

CARNEGIE MELLON UNIVERSITY

School of Architecture
College of Fine Arts

Thesis

Submitted in Partial Fulfillment of the requirements for the degree of

Master of Science in Computational Design

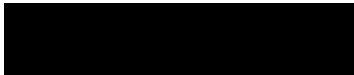
TITLE:

Just Listen:
Prototyping with TinyML to Augment Everyday Sound

AUTHOR:

Zhenfang Chen

ACCEPTED BY ADVISORY COMMITTEE:



Daragh Byrne

Principal Advisor

May 11th, 2022

DATE



Dina El-Zanfaly

Advisor

May 11th, 2022

DATE

Just Listen: Prototyping with TinyML to Augment Everyday Sound

“... the ways in which we intently listen to, look at, or concentrate on anything have a deeply historical character. Whether it is how we behave in front of the luminous screen of a computer or how we experience a performance in an opera house, how we accomplish certain productive, creative, or pedagogical tasks or how we more passively perform routine activities like driving a car or watching television, we are in a dimension of contemporary experience that requires that we effectively cancel out or exclude from consciousness much of our immediate environment.”¹

– Jonathan Crary, *Suspensions of Perception*

¹ Crary, J. (2001). *Suspensions of Perception: Attention, Spectacle, and Modern Culture*.

Abstract

This thesis presents an investigation into everyday sound through a series of speculative experiments and one final refined interactive prototype. The everyday sound here refers to non-verbal sounds, such as the sound of a door opening or someone sneezing. Based on the findings of a series of speculative experiments, the final prototype considers the ubiquitous everyday sound as a material for interaction design. The intention behind the final prototype is to help people become more aware of the prevalence of everyday sound and rethink what it might offer to us in designed interactions.

Sound can be thought of as an intangible material. Although it is invisible, it can be captured with the help of electronic sensors and sensory organs, then it can be modified and utilized as any other physical material. It is everywhere and permeable, whether you actively listen to it or not, it is there. Sound conveys certain information, but only through listening do people decode the information and understand the context it provides. The ability to decode and process sound is gained and polished as time passes; the more you have heard and experienced, the more you can learn from sounds.

Interestingly, as media technology develops, people seldom question the source of the sound they hear from the media. They might take it for granted, even if the sound is a fabricated sound effect. They decode the information from the sound in the way that is expected by the sound creator. People are less likely to ask the question *“Why does it sound like that?”* and *“What information can I get from that sound?”* Gradually, people tend to accept a “standardized” sound system in everyday life through all kinds of interfaces and regard other sounds as outliers. People certainly benefit from the “standardized” sound system, either because of its efficiency or sweet-sounding. With such a system, less attention will be paid to the everyday sound and less familiarity/intimacy people will relate themselves to the object (the sound source).

If we can become more aware of our relationships to sound/everyday sound, we, as a consequence, might be able to become more aware of the environment in which we dwell and realize the influence that we may pose on the people with whom we share the same space through the sound of our body movements.

The final project presented in the later chapters explores how new forms of intelligence might creatively respond to everyday sound/interactions. In this, I am particularly interested in non-verbal communication and expression as a design material for interaction design. A prototype illustrating this opportunity is presented: a

TinyML-enabled helper that augments the Google Home and enables it to recognize and respond to everyday phenomena 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..

Contents

Abstract	3
Contents	5
Introduction	7
Motivation	7
Research Question and Hypothesis	8
Overview of Thesis	9
Summary of Contributions	10
Background	11
Music, Culture, and Identity	11
Human-Computer Interaction, Action Mapping, and Voice Assistant	13
Art, Design, and Everyday Experience	15
Methods	20
Speculative Experiments	21
Play on Acoustic Barcodes	21
Listening Table	22
Sound Activated Shade	24
Mismatch	26
My Ears, My Sonic Memory	27
They Speak	29
Thing Ethnography	31
Summary of Speculative Experiments	32
Google Home, Listen	33
Context and Related Work	34
Design and Engineering	36

Evaluation and Design Reflections	40
Design Opportunities	40
Pilot Evaluation	40
Formal Evaluation	41
Process	42
Survey	42
Analysis	43
Discussion	48
Conclusion	51
Limitations, and Future Work	51
Bibliography	53

Introduction

Motivation

*“A cortical homunculus is a distorted representation of the human body, based on a neurological “map” of the areas and proportions of the human brain dedicated to processing motor functions, or sensory functions, for different parts of the body.”*² Taking our hands as an example, they can do a lot of things and we can sense a lot through them, so there is a large representation of hands in humans’ brains. The idea of cortical homunculus not only shows us what our body is capable of but it is also a snapshot of our body shaping our environment and activity.

