted nine times to change the temperature in their home with a voice command, however, their first three attempts did not work. Before picking up their tablet to look up the "If-This-Then-That" command that they specified on IFTTT.com, they said, "*The question is, did I forget the right command or didn't it understand me?*" Even after finding the phrases, they succeeded only the sixth time; it took them nearly 4.5 minutes to succeed. On the other hand, **PN2** had a structured approach:

> *"[. . .] I will go back through the hierarchy, like first look in the Alexa app, then look in the Sonos app or the Hue app [be]cause the data from the Alexa app comes from there. And then if that doesn't work, I have to look into the Sonos status reports"*

PN5 pointed out that they had long-term memory issues and found themselves occasionally forgetting exact phrases to use for adjusting lights in the different rooms. This shows how VUIs often do not adhere to common UI guidelines (e.g. reducing short-term memory load [45, 50, 61]), and reflects the importance of designing for dynamic diversity and universal usability [26, 61], given that users with a range of disabilities are using smart speakers [9, 57].

In summary, through our interpretivist approach, seven strategies emerged that describe how our participants address breakdowns (**RQ3**) (see Table 1). The individual strategies are used in various contexts and are sometimes used sequentially as pointed out by **PD12**:

"[. . .] *when* [HomePod] *doesn't trigger at first,* [. . .] *I turn around to see that.*" and *"When it's maybe really not picking up like sort of fourth time",* **PD12** confirmed that they would walk up to the smart speaker. This raises open questions, such as how much do users tolerate until they give up or move to a different product altogether. And as pointed out by **PD7:** "[. . .] *if you check the Google Home Reddit* [. . .] *you see the exact same posts over and over again. People who have lights that don't connect, or people who have a bulb that specifically doesn't* [want to] *update.*" and continues to emphasize the frustration that people get when they "[. . .] *get a device and want it to work immediately and perfectly, and you don't really give a device multiple chances, especially not a device that you need daily.*" This situation shows how difficult it can be to establish an IoT ecosystem in one's home without the help of the community of enthusiasts.

S1 and **S2** reflect that users indeed make use of common human-human communication practices to improve their directional interaction with the smart speaker, answering **RQ2** about facing and approaching the smart speaker. **S3** is among the common strategies used in overcoming NLP errors [38], which were also regularly experienced by our interviewees. This makes NLP error recovery strategies such as hyperarticulation, simplification, and adding more information to utterances the first type of strategies to be used [47]. However, they are of little use in overcoming infrastructural issues. **S4**, **S5** and **S7** reflect the limitations of current smart speakers' capabilities, where enthusiasts tend to exhibit interest in investigating such infrastructural breakdowns by looking for solutions that require external sources of information about how to overcome a particular breakdown with their smart speaker, as the smart speaker does not provide sufficient information. Finally, **S6** corresponds to voice interaction not always being ideal, hence the preferred physical interaction with smart speakers in some instances.

5.4.6 RQ4: Multi-User Scenarios. To answer **RQ4**: we set out to get a sense of who was responsible for the smart speaker(s) in the households and who felt ownership of the smart speaker(s). **PN1** said that setting up specific commands outside the smart speakers' built-in commands requires them as the primary user to instruct and update others in the household. Similarly, **PN2**, **3**, **5**, **6**, and **PD10** also indicated that they were the ones who introduced the smart speakers' built-in features to their family. While **PD7** and **PD9** did not live with family members, **PD7** shared their flat with a roommate who had to deal with the smart speaker regardless of their opinion about it, and **PD9** explained how they introduced smart speakers to their family (e.g. (grand)parents) in their homes. This role is in line with Mennicken and Huang's [43] observations about '*home technology drivers*' in smart homes, people who research, acquire, and implement home automation technology in their spare time. Mennicken and Huang also pointed out that most of the adult members of the households in their study fell into the '*passive users*' group: people who did not directly engage in planning, research, maintenance, and configuration, but did have some familiarity with the home automation system through use. This fits our observations of most interviewees describing their cohabitants and guests in line with passive users.

Participants were generally open to guests using the smart speakers for sources of information, entertainment, and controlling other smart devices. The smart speaker was considered a common device available for everyone to use, due to having limited or no personal information on the smart speakers. **PN4** pointed out how their family viewed this as *"Just one of dad's fun ideas"* (referring to themselves) and *"the wife thinks it's almost annoying, so far. But she has begun lately to give it commands."* **PN4**'s wife started using the countdown feature for when the children brush their teeth and

Table 1: Seven strategies enthusiasts used to recover from breakdowns.

