Exploring Physical Intelligibility and Control with Smart Speakers

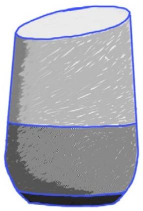

Figure 1: Google's smart speaker, Google Home with the Google Assistant.

Figure 2: Amazon's smart speaker, Amazon Echo with the Alexa assistant.

Mirzel Avdic
Department of Computer Science
Aarhus University
Aarhus N, 8200, Denmark
miavd18@cs.au.dk

DIS '19, June 23–28, 2019, San Diego, CA, USA.

ACM 978-1-4503-6270-2/19/06…$15.00
https://doi.org/10.1145/3301019.3324880

Abstract

Voice-controlled smart speakers with intelligent personal assistants (IPAs) are increasingly becoming ubiquitous in homes. They play a key role in home automation as a hub that interfaces with various appliances. Previous work has suggested that ubiquitous computing systems should be *intelligible* and *controllable*, by informing users about the system's underlying behavior and enabling users to intervene during breakdowns. In my PhD, I investigate *physical intelligibility* for smart speakers: the use of physical motion and interaction to provide intelligibility and control, informed by observations of users approaching and physically interacting with smart speakers.

Author Keywords

Intelligibility; control; physical interaction; voice user interface; breakdowns; explanations.

CSS Concepts

• Human-centered computing~Natural language interfaces • Human-centered computing~Empirical studies in HCI

Introduction

Systems that use sensors to understand the environment in which they are used and act on information they infer have long been discussed by

Table 1: Bellotti et al.'s Five Questions [2].

Address: How do I address one (or more) of many possible devices?

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 possible specify a target or targets for my action?

Alignment: How do I know the system is doing (has done) the right thing?

Accident: How do I avoid mistakes?

Table 2: Research Questions.

(1): What are users' mental models of smart speakers?

(2): How do users address their smart speakers?

(3): How do users recover from mistakes and system breakdowns?

(4): How do users use their smart speaker with others in households and/or when having visitors?

researchers under the name "context-aware systems" [5], and smart speakers are a recent rendition of those systems. In the early 2000s, researchers voiced concerns about context-aware systems' lack of intelligibility and control [4,6]; i.e. informing users of what the system infers, how it has inferred this, what it is doing with that information, and how users can take control over the system if it makes a mistake. Bellotti et al. [4] discussed design concerns regarding sensing systems and proposed five questions (see Table 1) that designers should consider to address key challenges with intelligibility and control. With recent advances in artificial intelligence (AI) and autonomous systems, these issues are again at the forefront of the HCI community [1]. Amershi et al. [2] presented 18 guidelines for human-AI interaction, although they observed that voice assistants were the least compatible with the guidelines, indicating that voice user interfaces (VUIs) might need additional guidelines.

Meanwhile, smart speakers' and IPAs' popularity (Figure 1–2) presents an interesting opportunity to study how an emerging ubiquitous presence in people's homes exposes challenges with respect to intelligibility and control. Porcheron et al. have observed that smart speakers at times exhibit 'black box' behaviour [14]. In their study, participants had difficulties using the smart speaker due to a lack of *interactional resources*, i.e. the lack of clear and informative responses by the smart speaker did not lead the participants in a fruitful direction to recover from breakdowns.

Researchers have demonstrated ways to make systems intelligible [10–12,15], however most designs emphasize textual and visual representations, less suitable to smart speakers due to their common lack of displays. An exception is Ripple Thermostat [13], which shows shape change as a promising alternative. Researchers have also observed an influence of physical movement on people's perception of autonomous artifacts' performance [7] and objects' intent [9]. Ju et al. demonstrated how a system can adapt its digital content on an electronic whiteboard depending on people's physical proximity to it, yet allow people to physically override the changes as they get within arm's reach [8]. In my research, I explore *physical intelligibility,* i.e. intelligibility provided through the artifact's physical motion. I will investigate this within the context of smart speakers, to explore when and how physical intelligibility can be used to convey smart speakers' underlying behaviour, and how it can allow users to intervene, recover, and learn from breakdowns to remain in control.

Research to Date

While Porcheron et al. [14] mention participants experiencing 'black box' behaviour with smart speakers, it is still unclear in which situations users encounter unintelligible behaviour and how they recover from it. I conducted a study that addresses four research questions (see Table 2) to better contextualize issues related to intelligibility and control of smart speakers. Research question 2 and 3 were directly inspired by Bellotti et al.'s questions on *address* and *accident* (Table 1). The study consisted of an online survey and 12 semi-structured interviews with smart speaker owners [3].

The 12 interviewees had four main explanations for how their smart speaker functioned: perceiving the device as a 'dumb' speaker, believing it was action-triggered, thinking most of the functionality was

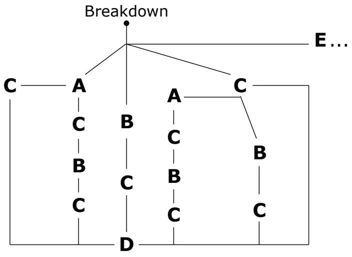

Figure 3: Common strategies participants in our study used to recover from breakdowns.

(A) Face the smart speaker.