With the development of modern technology, all kinds of digital interfaces have been introduced into our daily lives and we, at the same time, keep adjusting ourselves to these interfaces and making compromises in our activities and routines to interacting with them. As Mark Weiser points out in his famous paper, *“today’s multimedia machine makes the computer screen into a demanding focus of attention rather than allowing it to fade into the background.”*³ We are so addicted to all kinds of screens and devices. No matter how large or small in size they are, it seems like we are “forced” to enter their environment when we need to use them.

All of this makes me think about the figure of the modern homunculi and what do they look like? What are they more capable of and what abilities might they have lost when compared with the original cortical homunculus?

According to my observation and life experience, I could imagine one type of homunculi character who is the epitome of us. They look at their smartphone, touch it with several fingers, and keep staring at the screen most of the day. Thus, I would assume they have one large eye since viewing contents from a screen only requires 2D vision and large fingers with tactile sensors.

Staring at the screen and scrolling, these are the full extent of our interactions with most if not all of our devices. We compress the whole world of information into these tiny screens and decompress it again with our eyes and a couple of fingers. This act of decompressing our information takes a lot of attention and it is why

² Available at: https://en.wikipedia.org/wiki/Cortical_homunculus.

³ Weiser, M.D. (1991). The computer for the 21st century. Human-Computer Interaction, 933-940.

we cannot remove ourselves from our devices. We are stuck processing the information all the time. These devices are “emphasized”, not “disappeared”. I highly value what Mark Weiser suggested in his paper that *“the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.”*⁴

How can we escape from such a technological “trap”? How can we communicate more freely without confined boundaries? Can we get a sense of usefulness and enjoyment along with efficiency when interacting with devices?

*“... hearing and sound are gaining momentum as novel means for bridging the gap between expert and non-expert publics by both engaging and activating them and/or bringing attention to underrepresented dimensions and nuances of complex phenomena.”*⁵

Sara Lenzi and Paolo Ciuccarelli suggest using sound (in their case through data sonification) as a way to address complicated social issues. What I learned from them is that sound, as a sensory modality, is full of potential. Starting from this point, I am interested in cultivating a design research practice that aims at extending this sensory modality in physical interaction with interactive objects and devices.

Research Question and Hypothesis

The thesis works at the intersection of everyday non-verbal sound, design research and machine learning as a way to learn and experiment with opportunities for sound as a design material. The thesis approaches this through design research and speculative experiments. The research will explore two main questions:

1. *How can we leverage non-verbal sound as a design material for interaction design?*
2. *How to explore, understand, and imagine the affordances of everyday sounds with the help of edge computing technology, such as TinyML?*

Within this work, I focus on Tiny Machine Learning (TinyML). TinyML is an emerging field of study in Machine Learning and Embedded

⁴ Weiser, M.D. (1991). The computer for the 21st century. Human-Computer Interaction, 933-940.

⁵ Lenzi, S., & Ciuccarelli, P. (2020). Intentionality and design in the data sonification of social issues. Big Data & Society, 7.

Systems that explores the types of models you can run on small, low-powered devices like microcontrollers. It has three advantages, low latency of response time, low power consumption of the model, and a high degree of privacy.

The hypothesis for this work is that: TinyML can be used to augment everyday action/object for designers to quickly explore, understand, and imagine the richness and affordances of everyday sounds.

Overview of Thesis

The remainder of this thesis is organized into 5 chapters.

The Background Chapter is an overview of past research and precedent projects that revolve around the idea of using sound as a material to explore the opportunities of technology development, interaction design, and artistic practice. Many past experiments and approaches are introduced to reveal a variety of aspects of sound, to examine the relationship – either sounds relationship to our imagination or the relationship to our everyday practice – between it and ourselves,

In the Methods Chapter, I will introduce my research method and describe a series of speculative experiments I have conducted. In addition, I will describe the prototypes I have made within the development of this thesis. At the end of this chapter, a final project will be presented which is built upon the base that all the speculative experiments pave.

In the Evaluation Chapter, an evaluation – organized as a generative workshop and a participant survey to gather and explore design possibilities – of the final project will be reported. I will reflect on, analyze, describe findings, and draw conclusions based on the presented work and responses from the survey.

The Conclusion Chapter, is the overall conclusion of this thesis. I will discuss what I think can be improved upon in next steps as well as the other possibilities for future research, inquiry and experimentation.

Summary of Contributions

Technical contribution:

This work offers a pipeline of sound recognition, classification, and action mapping . This pipeline can be applied and extended to similar situations for sound interactions. In addition, this work offers a series of demonstrations of a system that applies this pipeline in a number of real world scenarios.

Design contribution:

A category of non-verbal sound gestures is created and examples are provided.