Strategy	Examples	Quotes from participants
S1: Turn towards the smart speaker.	**PD12** faces the smart speaker if it does not trigger.	"But when [HomePod] doesn't trigger at first, [. . .] I turn around to see that." – **PD12**
S2: Walk up to the smart speaker.	**PN4** walks up to the smart speaker to enunciate the utterance.	"[. . .] when it's doing something wrong, I go to it and then I face it quite often. So, there I address it directly and I go to it or come near to it, and then I repeat it slowly and loud again. So, what is kind of interesting because I could just also do it by my smartphone or something like that but I really go to the speakers and face it [. . .]" – **PN4**
S3: Retry request a number of times.	**PN1** tried nine times in a row to trigger an IFTTT command until they succeeded in changing the temperature in the kitchen.	"And sometimes, it's hard to remember precisely the sentence. That's the problem with if-this-then-that. Just [have to] think a moment. . . Because I haven't used it for many months because it was so hot. . ." "heat kitchen to the twen. . . [. . .] heat kitchen thirty. . . [. . .] heat kitchen 23 degrees [. . .] heat kitchen 23. . . [. . .] heat kitchen 23 degrees [. . .] kitchen 23. . . [. . .] kitchen. . . [. . .] kitchen 22 degrees [. . .] kitchen radio. . . [. . .] kitchen 22 degrees" – **PN1**
S4: Investigate the issue on a smart-phone or tablet.	**PN2** goes through a number of apps on their phone.	"I will ask [Alexa] to do a certain thing. And when it [doesn't] work I [will] then have to [. . .] debug it somehow [...] go back through the hierarchy. Like first look in the Alexa app, then look in the Sonos app or the Hue app [be]cause the data from the Alexa app comes from there. And then if that doesn't work, I have to look into the Sonos status reports" – **PN2**
S5: Complete task through companion app instead.	**PD12** uses their phone to stream music if the smart speaker fails to play music. **PN8** is more inclined to give up on the speaker if it fails to play music than if it fails to create a timer, as it is easier to select music on the phone.	"[. . .] some music that I want to play, sometimes [HomePod] doesn't recognize or I have to go to my phone and... look for it and stream it." – **PD12** "If I am setting a timer for example, I might try more times than if I am trying to play some music. In that case, I give up more easily and play the music through the app on my phone instead." – **PN8**
S6: Physically interact with the smart speaker.	**PN1** uses the smart speaker's physical buttons to lower the music volume.	"[. . .] when we have guests as yesterday, it's pretty loud, many people are speaking. Then it's difficult to communicate with the [smart speakers]. So, often you have to turn up and down the music. . . the volume, physically." – **PN1**
S7: Get help from the online community.	**PN2** checks online forums where other enthusiasts share their experiences.	"I always go online and check if there [are] some skills that some of the other guys are using and then perhaps try them if they make sense." – **PN2**

when the family is cooking, as well as asking the smart speaker trivia questions. Yet, participants pointed out that some passive users (usually guests) showed a reluctance in using the smart speakers. Meanwhile, **PN11** was worried that someone, such as a sibling or guest, would ask the HomePod to read aloud personal messages linked to their smartphones, while **PD12** did not connect their phone's messages to the smart speaker. Despite the voice recognition functionality, **PD9** pointed out that siblings with similar voices can trigger each other's calendar. Regarding *alignment* [4], users have the option to check their activity log in the mobile application and listen through all the latest activities to verify that nobody accidentally or intentionally accessed their personal information (e.g. calendar). Yet, none of our participants mentioned doing so every day nor after having guests over.

6 SUMMARY

RQ1 – How do enthusiasts conceptualize their smart speaker's behaviour? The data suggests that enthusiasts vary in their interpretation of how their smart speaker works depending on their experience and background. As discussed previously, less tech-savvy enthusiasts without a technical background rationalize the smart speaker's behaviour as primitive with no AI, while the enthusiasts with some technical background within the IT industry showed some understanding of machine learning. Finally, the developers provided a relatively complete description of how smart speakers work.

RQ2 – How do enthusiasts address their smart speaker? Do they approach it and face it, and what other modalities do they use besides speech? From our interviews, it was clear that speech and hands-free interaction with smart speakers was preferred due to the natural language interface. However, our data has shown that majority of our participants experience the need to face their smart speakers in times of breakdowns and in some cases even felt forced to walk up the smart speaker to ensure it understood them. While physical interaction with the smart speakers was reported to be infrequent, in some situations a quick 'tap' on the smart speaker allowed for a faster interaction and was thus occasionally preferred over speech. This is indicative of a possible preference for a multimodal interface over a strictly unimodal interface.

