

Google Home, Listen: Building Helper Intelligences for Non-Verbal Sound

Zhenfang Chen
zhenfanc@andrew.cmu.edu
School of Architecture, Carnegie
Mellon University
Pittsburgh, Pennsylvania, USA

Daragh Byrne
daraghb@andrew.cmu.edu
School of Architecture, Carnegie
Mellon University
Pittsburgh, Pennsylvania, USA

Dina EL-Zanfaly
delzanfa@andrew.cmu.edu
School of Design, Carnegie Mellon
University
Pittsburgh, Pennsylvania, USA

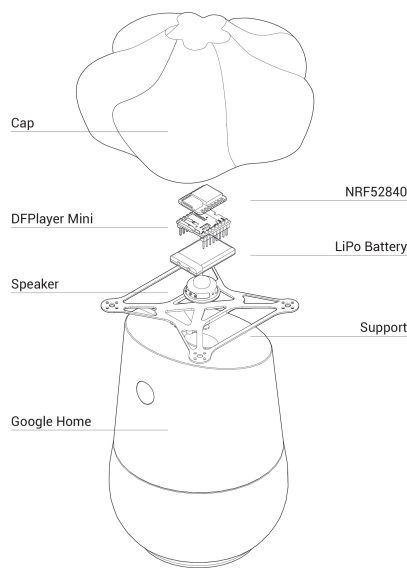


Figure 1: The 'helper' cap and the elements inside.

ABSTRACT

Voice assistants have become a mainstay of our increasingly smart domestic spaces. Within this design-through-research inquiry, how new forms of intelligence might creatively respond to everyday interactions. In this, we are particularly interested in non-verbal communication and expression as a design material for interaction

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

C&C '22, June 20–23, 2022, Venice, Italy

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9327-0/22/06.

<https://doi.org/10.1145/3527927.3535202>

design. In this demonstration, we present a prototype illustrating this opportunity: a tinyML-enabled helper that augments the Google Home and enables it to recognize and respond to everyday phenomenon such as a sneeze, a door opening, or rapping on a table. The helper offers a creative toolkit for designers and end-users to quickly explore, understand, and imagine the affordances of everyday sounds.

CCS CONCEPTS

• Human-centered computing → User interface design; Ambient intelligence; Sound-based input / output.

KEYWORDS

voice assistants, creativity, non-verbal communication, design research

ACM Reference Format:

Zhenfang Chen, Daragh Byrne, and Dina EL-Zanfaly. 2022. Google Home, Listen: Building Helper Intelligences for Non-Verbal Sound. In *Creativity and Cognition (C&C '22)*, June 20–23, 2022, Venice, Italy. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3527927.3535202>

1 INTRODUCTION

Current smart home devices, such as Google Home or Amazon Echo are designed as a voice-based virtual assistant. They help people with information requests, home automation, and entertainment needs, and are generally available in people's homes. These voice assistants operate by responding to structured commands issued as spoken invocations. To use Google Home, for example, people have to say the wake-up words first, "Ok Google" or "Hey Google", and then issue a command, such as "Play some music for me." The voice assistant will then respond to the invocation. It always follows this pattern. It sits there, listening *but only to spoken verbal commands*.

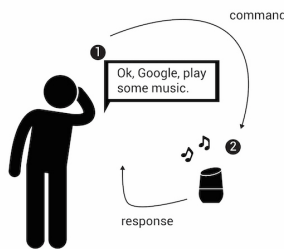


Figure 2: The traditional way of communicating with a smart assistant.

Within this design research project, we ask: how can we make a voice assistant more "lively"? Rather than being constrained to spoken invocations could it understand and interpret everyday phenomenon, routines, habits, and practices (e.g. a door closing, a human sneezing, strumming on a table top)? What design opportunities arise when it might observe and understand the environment or contexts around it?

Responding to these questions, this project – and the demonstration we present – aims to create *prosthetic intelligence*; helpers that can augment a smart home device and enable it to better understand the environment and activities taking place around the device and to respond to these cues. Within this project, we focus specifically on the design opportunities for intelligent understanding within non-verbal audio.

