Figure 1: The Amazon Echo, a popular smart speaker. The Echo has four physical buttons on top (volume up, volume down, mute and action) and visual feedback is limited to an LED ring that can pulsate and change color [2].

The Curious Case of Providing Intelligibility for Smart Speakers

Jo Vermeulen[1], Brian Lim[2], Mirzel Avdic[1], Danding Wang[2], and Ashraf Abdul[2]
[1]Department of Computer Science, Aarhus University, Aarhus, Denmark
[2]National University of Singapore, Singapore, Singapore
[1][jo.vermeulen, miavd18]@cs.au.dk
[2][brian.lim, wangdanding, ashrafabdul]@comp.nus.edu.sg

ABSTRACT

AI techniques are increasingly incorporated into everyday devices and appliances. Explainable AI (XAI) is an approach to improve algorithmic transparency, in which systems explain how they arrive at their decisions. Smart speakers are a specific AI technology that is gaining popularity in the home and is affecting people in their everyday lives. While useful for entertainment purposes and for information queries, studies have also pointed out that people often still experience problems in using this nascent technology. From interviews with smart speakers users, we identified issues people face with their smart speakers and how they go about recovering from these breakdowns. In this position paper, we discuss how smart speakers – as a popular smart device with limited visual feedback – provides an interesting case to investigate how to formally provide explanation support and improve intelligibility.

KEYWORDS

Intelligibility, Explainable AI, Smart Speakers

INTRODUCTION

There is increasing interest in bridging the AI and HCI communities to ensure transparency of systems that rely on machine learning, and to put forward a research agenda for HCI and AI. In our previous work [1], we provided an overview of the relationships between various subcommunities in HCI and AI, and revealed fading and burgeoning trends in explainable systems, as well closely connected and isolated domains. For instance, we suggested the use of real-world data with functional

Where is the Human? CHI '19 Workshop, May 04, 2019, Glasgow, Scotland

BREAKDOWN RECOVERY STRATEGIES

One of our findings relates to strategies participants used to recover from breakdowns. Several participants mentioned they oriented themselves towards the smart speaker or walked to the smart speaker to repeat commands when trying to recover from breakdowns, though not always with success. We hypothesize that the conversational nature of the interface may lead users to react analogously to when they are conversing with another person who doesn't understand them – with a natural response to speak louder and clearer (and perhaps get closer). This implies a possibly incorrect mental model, as there are many levels at which the smart speaker may fail to understand a user. For instance, the particular command may not have been part of the smart speaker's set of supported actions. Other factors that contribute are e.g. accents, incorrect pronunciation, wrong grammar, the use of incorrect commands or incorrect names of connected appliances. This suggests that an approach to provide intelligibility or explanations for smart speakers needs to be able to deal with all of these possible causes for breakdowns.

Sidebar 1: Breakdown recovery strategies we observed in our study on intelligibility issues with smart speakers.

complex models, to draw from the rich body of research in HCI to improve usability of intelligible or explainable interfaces, and to draw from theoretical work on the psychology of explanations to arrive at explanations that are easier to interpret. In this position paper, we focus on providing explanations for the behaviour of smart speakers. We see smart speakers as an interesting case to explore research challenges regarding Intelligibility and Explainable AI (XAI). In ongoing studies, we observed that users incorrect mental models may lead to ineffective ways of attempting to recover from breakdowns. To address this challenge, we argue that smart speakers require several types of explanations at different layers of reasoning. We also highlight the challenge of providing explanations in a faceless conversational interface with very limited visual feedback, and discuss possible research directions.

STUDYING BREAKDOWNS WITH SMART SPEAKERS USERS

Recent studies of smart speaker usage have suggested that people experience instances of 'black box' behaviour [8]. In our ongoing work [3], we are studying people's experiences with smart speakers to better understand situations in which these devices exhibit unintelligible behaviour and how they recover from it. To answer these questions, we conducted an online survey (N=117) and semi-structured interviews (N=12) with smart speaker owners. We observed various intelligibility issues. One participant experienced that his smart speaker turned on the smart lights while he was away on vacation. On his smartphone, he was able to see that the lights were turned on while but he had no means to not being able to do anything about it; the lights stayed on for a week. This demonstrates additional issues when smart speakers interface with other Internet-of-Things (IoT) technologies, as people might not be aware that the system executed an action. Other participants mentioned it was difficult to keep track of connected devices and the different commands that the speaker could recognize, i.e., when the home automation setup expanded to a point where the participant may not be able to remember all connected devices, their names, or location (e.g. names of smart lightbulbs) or particular custom commands they configured (e.g. IFTTT rules).

THEORY-DRIVEN USER-CENTRIC EXPLAINABLE AI

Recently, we have proposed a theory-driven, user-centric XAI framework that connects XAI explanation features to underlying reasoning processes that people have for explanations [11]. With this framework (Figure 2), it is possible to identify pathways for how specific explanations can be useful, how certain reasoning methods fail due to cognitive biases, and how to apply different elements of XAI to mitigate these failures. By articulating a detailed design space of technical features of XAI and user requirements of human reasoning, this framework aims to help developers build more user-centric explainable AI-based systems with targeted XAI features. Using our framework, previously explored intelligibility explanation types such as "why not" questions and "how to" questions [5, 6, 10], can be explained as

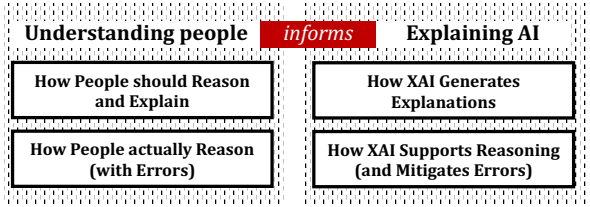

Figure 2: Conceptual framework for Reasoned Explanations that describes how human reasoning processes (left) informs XAI techniques (right) [11]. Theories of human reasoning inform XAI features, and there are further inter-relations between different reasoning processes and associations between XAI features.