RQ3 – What strategies do enthusiasts employ to recover from mistakes and system breakdowns? Table 1 shows seven strategies used by our enthusiast participants during smart speaker breakdowns. It is important to point out that despite the developers' stronger grasp of the smart speaker's inner workings, they still experienced uncertainty regarding breakdowns in the context of

IoT ecosystems. Our results suggest that there is a lack of transparency in a home's IoT ecosystem, even for users with advanced knowledge on smart speakers, and that enthusiasts draw a lot on online communities for help.

RQ4 – How do enthusiasts use their smart speaker with others in their households and/or when having visitors? Our participants made it very clear that smart speakers are part of the home, especially if they can control other smart home appliances. They should be accessible to others, both household members and people visiting their home, making the smart speaker a shared technology. However, enthusiasts mentioned that they bear the brunt of this burden, becoming responsible for maintaining the devices, informing passive users about the changes as well as helping them during breakdowns.

Our findings extend prior work by highlighting challenges that enthusiasts encounter in their interactions with smart speakers within IoT ecosystems, beyond conversational challenges with smart speakers alone. Importantly, our results also suggest that technically skilled users and developers who likely have a better understanding of smart speakers and smart home setups, like some of our interviewees, still experience difficulties locating and overcoming infrastructural errors. This suggests that there is a need for smart speakers to be more intelligible and support users in overcoming such issues regardless of the user's experience or background, in particular when smart speakers are interacting with a larger IoT infrastructure. Our interviews show that our participants often ended up in a position in which they become the leading experts in their household, as they tend to be the primary users in charge of the smart home transformation in their home. This might become an extra burden on the primary users, as they are the responsible ones who have to maintain the system in place.

7 DISCUSSION

From our findings, we synthesized three discussion points for future research that should be considered given that smart speakers are increasingly pervasive in people's homes.

7.1 Infrastructural Breakdowns as Learning Opportunities

In addition to conversational breakdowns, a common source of errors that our interviewees experienced are infrastructural breakdowns due to the speaker being part of a larger ecosystem of services and appliances. It is hard to recover from infrastructural breakdowns due to issues with infrastructural intelligibility [23]: difficulties in understanding how the individual IoT devices and services work together to form a large ecosystem. This is even experienced by enthusiasts who have a strong grasp of how smart speakers work conceptually such as several of the developers among our participants. As we will explain below, we suggest that smart speakers could help users in identifying these infrastructural breakdowns and suggest possible ways to overcome them.

While prior work has suggested to use smart speakers' responses as a medium to help recover from communication breakdowns [6, 56], our interviewees pointed out the limitations of voice interaction to overcome breakdowns. For example, they had concerns about having to listen to long lists of options through an IPA's voice (Section 5.4.5). Currently, neither the visual nor the auditory modality is ideal to provide information to recover from infrastructural breakdowns: few smart speakers have displays, speakers are often hidden from direct view [60], and speech can be overwhelming [45]. Yet, visually impaired users may prefer information delivery via high-pace speech [9]. For all users to learn from breakdowns, a "one-size fits all" approach to intelligibility is not desirable.

Users tend to prefer comprehensible and practical information that helps and improves their daily interactions with intelligent interactive systems such as recommendation algorithms [11]. This is reflected in **PD10's** attempt to make a personal overview of the connectivity between their IoT devices and the devices capabilities. The streamlined, minimalistic and seamless design of smart speakers might in fact make the technology less accessible and intelligible. Despite this, our participants did not portray their smart speakers entirely as black boxes nor did they suggest that it was useful to think about it in such a way. Instead, they portrayed their smart speakers as less capable during infrastructural breakdowns and felt that the limited information about the respective breakdowns could be improved to enable users to act appropriately.

This discussion emphasizes the value of designing technology through a seamful [13] and upfront approach about the issues that arise regarding the smart speaker's internal behaviour. We do not argue that all technical errors need to be presented to the user's immediate attention, however, we do argue that those choices should be the user's decision to make. The aim is to not only reveal imperfections in the interactions between user and smart speaker, but also to enable users to identify which issues are related to the larger IoT infrastructure [23], and which of those are within and beyond their control.