A potential audience for this work includes designers or researchers who are interested in leveraging sound in interaction design. Through my thesis, they will gain an understanding of sound as a design material through my experiments and experiences, as well as, a range of potentials for sound in interaction design.

Background

In this chapter, I will cover different aspects of sound as a design material. The topics range from the discipline of art to human-computer interaction, and from sound design to data sonification. They may not directly contribute to my research but they do help me to better understand sound and they provide guidance as to the directions that I can follow and explore, and that help to inform my own experiments.

Music, Culture, and Identity

“The sensory faculties – seeing, hearing, touching, smelling, and tasting – played a crucial role in the flurry of European exploration that began in the fifteenth century. Making sense of encounters in Africa, Asia, and the Americas required expanding the limits of the known to encompass elements of human and nonhuman life that challenged the boundaries of European imagination. Listening to human and nonhuman sounds and documenting them in writing for audiences back home helped explorers to conceive of the world outside Europe. It also served to justify European domination and subjugation of that “new world”; what Europeans heard as the wild voices of uncivilized places seemed to cry out for Europe’s guiding and civilizing hand. This took the form of colonization: a centuries-long global project with economic, juridical, theological, epistemological, and sensory dimensions.”⁶

The idea of geographical innovation/power centers has a long history dating back to European exploration and has gradually changed people’s way of thinking^{7,8}. I want to address the importance of respecting each unique culture group and individual. As Ahmed Ansari expressed in his Twitter threads⁹: *“White supremacy constructs the non-white other as ‘exotic’, ‘authentic’, ‘mysteriously productive’ etc. while at the same time asserting itself as the non-mystical, ‘rational’ self. This is orientalism at work through knowledge production. Similarly, the repeated claims, across many texts on ‘fundamental’ design principles, about how ‘universal’ design*

⁶ Available at:
<https://thequietus.com/articles/29445-sound-colonialism-and-decolonial-listening-focus-on-sound-emily-hansell-clark>

⁷ Bell, G., & Kaye, J. (2002). Designing Technology for Domestic Spaces: A Kitchen Manifesto. *Gastronomica*, 2, 46-62.

⁸ Suchman, L.A. (2011). Anthropological Relocations and the Limits of Design. *Annual Review of Anthropology*, 40, 1-18.

⁹ Available at:
<https://twitter.com/aansari86/status/1293262346761179137>.

principles, theories, and methods are, or even that people are the same.”

This makes me think of the music production industry. The popular software people use in this field are all from western-based giant technology/music companies. This software is mainly aimed at people who have been educated under western music theories and at people who are in their target culture markets. African musicians or musicians from less represented cultural groups, contexts or populations, as a consequence, have much less opportunity to express their own culture compared with western musicians. Tuning systems and instruments from these less represented cultural groups have little to no chance to be learned or heard by people from outside. Gradually, they could come to an end of extinction. I would assume the idea of ‘universal’ is for and advocated by the group from which people bring this idea but not for other groups from which their voices are not paid attention by people from the former group.

It is obviously non-Anglo-European cultures that have been mostly affected and disfigured by 12-tone equal temperament, but it is not exclusive to them. Anglo-European cultures have been affected too, as have the efforts of 20th century experimental Anglo-European composers. Tuning affects all music, and although the difference in comparison to 12-tone equal temperament is sometimes subtle, it is precisely these subtleties that make a difference. In this sense, “transcultural” could be a good word to describe the nature of these tools.

Leimma¹⁰ and Apotome¹¹, created by Khyam Allami and Counterpoint, are two web based tools designed to allow users to create their own tuning systems and a generative music environment based on the subsets of the created tuning system.

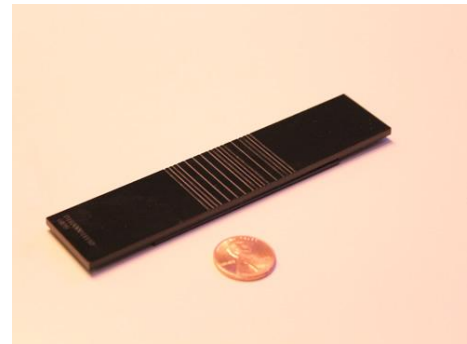
By implementing the relationship between tuning systems and their subsets in a simple and intuitive way, and utilizing both web synths and MIDI, Leimma and Apotome can be integrated into any music-making or music-learning environment. They allow the user to learn then make, or make then learn. To me, applications like these two tools not only provide opportunities for music creation but a platform for any sound (tuning system) to be heard and known by anyone from the earth. Some thoughts come from this, can a sound (not a musical one) represent a “local culture”? How to retrieve the “local culture” from it?