1. **Sound and Speech Recognition**
2. **Natural Language Understanding**
3. **Context Inference**
4. **Decision To Act**

Sidebar 2: Four suggested layers of reasoning for smart speaker XAI.

facilitating contrastive and counterfactual reasoning respectively [7]. This framework provides us a basis to select explanations for smart speakers, which we discuss next.

TOWARDS XAI FOR SMART SPEAKERS

Our study of smart speaker breakdowns [3] revealed some potential mismatches in how people understand underlying causes of breakdowns with their smart speakers. We found that after the initial novelty period, most annoyance happened when the system misbehaved or did nothing. Hence, users were interest in *fault finding* to understand about the system misbehavior and we focus on this explanation goal for determining appropriate explanations [11]. While working to apply the framework for smart speakers, we encountered several challenges. First, as an AI-driven ubiquitous computing device, there are multiple layers of reasoning it performs and each layer may require different explanations. A second issue is the faceless interface of smart speakers as an IoT device.

A Layered Approach to Explainability

We argue it is promising to consider a layered approach to explainability for smart speakers, where explanations are provided at different levels, depending on what users want to know. Consider a breakdown that the system does not respond while they're cooking in a noisy environment with the hood fan. The user says "Alexa, turn on the kitchen light", but nothing happens. We propose four layers of reasoning (Sidebar 2). The first layer concerns sound and speech recognition. If things go wrong in this layer, the problem may be that a wrong word is recognized with 60% confidence. A possible explanation technique that could be used is showing what was recognized together with the **confidence** level, indicating that the smart speaker recognized the wrong word. The second layer concerns natural language understanding. Returning to the same example, a **feature attribution** explanation could indicate that the word "turn" was influential for the action "turn off light", but the user would have expected "on" to be influential instead. user may also ask the **counterfactual** question of whether the system would turn on the light if he said"brighten" or "switch on". The third layer concerns modeling the rules of the smart home and context inference, i.e. inferring the current state or user intent. Perhaps, the user learns that the smart speaker thought the living room light should be turned on, instead of the kitchen light. The user may ask the **contrastive** question of why the kitchen was not prioritized. the kitchen light is actually called "kitchen lamp" in the smart home model. The final layer concerns how the smart speaker decides to act, based on decision theoretic and mixed-initiative methods [4]. For instance, the speaker may have thought the user wanted the oven to be turned on, but this would be costly for energy use, so it erred to do nothing instead. The user could ask the speaker what it would take to increase the chances of it correctly responding to the kitchen light request. Other than handling errors and queries at each individual layer, we see that the user may just ask a question without specifying the reasoning layer. This raises research challenges

for the system to infer about which layer is likely to be faulty or which layer the user is querying. This also presents opportunities for new UX design to support ways for users to ask specific layers.

Explanations for Faceless Interfaces

Another challenge for designing intelligible smart speakers is their lack of visual display[1]. Most explanations are conveyed visually and there is much work on visualization methods. In this case, how can we convey explanations for faceless interfaces without screens? Explanations may require a combination of sounds, text-to-speech, LEDs, mobile screen displays, or even ambient displays [9]. Furthermore, users may use smart speakers peripherally. We observed that participants in our study often used the smart speaker while they were doing other activities, and they rarely faced the speaker. This poses further challenges for capturing and maintaining attention while explaining.

[1]Some exceptions include Google's Home Hub or Amazon's Echo Spot

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 740548).

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 582.

[2] Amazon. 2019. Alexa Voice Service (AVS) UX Attention System. https://developer.amazon.com/docs/alexa-voice-service/ux-design-attention.html Accessed February 15, 2019.

[3] Mirzel Avdic and Jo Vermeulen. 2019. In preparation.

[4] Eric Horvitz. 1999. Principles of Mixed-initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166. https://doi.org/10.1145/302979.303030

[5] Brian Y Lim and Anind K Dey. 2009. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th international conference on Ubiquitous computing (Ubicomp '09)*. ACM, 195–204.

[6] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 2119–2128.

[7] Brian Y Lim, Qian Yang, Ashraf Abdul, and Danding Wang. 2019. Why these Explanations? Selecting Intelligibility Types for Explanation Goals *(ExSS '19 Workshop)*.

[8] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. https://doi.org/10.1145/3173574.3174214

[9] Jo Vermeulen, Jonathan Slenders, Kris Luyten, and Karin Coninx. 2009. I bet you look good on the wall: Making the invisible computer visible. In *European Conference on Ambient Intelligence (Aml '09)*. Springer, 196–205.

[10] Jo Vermeulen, Geert Vanderhulst, Kris Luyten, and Karin Coninx. 2010. PervasiveCrystal: Asking and answering why and why not questions about pervasive computing applications. In *2010 Sixth International Conference on Intelligent Environments (IE '10)*. IEEE, 271–276.

[11] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 601, 15 pages. https://doi.org/10.1145/3290605.3300831