One way of designing seamful smart speakers would be to involve the community in the process of learning from infrastructural breakdowns, since enthusiasts already seek out help on online forums about the immediate issues they face with their smart speakers. Smart speakers could leverage knowledge from online communities or a new dedicated platform where smart speaker enthusiasts share their experiences, solutions, and challenges, which the smart speaker could present to users. This could make smart speakers and users collaborators in overcoming infrastructural breakdowns in their homes. Similar work has been done in the context of providing community assistance with command recommendations in novice-to-expert transitions in complex software applications [24, 40]. Trigger-action commands [65] through services such as If-This-Then-That (IFTTT.com) are already a common way to take advantage of "recipes" shared by other people in the smart speaker community. In addition to sharing such rules or commands, we argue that the communities' expertise and experience with smart speakers in smart homes could be utilized as well, similar to how others have shown the potential of crowdsourcing contextual help for web applications [14].

Another possible solution would be to break the conversational metaphor, entering a diagnostics mode where user and smart speaker collaboratively investigate the infrastructural breakdown at hand, bringing forth the seams in the technologies within the smart home. This diagnostics mode could consist of questions and answers between the user and the smart speaker, allowing the device to narrow down the scope of possible issues and solutions

(e.g. [52]). Explanatory approaches have been researched extensively for complex software [46], context-aware systems [37] and end-user debugging of machine-learned systems [31]. In particular, explanations have been shown to improve users' understanding of semi-autonomous interactive behaviour [37]. This supports the idea that users of semi-autonomous interactive systems could learn about some of the underlying system behaviour through which (infrastructural) breakdowns are caused. In this case, smart speakers could be designed to become proactive in their level of engagement with users [12], offering explanations when a recent request failed.

Our findings point to an opportunity in investigating how and when smart speakers can expose underlying layers of their and interconnected IoT devices' system behaviour and simultaneously reveal possible actions to recover from infrastructural breakdowns.

7.2 Non-Verbal Communication as a Design Opportunity

Our findings point to our participants handling communication with their smart speakers in a similar way to how they would communicate with an actual person. While their overall understanding of how the smart speaker works varies depending on their technical expertise and experience with smart speakers, it was clear from our analysis that the majority of people tend to face their smart speakers *during* breakdowns. People act in a similar fashion during conversational breakdowns with collocutors to improve their communication, for example, by getting closer to the other person and establishing eye contact. We suspect that enthusiasts may adopt a similar subconscious understanding—anthropomorphizing their devices—due to the IPA's voice, as pointed out by **PD7**. Indeed, people attributing intent, social cues, and other anthropomorphic characteristics to technology has been observed previously to varying degrees [3, 29, 53, 63] and with smart speakers in particular [16, 22].

However, this metaphor of human-human communication also has downsides. While anthropomorphism can make use of metaphors and interactions familiar to people and thereby offer familiar action possibilities, anthropomorphism has also a tendency to mislead people and raise their expectations towards technology. Prior findings point towards this misalignment between users' expectations and the IPA's capabilities [19, 39]. While this might persist in first-time users' experiences, we observed that smart speaker enthusiasts adapt their expectations over time and accept the limitations of smart speakers' capabilities. More importantly though, our study highlights a possible subconscious behaviour from the enthusiasts (approaching the device or speaking up when experiencing a breakdown), which might change little over time, if at all.

Our results suggest that enthusiasts' behaviour around smart speakers, such as approaching or facing the device, could be leveraged as a signal that the user is attempting to recover from a breakdown. An interesting future direction is to explore whether this could be recognized by the smart speaker (e.g. through proxemic dimensions [25], or gaze as in Tama [41]) to offer incidental intelligibility [73] and reveal action possibilities to recover from these breakdowns. In addition, smart speakers could be designed with multimodal action possibilities, offering users more and different ways of communicating with IPAs in smart speakers, such as gesturing and physical interaction.

Building on this, future research could also investigate whether there are other user behaviours resulting from anthropomorphic attributions to smart speakers that could be leveraged for the design of smart speakers.

7.3 Intelligibility and Control Needs of Passive Users

Our results show passive users are common but can be left out of interaction with smart speakers or planning of smart home integrations (Section 5.4.6). Passive users may be uncomfortable interacting with smart speakers or unaware that they are there. While passive household members increase their frequency of use over time due to convenience and frequent exposure to the smart speaker, interviewees did not report the same tendency for guests. Overall, passive users have little means of knowing what they can do and how to recover from errors since not everyone uses smartphones nor is willing to install and configure companion apps, as observed by Lau et al. [32]. Given a future of smart homes with smart speakers as central interfaces for smart home appliances, it would be problematic if passive users had little to no means of controlling these appliances. Our interviewees, as primary smart speaker users, acted as system administrators that passive users have to rely on (in line with similar findings in smart homes [43]). With growing numbers of IoT appliances that can interface with smart speakers, this burden on primary users will only increase, since they already struggle in their own interactions with their smart speakers.