¹⁰ Available at: <https://isartum.net/leimma>.

¹¹ Available at: <https://isartum.net/apotome>.

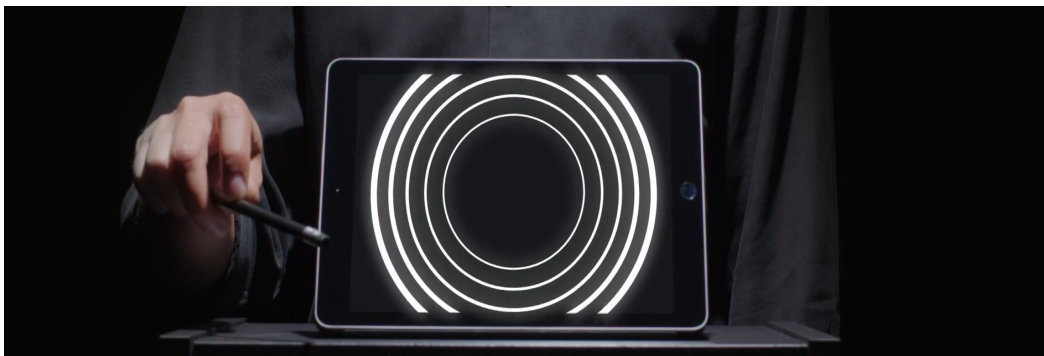
Human-Computer Interaction, Action Mapping, and Voice Assistant

A lot of research relevant to sensing and sound recognition^{12,13,14} has been done in the field of human-computer interaction. The one that typically draws my attention is called “Acoustic Barcodes”¹⁴, by Chris Harrison, Robert Xiao, and Scott Hudson. Acoustic Barcodes are a series of 3D printed materials that produce unique sonic signatures, or barcodes, with structured patterns of physical notches. When a fingernail swipes on them, a complex sound will be generated that can be resolved to a binary ID. A microphone attached to the surface or the object is used to capture the sound that is generated from an interaction with the barcodes. Acoustic Barcodes can be used for information retrieval or to trigger interactive functions. This project provides a good model to me, that based on acoustic sensing I can use non-verbal sound as an input to trigger certain interactions. It is also plausible to apply this technology to different surfaces making it more accessible and approachable.



Acoustic Barcodes
Image Credit: Harrison et al

The company, HYPERSURFACES¹⁵, uses a patented Deep Learning algorithm to convert any object of any material and shape into an intelligent surface that listens to users' inputs, such as a pencil knocking on a wooden table surface.



An intelligent surface listens to a pencil knocking sound.
Image Credit: HYPERSURFACES.

¹² Harrison, C., & Hudson, S.E. (2008). Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. UIST '08.

¹³ Laput, G., Brockmeyer, E., Hudson, S.E., & Harrison, C. (2015). Acoustruments: Passive, Acoustically-Driven, Interactive Controls for Handheld Devices. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.

¹⁴ Harrison, C., Xiao, R., & Hudson, S.E. (2012). Acoustic barcodes: passive, durable and inexpensive notched identification tags. Proceedings of the 25th annual ACM symposium on User interface software and technology.

¹⁵ Available at: <https://www.hypersurfaces.com/>.

Another similar research project is called “Thumprint”¹⁶, by Sauvik Das et al, introduces one group authentication method that authenticates groups based on a shared, three-second knock on a surface instrumented with (or containing) an accelerometer and microphone. Since an accelerometer is employed, this kind of method increases the accuracy of detecting sounds from a moving object.



A smartphone listens to a knocking sound.

Image Credit: Das et al.

Some other interesting projects can be found within the study and development of assistive technology for people with visual impairments since their most reliable senses are auditory and tactile sensations. The two projects^{17,18} – Asakawa et al’s interactive museum experiences and Zhao augmented haptic and auditory cane for navigating VR spaces – I found provide a basic workflow that illustrates how sound feedback and devices can assist people with visual impairments in their daily activities. Here, voice assistants, sound cues, and haptic feedback are three most commonly used techniques.

¹⁶ Das, S., Laput, G., Harrison, C., & Hong, J.I. (2017). Thumprint: Socially-Inclusive Local Group Authentication Through Shared Secret Knocks. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.

¹⁷ Asakawa, S., Guerreiro, J.P., Sato, D., Takagi, H., Ahmetovic, D., Gonzalez, D., Kitani, K.M., & Asakawa, C. (2019). An Independent and Interactive Museum Experience for Blind People. Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web.

¹⁸ Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R., & Sinclair, M. (2018). Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

Another example comes from, VizWiz¹⁹, by Jeffrey Bigham et al., helps people with visual impairments to know the visual information in three simple steps: 1) take a photo of the object or any visual information they want to know; 2) ask a question; 3) receive the answer from the crowd workers.