(B) Walk up to the smart speaker (facing it).

(C) Retry the request 'X' amount of times.

(D) Try to find the issue on smartphone, use an app to complete the task, or give up entirely.

(E) Other approaches.

happening in the cloud, and finally, understanding most of the process step-by-step for participants with a technical background. While participants generally preferred to address the smart speaker using voice alone to keep their hands free, some participants reported situational physical interactions with the smart speakers depending on their proximity to the device. Moreover, the majority of participants tended to either use their phone, face or walk up to their smart speaker to recover from breakdowns, though not always with success (see Figure 3). Finally, we noticed a commonality among the participants feeling fairly comfortable about sharing their device with household members and visitors. While some visitors and household members were easing into using the smart speaker, they noted others (usually visitors), showed a reluctance in using the devices.

Similar to Yang and Newman's observation of *incidental intelligibility*—a way for the system to explain itself as part of a user's ongoing task [16]—my observations also seem to suggest an opportunity with respect to situational physical interactions and proximity during breakdowns: providing intelligibility in proximity.

Future Research

The Design of Physical Intelligibility. I am planning on developing physical prototypes based on Google's AIY Voice Kit that allows me to create connections between voice input/output, physical actuation, and visual & audio feedback. I intend to explore this design space through multiple prototypes. I then consider comparing one or more of the prototypes to commercially available smart speakers in a controlled lab study, to identify benefits and limitations of physical intelligibility and how much information physical intelligibility can convey. I want to explore whether voice and textual explanations for an IPA's underlying behaviour can be conveyed through physical motion, to either fully or partially explain itself to the user as needed. Physical intelligibility could also expand on people's existing physical interactions with speakers as physical motion with input capabilities could open up new directions for controlling smart speakers and IoT devices alike.

Long-Term Deployment. Informed by the findings from the controlled lab study above, I plan to deploy one of the physical prototypes at smart speaker users' homes for a longer period of time, replacing their original smart speakers. My aim is to investigate whether physical intelligibility is beneficial in allowing users to understand and control the device in their own homes over a longer period of time, and to investigate the consequences of physical intelligibility on the interactions between user and device.

Acknowledgements

I would like to thank my advisors Jo Vermeulen and Susanne Bødker for their continued guidance and support. 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).

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 *Proc.* CHI '18, 582:1–582:18. https://doi.org/10.1145/3173574.3174156

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proc.* CHI ’19, 19.

[3] Mirzel Avdic and Jo Vermeulen. 2019. What Alexa, Siri and Google Don’t Tell You: Understanding Intelligibility Issues with Smart Speakers. In *Submission*.

[4] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. In *Proc.* CHI ’02, 415–422. https://doi.org/10.1145/503376.503450

[5] Anind K. Dey. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing* 5, 1: 4–7. https://doi.org/10.1007/s007790170019

[6] W. Keith Edwards and Rebecca E. Grinter. 2001. At Home with Ubiquitous Computing: Seven Challenges. In *Proc*. *Ubicomp '01* (Lecture Notes in Computer Science), 256–272. https://doi.org/10.1007/3-540-45427-6_22

[7] Pedro Garcia Garcia, Enrico Costanza, Sarvapali D. Ramchurn, and Jhim Kiel M. Verame. 2016. The Potential of Physical Motion Cues: Changing People’s Perception of Robots’ Performance. In *Proc.* UbiComp ’16, 510–518. https://doi.org/10.1145/2971648.2971697

[8] Wendy Ju, Brian A. Lee, and Scott R. Klemmer. 2008. Range: exploring implicit interaction through electronic whiteboard design. In *Proc. CSCW ’08*, 17. https://doi.org/10.1145/1460563.1460569

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

[10] Brian Y. Lim and Anind K. Dey. 2010. Toolkit to support intelligibility in context-aware applications. In *Proc. Ubicomp ’10*, 13. https://doi.org/10.1145/1864349.1864353

[11] Brian Y. Lim and Anind K. Dey. 2011. Design of an Intelligible Mobile Context-aware Application. In *Proc.* MobileHCI ’11, 157–166. https://doi.org/10.1145/2037373.2037399

[12] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems. In *Proc.* CHI ’09, 2119–2128. https://doi.org/10.1145/1518701.1519023

[13] Anke van Oosterhout, Miguel Bruns Alonso, and Satu Jumisko-Pyykkö. 2018. Ripple Thermostat: Affecting the Emotional Experience through Interactive Force Feedback and Shape Change. In *Proc. CHI ’18*, 1–12. https://doi.org/10.1145/3173574.3174229

[14] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proc.* CHI ’18, 640:1–640:12. https://doi.org/10.1145/3173574.3174214

[15] Jo Vermeulen, Jonathan Slenders, Kris Luyten, and Karin Coninx. 2009. I Bet You Look Good on the Wall: Making the Invisible Computer Visible. In *AmI* '09 (Lecture Notes in Computer Science), 196–205. https://doi.org/10.1007/978-3-642-05408-2_24

[16] Rayoung Yang and Mark W. Newman. 2013. Learning from a Learning Thermostat: Lessons for Intelligent Systems for the Home. In *Proc.* UbiComp ’13, 93–102. https://doi.org/10.1145/2493432.2493489