This highlights a need for control mechanisms for passive users. More care needs to be taken in the design of smart speakers to avoid unintended effects such as excluding particular users and a heavy reliance on primary users.

One way of designing smart speakers so that they are more accessible to passive users would be to include the devices in the introduction phase, making them key actors in "showing" users where smart appliances are located, what they can be used for, and how to control them. This will likely need to be set up and maintained by the primary user(s), however, this would allow primary users to delegate introductory tasks to smart speakers so the smart speakers themselves can teach passive users about (new) features and functionalities. Furthermore, we suggest to design this without relying on smartphones necessarily, as they need to be configured with smart speakers. Instead, we argue that for voice-enabled smart speakers to be more discoverable and accessible, the devices would benefit from a more diverse set of interaction channels through which users, in particular passive users, can input their requests and control some aspects of a smart home. While existing smart speakers with displays could utilize the graphical user interface [47], another interaction channel could be spatial interaction [27], where physical interaction exists within real space, which uses movement as input in the space as discussed in section 7.2. This has been seen with other technologies such as the Nest Thermostat [49], which comes with a mobile application, yet that application is not necessary for passive users. The users can control the Nest Thermostat by rotating it to the desired temperature setting while

also lighting up if the device detects the user nearby, indicating that they can interact with the device.

In summary, it is a promising direction to study the intelligibility and control needs of passive users and investigate how they can be included in future smart speaker experiences?

7.4 Limitations

Our interest in this study has been in developing an initial understanding of intelligibility issues faced by a specific group of smart speaker users – *smart speaker enthusiasts* – and how they recover from different types of breakdowns, in contrast with and complementing prior studies that have mostly focused on first-time users. Our methodological approach, in which we used an online survey and interviews, draws on respondents' overall experiences in using smart speakers. While this study might not provide details on specific observed instances of handling breakdowns with smart speakers, it complements existing work by contributing new insights into a different group of users and how their accumulated experience influences how they handle different breakdowns, in particular infrastructural intelligibility issues within IoT ecosystems.

In compliance with the EU's General Data Protection Regulation (GDPR) article 5(c), which states that data should be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')" [64], we only collected personal data that was deemed absolutely necessary for our study. In light of this and since our study did not focus on gender nor specific age differences, we did not collect participants' gender and specific age. While we acknowledge the significance of aspects such as gender, age, or socio-economic background on people's experiences with smart speakers [62, 70], an investigation of these aspects was deemed out of scope for this study.

Native English speakers may have a different experience than non-native English speakers with English as their only viable option to interact with their smart speaker (e.g. [72]). In this study, we had a mix of both. Most survey respondents were native English speakers, while the interviewees were only European smart speaker owners, due to constraints in the contract with our funding organization. As pointed out by prior research [72], non-native English speakers, in contrast to native speakers, might have trouble with respect to producing the right words for the IPA to understand. Having non-native English speakers use their smart speakers in English might have influenced our findings in relation to conversational breakdowns. However, it is unclear whether results from prior studies with native Mandarin speakers [72] would translate to native Germanic language speakers (as with our participants who used the smart speaker in English) as Germanic languages are closely related to English. Additionally, the recruitment of the participants from online fora may have biased the type of people we got, in particular those that were comfortable and frequent users of these fora, and likewise could have influenced the preference towards using online fora and the smart speaker community as a solution for issues with smart speakers.

Lastly, our notion of "passive users" consists of both members of the household and guests. Future work should look into the specific needs of people who indirectly interact with the smart speakers and people who only interact, but do not configure nor maintain the smart speakers. In addition, it would be equally important to know what the enthusiasts' view on sharing responsibility with respect to configuration and maintenance of smart speakers and homes, and to which extent enthusiasts would be willing to hand over more control to passive users.

8 CONCLUSION

Previous studies on smart speakers provide insights into how households integrate them into their lifestyles. We extend prior findings by: contributing insights into enthusiasts' understanding of their smart speakers; how they address the device; when they encounter unintelligible behavior, and strategies they use to recover from such breakdowns; and how these issues are handled in different multi-user settings. Based on our results, we propose three future research directions: considering infrastructural breakdowns as learning opportunities for understanding the smart speaker's behaviour; leveraging aspects of non-verbal communication as opportunities for design; and considering the intelligibility and control needs of passive users.

ACKNOWLEDGMENTS

We thank our participants for their contribution in making this work happen. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 740548). This project was approved by the Institutional Review Board at Aarhus University (serial number: 2019-01).

REFERENCES

[1] Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y. and Kankanhalli, M. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 582:1–582:18.