Although my main focus is not on assistive technology, this kind of technique reminds me of the existence of voice assistant devices that I am seldom aware of. They have become a mainstay of our increasingly smart domestic spaces. I wonder how new forms of intelligence might creatively respond to everyday interactions. Can voice assistants understand our everyday phenomenon such as a sneeze, a door opening, or rapping on a table?

Art, Design, and Everyday Experience

Traditionally, auditory processes are treated as a “secondary” sensory system behind vision. Most of the time, people are used to familiarizing themselves with the affordance of a tangible object through vision and tactile sensation. Auditory senses are less involved or are involved but are not explicitly recognized. From a designer’s perspective, this is a situation in which people are equipped and empowered but they are not aware of it. Some of the following projects aim to raise people’s awareness of auditory perception and realize the importance of it. Learning from them, I aim to closely examine the act of listening itself and to conceive sound as a design material throughout my thesis.

A project from Gershon Dublon and Edwina Portocarrero called “ListenTree”²⁰, is a transducer attached to the underground base of a tree (or trees), transforming the tree into a living speaker that channels audio through its branches and provides vibrotactile feedback. Using this infrastructure any kind of sound can be played through the tree, including live audio or pre-recorded tracks. By resting their head against the tree, people are able to both feel and hear crystal clear sound through bone conduction and might feel a slight vibration under their feet as they approach the tree. Sounding

¹⁹ Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R., Miller, R., Tatarowicz, A., White, B.A., White, S., & Yeh, T. (2010). VizWiz: nearly real-time answers to visual questions. Proceedings of the 23rd annual ACM symposium on User interface software and technology.

²⁰ Portocarrero, E., Dublon, G., Paradiso, J.A., & Bove, V.M. (2015). ListenTree: Audio-Haptic Display In The Natural Environment. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems.

canebrake²¹, is another project that uses a similar approach. These projects suggest the idea of thinking carefully about the act of listening itself. What does the act of listening offer to us? What affordances can we get from an object if we listen closely to it and only through listening? How much attention does listening consume compared with vision and tactile sensations?



A person is listening to the sound that is “played” by a tree.
Image Credit: Gershon Dublon and Edwina Portocarrero.

Not limited to using a tree or trees from an environment, many audio artists use soundwalks as an approach to reveal their relationship with the world. *“A soundwalk is a walk with a focus on listening to the environment. The term was first used by members of the World Soundscape Project under the leadership of composer R. Murray Schafer in Vancouver in the 1970s. Hildegard Westerkamp, from the same group of artists and founder of the World Forum of Acoustic Ecology, defines soundwalking as “... any excursion whose main purpose is listening to the environment. It is exposing our ears to every sound around us no matter where we are.”*²² What I understand from the construct of soundwalks is not just a sonic experience with an environment or with nature, but a way to think about what we can hear from the environment and what we can infer from the sound we hear. Our environment can also be heard as different sound sources when thinking of the information spread in space. I came up with some ideas that instead of me thinking of and sonifying the spatial and sound quality of an environment when

²¹ Available at:
<https://alessandroperini.com/portfolio/sounding-canebrake/>.

²² Available at:
<https://en.wikipedia.org/w/index.php?title=Soundwalk&oldid=1045141226>.

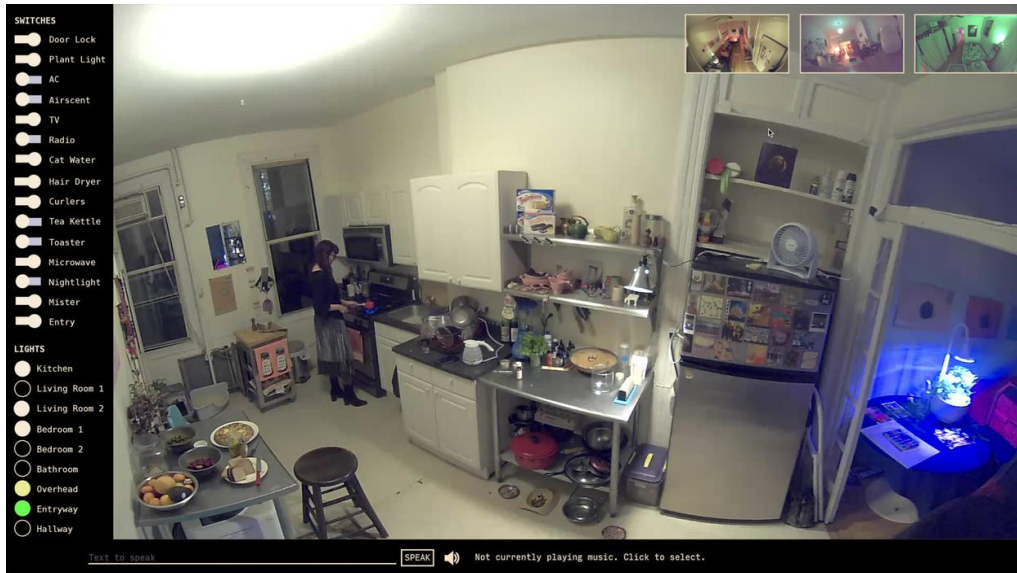
constructing a soundwalk, can we learn or hear anything from the environment we experience or dwell? What will this bring to us? What can we learn from the perspective of the environment?