2 BACKGROUND

Since the introduction of voice user interfaces (VUIs) like Alexa, Google Home, and Siri, many design studies have explored how they have shaped interactions and the experience of domestic spaces [2, 13–15]. Porcheron et al [14] conducted an ethnomethodologically informed study of conversations around the Amazon Alexa device. This works highlight how the Alexa 'becomes embedded in the

life of the home', the 'troublesome' nature of navigating responses, and the role that silence plays in these interactions. Beneteau et al [2] similarly studies implications of breakdowns in communication with voice assistants in the home. Voice assistants have been shown to shape thinking about incorporating and interacting with IoT devices, services and digital content [1]. Not only do VUIs shape perceptions and experiences in the home, but work by Sutton et al (2019) [16] highlights that such systems also shapes cognition of language and influence patterns of speech. They also suggest voice – and by extension the sounds that VUIS interpret – as a design material. Recent work by Desjardins et al. [4] and Brand et al. [3] suggest how design experimentation and research can investigate new opportunities, imaginaries and relationships with voice user interfaces and sound materials.

Within this work, we draw particular inspiration from the design speculations of Bjørn Karmann. Project Alias [9], by Karmann and Knudsen, is a teachable 'parasite' or an add-on to smart home assistants that allows for customization of wake words and increased privacy (see Figure 3). Without saying a customized wake-up word, the Alias device will play white noise to interfere with the assistant and interrupt its recognition of users' voices. It also suggests how 'parasites' (or 'helpers' in our case) may be designed to augment VUIs.



Figure 3: The 'parasite'-like design of Alias. Image Credit: Bjørn Karmann & Tore Knudsen CC BY-NC-ND 2.0.

As VUIs are largely organized around structured verbal commands, there are many opportunities for augment VUIs interactions. There is much existing research that examines the affordances of non-verbal sonic gestures as inputs to control and configure interaction. For example, Harrison et al. construct 3D printed acoustic barcodes that can be swiped to produce a distinctive audio signature and in turn mapped to interactive events [7]. Similarly, Knocker [5] recognizes the acoustic characteristics of an object, such as a bottle, as it is knocked against a smartphone while SoundCraft [6] explores non-verbal acoustics produced by hand movements (e.g. finger clicks, body taps, etc.) detected by a close proximity microphone in a smart watch. Recently, Jeong et al. [8] used a 4 microphone array to augment wood and glass panels to detect and

localize knocking and tapping gestures. Kuo & Rawn [10] demonstrated everyday materials – namely the sound of velcro ripping apart – as an interactive input. Kurosawa et al. [11] show that everyday phenomena such as the clacking of a keyboard being typed offer similar potential. These studies highlight non-verbal audio as a rich, and under exploited, resource for designing interactions.

3 OUR DESIGN PROCESS

Our prototype was informed through a range of several speculative experiments into two aspects of non-verbal sounds and communication. We explored what non-verbal sounds might be used for and how people perceive them. Early experiments included augmenting everyday objects to respond to sonic gestures performed on their surfaces. For example, we developed a “listening” lampshade that when scratched or tapped, the color of the lamp changes. We next inventoried everyday sounds and developed a non-verbal sound library. This was developed, in part, by recording ambient audio from a range of rooms within a domestic space. We explored the qualitative and experiential aspects of these sounds by asking people to listen to sounds from the library and describe their responses and feelings of the sound. In addition, we explored sounds produced by everyday gestures and interactions with objects by creating short filmings, documenting interactions with materials, and rapidly prototyping interactive devices. These experiments contributed to the approach of this prototype.

4 DESIGN AND ENGINEERING

Within this section, we describe our approach to a ‘helper’ intelligence for the Google Home and how it enables recognition, interpretation, and response to everyday non-verbal audio. This approach is illustrated in Figure 4.

Drawing inspiration from Project Alias, the ‘helper’ is designed as a small add-on unit that sits on top of the Google Home and is depicted in Figure Figure 1.