[2] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. 2019. Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 19.

[3] Anderson-Bashan, L., Megidish, B., Erel, H., Wald, I., Hoffman, G., Zuckerman, O. and Grishko, A. 2018. The Greeting Machine: An Abstract Robotic Object for Opening Encounters. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Aug. 2018), 595–602.

[4] Bellotti, V., Back, M., Edwards, W.K., Grinter, R.E., Henderson, A. and Lopes, C. 2002. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2002), 415–422.

[5] Bellotti, V. and Edwards, K. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human–Computer Interaction*. 16, 2–4 (Dec. 2001), 193–212. DOI:https://doi.org/10.1207/S15327051HCI16234_05.

[6] Beneteau, E., Richards, O.K., Zhang, M., Kientz, J.A., Yip, J. and Hiniker, A. 2019. Communication Breakdowns Between Families and Alexa. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), 243:1–243:13.

[7] Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B. and Lottrjdge, D. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2, 3 (Sep. 2018), 1–24. DOI:https://doi.org/10.1145/3264901.

[8] Blandford, A., Furniss, D. and Makri, S. 2016. *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool.

[9] Branham, M., S. and Roy Rishin Mukkath, A. 2019. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. *ASSETS '19* (Oct. 2019).

[10] Brinkmann, S. and Kvale, S. 2014. *Interviews - Learning the Craft of Qualitative Research Interviewing*. Sage Publications Inc.

[11] Bunt, A., Lount, M. and Lauzon, C. 2012. Are explanations always important?: a study of deployed, low-cost intelligent interactive systems. *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces* (Feb. 2012), 169–178.

[12] Cha, N., Kim, A., Park, C.Y., Kang, S., Park, M., Lee, J.-G., Lee, S. and Lee, U. 2020. "Hello There! Is Now a Good Time to Talk?": Opportune Moments for Proactive Interactions with Smart Speakers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 4, 3 (Sep. 2020), 28. DOI:https://doi.org/10.1145/3411810.

[13] Chalmers, M. 2003. Seamful Design and Ubicomp Infrastructure. *Proceedings of Ubicomp 2003 workshop at the crossroads: The interaction of HCI and systems issues in Ubicomp.* (2003).

[14] Chilana, P.K., Ko, A.J. and Wobbrock, J.O. 2012. LemonAid: selection-based crowdsourced contextual help for web applications. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (Austin, Texas, USA, 2012), 1549.

[15] Cho, J. 2018. Mental Models and Home Virtual Assistants (HVAs). *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), SRC05:1–SRC05:6.

[16] Cho, M., Lee, S. and Lee, K.-P. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home Are Forgotten. *Proceedings of the 2019 on Designing Interactive Systems Conference* (New York, NY, USA, 2019), 1557–1569.

[17] Chuang, Y., Chen, L. and Liu, Y. 2018. Design vocabulary for human-IoT systems communication. *CHI 2018 - Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems: Engage with CHI* (Apr. 2018).

[18] Corbin, J. and Strauss, A. 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, Inc.

[19] Cowan, B.R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D. and Bandeira, N. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2017), 43:1–43:12.

[20] Dey, A.K. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing*. 5, 1 (Jan. 2001), 4–7. DOI:https://doi.org/10.1007/s007790170019.

[21] Dey, A.K. and Newberger, A. 2009. Support for context-aware intelligibility and control. (Apr. 2009), 859–868.

[22] Druga, S., Williams, R., Breazeal, C. and Resnick, M. 2017. "Hey Google is It OK if I Eat You?": Initial Explorations in Child-Agent Interaction. *Proceedings of the 2017 Conference on Interaction Design and Children* (New York, NY, USA, 2017), 595–600.

[23] Edwards, W.K., Newman, M.W. and Poole, E.S. 2010. The infrastructure problem in HCI. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (Atlanta, Georgia, USA, 2010), 423.

[24] Fraser, C.A., Dontcheva, M., Winnemöller, H., Ehrlich, S. and Klemmer, S. 2016. DiscoverySpace: Suggesting Actions in Complex Software. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16* (Brisbane, QLD, Australia, 2016), 1221–1232.

[25] Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R. and Wang, M. 2011. Proxemic interactions: the new ubicomp? *interactions*. 18, 1 (Jan. 2011), 42. DOI:https://doi.org/10.1145/1897239.1897250.

[26] Gregor, P., Newell, A.F. and Zajicek, M. 2002. Designing for Dynamic Diversity: Interfaces for Older People. *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (New York, NY, USA, 2002), 151–156.

[27] Hornecker, E. and Buur, J. 2006. Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), 437–446.