Interestingly, I found one research project called “Thing Ethnography”²³, which is organized as a design research that aims to provide a new way of thinking about our relationship with daily objects. This work operates from the perspective of a ‘thing’ that is actively situated within relations with other entities, and that has different trajectories than humans. It provides opportunities for people to think from a non-human perspective. People can use this approach to study and to understand what a “thing” experiences in daily life. The insights gathered from such object-oriented ethnography may offer new perspectives on how to shape the relationship between humans and the non-human “thing”. This makes me think about using a living space as a field for possible experiments.

Some similar ideas can be found in the performance LAUREN²⁴, by Lauren Lee McCarthy. This work also examines the relationship between human and object (machine), but more to explore about the interrelated concepts of intimacy, convenience, and privacy. During the performance, Lauren plays the role of the human version of an Amazon Alexa. She aims to serve people better than an AI assistant by watching them in their houses 24/7 and understanding their needs from a human perspective. Based on her interpretation, she then controls all aspects of the house to respond to the occupant’s needs. The performance reveals the ambiguity of interpreting intention as a human and sheds light on issues and potential futures of home automation.

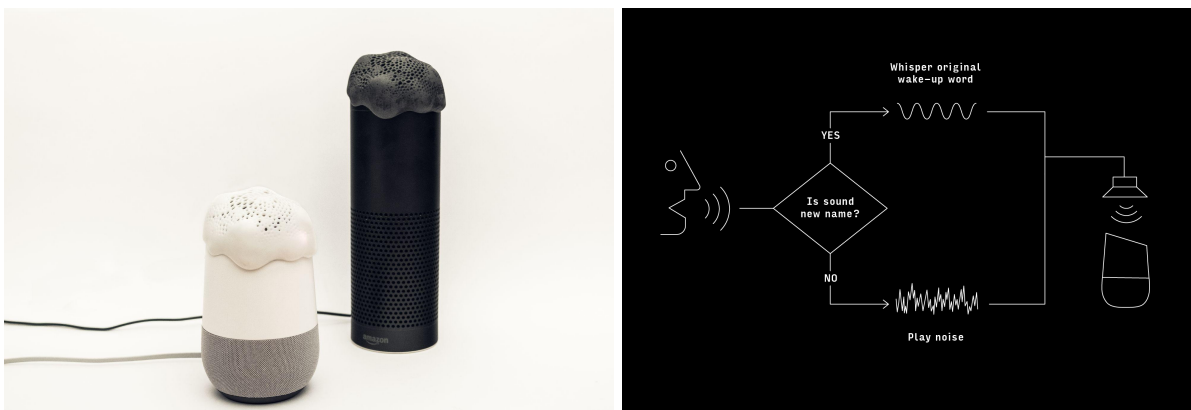
²³ Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016). Thing Ethnography: Doing Design Research with Non-Humans. Proceedings of the 2016 ACM Conference on Designing Interactive Systems.

²⁴ Available at: <https://lauren-mccarthy.com/LAUREN>.



Lauren remotely watches over the person 24/7 and controls all aspects of their home.
Image Credit: Lauren Lee McCarthy.

Another relevant design project aiming at highlighting privacy issue embedded in the modern smart home is called Project Alias²⁵, by Bjørn Karmann and Tore Knudsen. It is an add-on to smart home assistants that aims to give users more control over their privacy and opportunity for customization of the voice assistant. For example, an app is developed that allows someone to change the wake-up word of the assistant. Without saying the customized wake-up word, Alias constantly plays some whitenoise to interfere with the listening process of the assistant. This makes it impossible for the voice assistant to recognize users' voices, and thereby protects their privacy. If Alias recognizes the customized wake-up word, it will stop the noise, it plays a recording of the original wake-up word, and then the assistant functions normally.



Left: The 'parasite'-like design of Alias. Right: Input and output logic of Alias.
Image Credit: Bjørn Karmann and Tore Knudsen.