This “helper” for Google Home deploys Tiny Machine Learning (TinyML) on an NRF52840 board which has an on-board microphone. To train the device, sound samples – such as sneeze sound and the sound of opening a window – are collected. A model is developed using TensorFlow, converted to the TensorFlowLite format and then can be deployed to the microcontroller unit (MCU). With an on-board pulse density modulation (PDM) digital microphone, the TensorFlowLite model classifies environmental sounds. When a target sound is recognized, the MCU plays a pre-recorded voice command using an MP3 Player unit and attached speakers. For example, if the sound of opening a window is recognized, a corresponding voice command “OK Google, what’s the weather like today?” will be played at a low volume that is only perceptible to the Google Home. Scenarios are illustrated in Figure 5.

The “helper” is organized as a cap that hides all other components inside, an NRF52840 board running a TinyML sound recognition model, a DFPlayer Mini which stores the pre-recorded voice commands, a 2000mAh LiPo battery that powers up the NRF52840 board, a speaker that plays the stored voice command from DFPlayer Mini, and a support frame that holds other components above it.

The idea of a trumpet shape leads to the design of the cap. The trumpet shape can be found in many sound art installations, such as

Sonic Playground by Yuri Suzuki [17] and Hypothetisches Gebilde by Alicja Kwade [12]. The final design of the cap is twisted to some degree to better fit itself on top of the Google Home. It is 3D printed using white filament to better match the color and aesthetics of the Google Home. The top of the cap is open to allow the environmental sound to pass through and be captured by the PDM digital microphone.



Figure 4: The working principle behind the prototype.

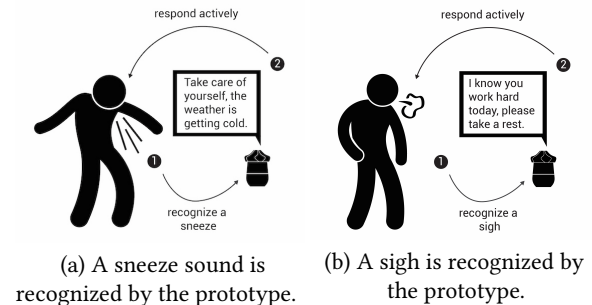


Figure 5: Illustrations of two interactive scenarios.

5 DESIGN OPPORTUNITIES

By adding a helper intelligence with broader forms of sound recognition model to Google Home, this opens up many opportunities to reconsider our relationship to voice assistants and new and creative encounters with these devices. This proposition – and the main contribution of this prototype – is the pipeline of customizing a smart assistant to encounter and react to everyday phenomena, habits, and routines.

It also opens a design space where anyone can fine-tune a smart assistant to better serve their needs. Our prototype offers a platform for anyone to experiment with non-verbal, trainable intelligences; they can quickly train the helper to recognize a wide variety of non-verbal and environmental sound cues and to respond with specific actions.

We conducted a preliminary evaluation of these opportunities. We interviewed four people and gathered anecdotal evidence for these opportunities. We introduced them to the project, how it functions and solicited ideas for the non-verbal sound gestures they would like to be recognized and corresponding responses. For example, one person likes cooking and wants the prototype to recognize the sound of knocking a pot so that the Google Home would create a 15-minute timer. Generally, interviewees enjoyed the playfulness of this project. It suggests many opportunities to gather and co-design interactions with this platform.

As such, we envision this as a resource for future design inquiry that will explore the affordances of everyday sounds in augmenting the capacities of currently voice-enabled intelligences. As part of future work, we plan a series of workshops that use the helper as a generative design tool to investigate our relationship to smart home automation. We also plan to deploy the platform to homes to examine its affordances in rethinking how we encounter devices that occupy our domestic spaces. We also imagine that our approach has much potential in educational settings and could help designers explore, understand, and imagine the potential of environmental and contextual cues in future scenarios interactive experiences.