[28] Jiang, J., Jeng, W. and He, D. 2013. How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2013), 143–152.

[29] Ju, W. and Takayama, L. 2009. Approachability: How People Interpret Automatic Door Movement as Gesture. *International Journal of Design*. Vol. 3(2), Design and Emotion (Aug. 2009), 15.

[30] Klein, L. 2015. *Design for Voice Interfaces*. O'Reilly Media, Inc.

[31] Kulesza, T., Burnett, M., Wong, W.-K. and Stumpf, S. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15* (Atlanta, Georgia, USA, 2015), 126–137.

[32] Lau, J., Zimmerman, B. and Schaub, F. 2018. Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.*2, CSCW (Nov. 2018), 102:1–102:31. DOI:https://doi.org/10.1145/3274371.

[33] Lim, B.Y. and Dey, A.K. 2009. Assessing demand for intelligibility in context-aware applications. *Proceedings of the 11th international conference on Ubiquitous computing - Ubicomp '09* (Orlando, Florida, USA, 2009), 195.

[34] Lim, B.Y. and Dey, A.K. 2011. Design of an Intelligible Mobile Context-aware Application. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2011), 157–166.

[35] Lim, B.Y. and Dey, A.K. 2011. Investigating Intelligibility for Uncertain Context-aware Applications. *Proceedings of the 13th International Conference on Ubiquitous Computing* (New York, NY, USA, 2011), 415–424.

[36] Lim, B.Y. and Dey, A.K. 2010. Toolkit to support intelligibility in context-aware applications. *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10* (Copenhagen, Denmark, 2010), 13.

[37] Lim, B.Y., Dey, A.K. and Avrahami, D. 2009. Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), 2119–2128.

[38] Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q. and Martinez, A. 2018. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*. (Mar. 2018), 096100061875941 4. DOI:https://doi.org/10.1177/0961000618759414.

[39] Luger, E. and Sellen, A. 2016. "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* (New York, NY, USA, 2016), 5286–5297.

[40] Matejka, J., Li, W., Grossman, T. and Fitzmaurice, G. 2009. CommunityCommands: command recommendations for software applications. *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST '09* (Victoria, BC, Canada, 2009), 193.

[41] McMillan, D., Brown, B., Kawaguchi, I., Jaber, R., Solsona Belenguer, J. and Kuzuoka, H. 2019. Designing with Gaze: Tama – a Gaze Activated Smart-Speaker. *Proceedings of the ACM on Human-Computer Interaction*. 3, CSCW (Nov. 2019), 1–26. DOI:https://doi.org/10.1145/3359278.

[42] Mennicken, S., Brillman, R., Thom, J. and Cramer, H. 2018. Challenges and Methods in Design of Domain-Specific Voice Assistants. *AAAI Spring Symposium Series*. (2018), 5.

[43] Mennicken, S. and Huang, E.M. 2012. Hacking the Natural Habitat: An In-the-Wild Study of Smart Homes, Their Development, and the People Who Live in Them. *Pervasive Computing*. J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger, eds. Springer Berlin Heidelberg. 143–160.

[44] Mennicken, S., Vermeulen, J. and Huang, E.M. 2014. From today's augmented houses to tomorrow's smart homes: new directions for home automation research. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct* (Seattle, Washington, 2014), 105–115.

[45] Murad, C., Munteanu, C., Cowan, B.R. and Clark, L. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing*. 18, 2 (Apr. 2019), 33–45. DOI:https://doi.org/10.1109/MPRV.2019.2906991.

[46] Myers, B.A., Weitzman, D.A., Ko, A.J. and Chau, D.H. 2006. Answering why and why not questions in user interfaces. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* (Montréal, Québec, Canada, 2006), 397.

[47] Myers, C., Furqan, A., Nebolsky, J., Caro, K. and Zhu, J. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (Montreal QC, Canada, 2018), 1–7.

[48] Myers, C.M., Grethlein, D., Furqan, A., Ontañón, S. and Zhu, J. 2019. Modeling Behavior Patterns with an Unfamiliar Voice User Interface. *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization - UMAP '19* (Larnaca, Cyprus, 2019), 196–200.

[49] Nest Learning Thermostat: 2020. *https://store.google.com/us/product/nest_learning_thermostat_3rd_gen?hl=en-US*. Accessed: 2020-08-14.

[50] Nielsen, J. 1994.10 Heuristics for User Interface Design: Article by Jakob Nielsen.