²⁵ Available at: https://bjoernkarmann.dk/project_alias.

There are some designs of personal assistants and ways of home automation that I appreciate for their uniqueness and playfulness. These have also been a part of my inspiration in developing my final project for the thesis.

Nabaztag²⁶, created by the French company Violet, is one of the very first personal assistants designed in a rabbit shape. Its sound and behavior are carefully designed which makes it adorable. With the 2020 new release kit, people can customize and program the ears' behaviors as they like.

Zelda Ocarina Controlled Home Automation²⁷, by Allen Pan, is a home automation system that is based on the melodies from the Zelda game series using an ocarina. Ocarina of Time is a critical item in the game, Allen cleverly uses an ocarina to salute the Zelda game series. A center device listens to the environment and when it recognizes a melody a corresponding action signal will be sent to the peripheral device. Interestingly, for example, when the melody of Song of Storms is recognized, the humidifier starts to work; when the melody of Song of Time is recognized, the electronic clock tells the current time, etc.



Nabaztag
Image Credit: Rama



Allen Pan is performing Song of Storms to active the humidifier.
Image Credit: Allen Pan.

²⁶ Available at: <https://en.wikipedia.org/wiki/Nabaztag>.

²⁷ Available at: <https://www.youtube.com/watch?v=glZnkpIDWSE>.

Methods

The main focus of this thesis is to explore how to use sound as a material for interaction design, and if possible to reveal the information embedded with that sound.

Within my thesis, a final project uses sound as input to discuss how a built environment can use sensed signals to better serve its occupants and how occupants would enrich the experience of a built environment.

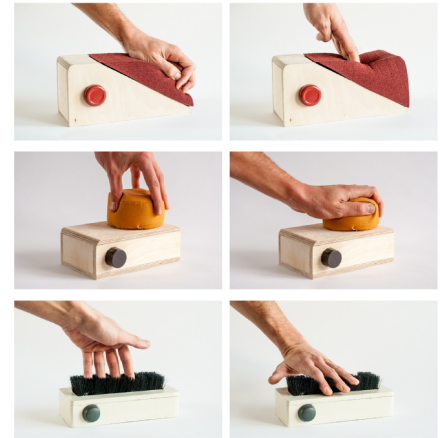
The final project resulted from and was informed by a range of the experiments I conducted in the Fall of 2021.

Before describing my main project, I will introduce a series of experiments that have expanded and scoped my area of study. Although some of the experiments did not directly contribute to the final projects — either technically or theoretically — some aspects of them helped to inform the final outcomes.

Within this work, I use “The Tuning of Materials”²⁸ as a precedent approach to guide my speculative experiments. The authors of the paper conducted a series of material-oriented experiments to help understand the simple tuning action and what materials can facilitate this action. They state:

“Accordingly, the notion of materials experience suggests that designers should qualify the material not only for what it is, but also for what it does, what it expresses to us, what it elicits from us, what it makes us do.”

I was inspired by their approach that simply starts from the study of a material and then gradually begins to explore a range of different ways as to how people would interact with the materials. I decided to adapt this approach to explore the physical and spatial characteristics of sound. Within this, I treat sound as design material, and conduct design-led experiments as an initial inquiry into my own field of study.



Material explorations
Image Credit: Karana et al

²⁸ Karana, E., Giaccardi, E., Stamhuis, N., & Goossensen, J. (2016). The Tuning of Materials: A Designer's Journey. Proceedings of the 2016 ACM Conference on Designing Interactive Systems.

Speculative Experiments

In this section, I will present 7 speculative experiments prepared between September 2021 and February 2022. Within them, I use a number of off-the-shelf toolkits for rapid prototyping, including Teachable Machine²⁹. A range of physical materials have also been used, such as PLA (Polylactic Acid) and air-dry clay.

Play on Acoustic Barcodes (Teachable Machine, 3D Printed forms, 09/2021-10/2021)

I formed the idea for this experiment after reviewing the paper “Acoustic Barcodes”, in which the authors mapped the sounds of scratching barcodes as triggers of corresponding activities when these sounds were recognized and classified by their system.

The main aim of this experiment was to study what parameters will affect the sound recognition and classification processes and how to construct the sound so that it will be distinguishable from other sounds.

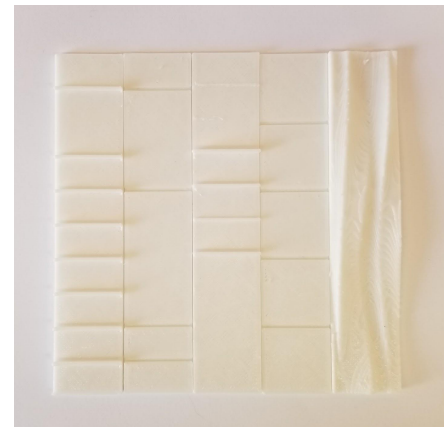
I use an Ender 3 printer to print out all the samples and Ultimaker Cura (version 4.11.0) for slicing.