6 LIMITATIONS, FUTURE WORK, AND CONCLUSIONS

In the current prototype, the TinyML model is trained on sparse data. There is the strong possibility that an environmental or non-verbal sound may be mis-recognized and may inadvertently trigger an unexpected response from Google Home. Future work will aim to assess and improve the robustness of classification. In addition, end-users must manually gather, classify and train environmental sounds as well as record spoken invocations to be played when they are recorded. We plan to develop a companion tool for simplifying this process and to enable creative exploration by those unfamiliar with microcontroller programming. The form and materiality of the cap enclosure can be further studied and improved too. Finally, many opportunities for improvement of recognition of non-verbal sound recognition exist. As this project employs an NRF52840 board that is capable of Bluetooth Low Energy (BLE) communication, it is feasible to network a series of helpers to cooperate and interact. Drawing inspiration from sound-gesture research such as Knocker [5], SoundCraft [6] and Knock&Tap [8], helpers could be embedded in specific surfaces, on-body wearables or trained to recognize audio gestures that are not in the immediate vicinity of the Google Home, that require near proximity or contact with the microphone, or relate to body gestures such as tapping a forearm. Such approaches would increase the reliability of detection in specific non-verbal scenarios and broaden the range of possibilities for interaction.

Importantly, this project does not aim to make a technical product, but to explore alternative modes of how a smart assistant can operate and function in everyday experiences. Through the implementation and demonstration described, we highlight one such opportunity: a trainable helper for voice assistants that expands their capacity to creatively engage in their contexts. It also examines the potential for intimate, poetic and expressive interactions between an individual and a smart assistant, thinking about the

identities of both sides, we humans are more than some user personas that are generated by the data collected by these assistants, and the assistant can be more than a device that programmed by a group of people.

REFERENCES

- [1] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (apr 2019), 28 pages. <https://doi.org/10.1145/3311956>
- [2] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
- [3] Nico Brand, William Odom, and Samuel Barnett. 2021. A Design Inquiry into Introspective AI: Surfacing Opportunities, Issues, and Paradoxes. In *Designing Interactive Systems Conference 2021 (Virtual Event, USA) (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1603–1618. <https://doi.org/10.1145/3461778.3462000>
- [4] Audrey Desjardins, Afroditi Psarra, and Bonnie A. Whiting. 2021. Voices and Voids: Subverting Voice Assistant Systems through Performative Experiments. In *Creativity and Cognition (Virtual Event, Italy) (C&C '21)*. Association for Computing Machinery, New York, NY, USA, Article 29, 10 pages. <https://doi.org/10.1145/3450741.3466807>
- [5] Taesik Gong, Hyunsung Cho, Bowon Lee, and Sung-Ju Lee. 2018. Identifying Everyday Objects with a Smartphone Knock. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [6] Teng Han, Khalad Hasan, Keisuke Nakamura, Randy Gomez, and Pourang Irani. 2017. Soundcraft: Enabling spatial interactions on smartwatches using hand generated acoustics. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 579–591.
- [7] Chris Harrison, Robert Xiao, and Scott Hudson. 2012. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 563–568.
- [8] Jae-Yeop Jeong, Jung-Hwa Kim, Ha-Yeong Yoon, and Jin-Woo Jeong. 2021. Knock&Tap: Classification and Localization of Knock and Tap Gestures using Deep Sound Transfer Learning. In *Companion Publication of the 2021 International Conference on Multimodal Interaction*. 1–6.
- [9] Björn Karmann. 2018. Project Alias. https://bjoernkarmann.dk/project_alias
- [10] Tzu-Sheng Kuo and Eric Rawn. 2020. Let It Rip! Using Velcro for Acoustic Labeling. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 28–30.
- [11] Toshifumi Kurosawa, Buntarou Shizuki, and Jiro Tanaka. 2013. Keyboard clawing: input method by clawing key tops. In *International Conference on Human-Computer Interaction*. Springer, 272–280.
- [12] Alicja Kwade. 2015. Hypothetisches Gebilde. https://alicjakwade.com/works/hypothetisches_gebilde
- [13] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–31.
- [14] 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*. 1–12.
- [15] Alex Sciuto, Armita Saini, Jodi Forlizzi, and Jason I Hong. 2018. "Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 designing interactive systems conference*. 857–868.
- [16] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300833>
- [17] Yuri Suzuki. 2018. Sonic Playground. <https://www.yurisuzuki.com/artist/sonicplayground>