[51] Niemantsverdriet, K., Broekhuijsen, M., van Essen, H. and Eggen, B. 2016. Designing for Multi-User Interaction in the Home Environment: Implementing Social Translucence. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (New York, NY, USA, 2016), 1303–1314.

[52] Norval, C. and Singh, J. 2019. Explaining automated environments: interrogating scripts, logs, and provenance using voice-assistants. *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers - UbiComp/ISWC '19* (London, United Kingdom, 2019), 332–335.

[53] Nowacka, D., Wolf, K., Costanza, E. and Kirk, D. 2018. Working with an Autonomous Interface: Exploring the Output Space of an Interactive Desktop Lamp. *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '18* (Stockholm, Sweden, 2018), 1–10.

[54] van Oosterhout, A., Bruns Alonso, M. and Jumisko-Pyykkö, S. 2018. Ripple Thermostat: Affecting the Emotional Experience through Interactive Force Feedback and Shape Change. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (Montreal QC, Canada, 2018), 1–12.

[55] Pelikan, H.R.M. and Broth, M. 2016. Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), 4921–4932.

[56] Porcheron, M., Fischer, J.E., Reeves, S. and Sharples, S. 2018. Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 640:1–640:12.

[57] Pradhan, A., Mehta, K. and Findlater, L. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 459:1–459:13.

[58] Reeves, S. 2019. Conversation considered harmful? (Aug. 2019), 10.

[59] Reeves, S., Porcheron, M. and Fischer, J. 2018. "This is Not What We Wanted": Designing for Conversation with Voice Interfaces. *Interactions*. 26, 1 (Dec. 2018), 46–51. DOI:https://doi.org/10.1145/3296699.

[60] Sciuto, A., Saini, A., Forlizzi, J. and Hong, J.I. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 Designing Interactive Systems Conference* (New York, NY, USA, 2018), 857–868.

[61] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N. and Diakopoulos, N. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson.

[62] Søndergaard, M.L.J. and Hansen, L.K. 2018. Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18* (Hong Kong, China, 2018), 869–880.

[63] Sung, J.-Y., Guo, L., Grinter, R.E. and Christensen, H.I. 2007. "My Roomba Is Rambo": Intimate Home Appliances. *UbiComp 2007: Ubiquitous Computing* (2007), 145–162.

[64] The European Parliament and the Council of the European Union 2020. *Regulations*.

[65] Ur, B., McManus, E., Pak Yong Ho, M. and Littman, M.L. 2014. Practical trigger-action programming in the smart home. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (Toronto, Ontario, Canada, 2014), 803–812.

[66] U.S. Smart Speaker Ownership Rises 40% in 2018 to 66.4 Million and Amazon Echo Maintains Market Share Lead Says New Report from Voicebot: 2019. *https://voicebot.ai/2019/03/07/u-s-smart-speaker-ownership-rises-40-in-2018-to-66-4-million-and-amazon-echo-maintains-market-share-lead-says-new-report-from-voicebot/*. Accessed: 2019-05-20.

[67] Vermeulen, J., Slenders, J., Luyten, K. and Coninx, K. 2009. I Bet You Look Good on the Wall: Making the Invisible Computer Visible. *Ambient Intelligence* (2009), 196–205.

[68] Vermeulen, J., Vanderhulst, G., Luyten, K. and Coninx, K. 2010. PervasiveCrystal: Asking and Answering Why and Why Not Questions about Pervasive Computing Applications. *2010 Sixth International Conference on Intelligent Environments* (Jul. 2010), 271–276.

[69] Weiser, M. and Brown, J.S. 1997. The Coming Age of Calm Technology. *Beyond Calculation: The Next Fifty Years of Computing*. P.J. Denning and R.M. Metcalfe, eds. Springer New York. 75–85.

[70] West, M., Kraut, R. and Ei Chew, H. 2019. *I'd blush if I could: closing gender divides in digital skills through education*. UNESCO.

[71] Winograd, T., Flores, F. and Flores, F.F. 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Intellect Books.

[72] Wu, Y., Rough, D., Bleakley, A., Edwards, J., Cooney, O., Doyle, P.R., Clark, L. and Cowan, B.R. 2020. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg Germany, Oct. 2020), 1–9.

[73] Yang, R. and Newman, M.W. 2013. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2013), 93–102.

[74] Yankelovich, N. 1996. How do users know what to say? *interactions*. 3, 6 (Dec. 1996), 32–43. DOI:https://doi.org/10.1145/242485.242500.

[75] You Can Now Use Google Home as an Intercom: 2017. *https://uk.pcmag.com/news-analysis/91988/you-can-now-use-google-home-as-an-intercom*. Accessed: 2019-07-11.