The parameters that I want to study are the width, length, thickness, and surface texture pattern of a barcode.

In the very first test, the prints share the same width, and thickness, and have no surface texture. The only difference is the length. The length of the printed tiles starts from 10 mm to 100 mm, 10 mm is set for the intervals.

I follow the same procedure to study the other parameters.

I use Teachable Machine to train a sound recognition model. This tool truncates the samples into 1 second long consecutive pieces before starting the training process. In order to get a better result, I had to carefully handle the original sound sample when recording it and select the “good” 1 second long samples for training, otherwise, the sound sample information will be located in two consecutive 1 second long pieces which are regarded as two samples instead of one. Besides, in order to make sure that the original sound sample would not be divided into two pieces, I have to scratch the surface



Different barcode patterns

²⁹ Available at: <https://teachablemachine.withgoogle.com/>.

really fast. It reduces the possibility and opportunity of detecting slow motion or using larger surfaces and increases the noises for the training process.

Different ways of interacting with the surfaces are studied, such as brushes, combs, etc, because I think this will greatly expand the possibilities of applying this kind of technology in everyday life.

From this experiment, I learned that it is possible to recognize and classify sounds of scratching different surfaces. Texture plays the most important role in distinguishing a sound from other sounds. Using a complex surface can generate multiple sounds if scratched from different directions and they can be recognized as different triggers. Besides, it is possible to construct a sound using multiple tiles and it will open up the opportunity for this kind of system.

I think scratching is an intuitive way of interacting with a surface, but what about other hand gestures? What about other materials? Are there any other tools available that I can employ for the classification work?

To answer these questions, I conducted my second experiment.

Listening Table (Processing³⁰, Scikit-learn³¹, 10/2021-11/2021)

The main aim of this experiment is to further study some possibilities uncovered by the first experiment. I do not want to confine the system to barcodes, I want to explore the opportunities with everyday surfaces and other intuitive hand gestures.

I used a smartphone that leverages tap detection and scratch detection using a built-in mic instead of any other peripherals. It is interesting to me that I can cooperate with something that already exists in the environment and utilize it.

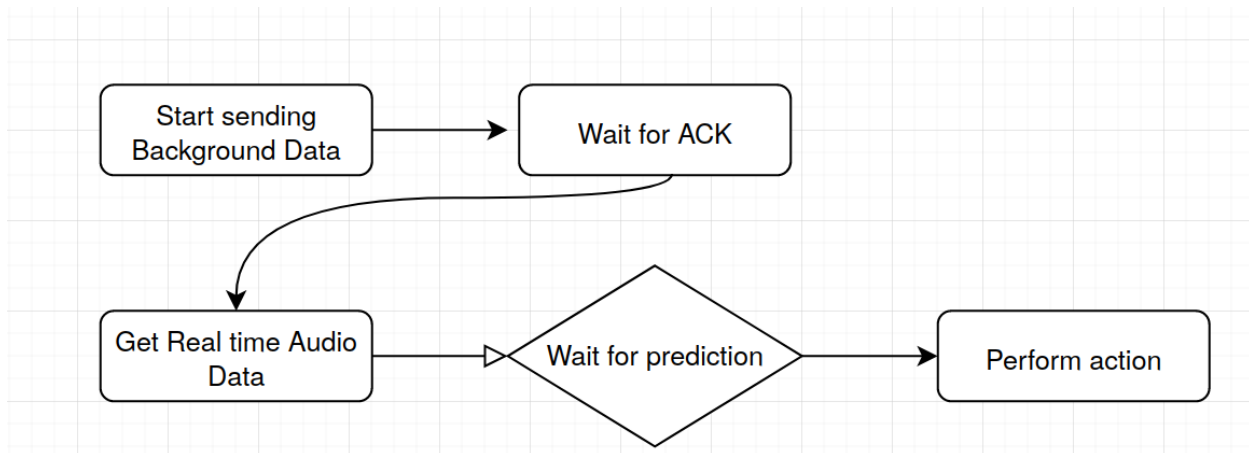
Within this experiment, I used a table surface to control a music player app that is created intentionally to test the feasibility of the idea. When scratch from away to close to the mic, the volume of the music will increase, when scratch from close to the mic to a distance away, the volume of the music will decrease.

The pipeline for the app side: 1) smartphone once connected to python, sends background data first, for about 10 samples; 2) once

³⁰ Available at: <https://processing.org/>.

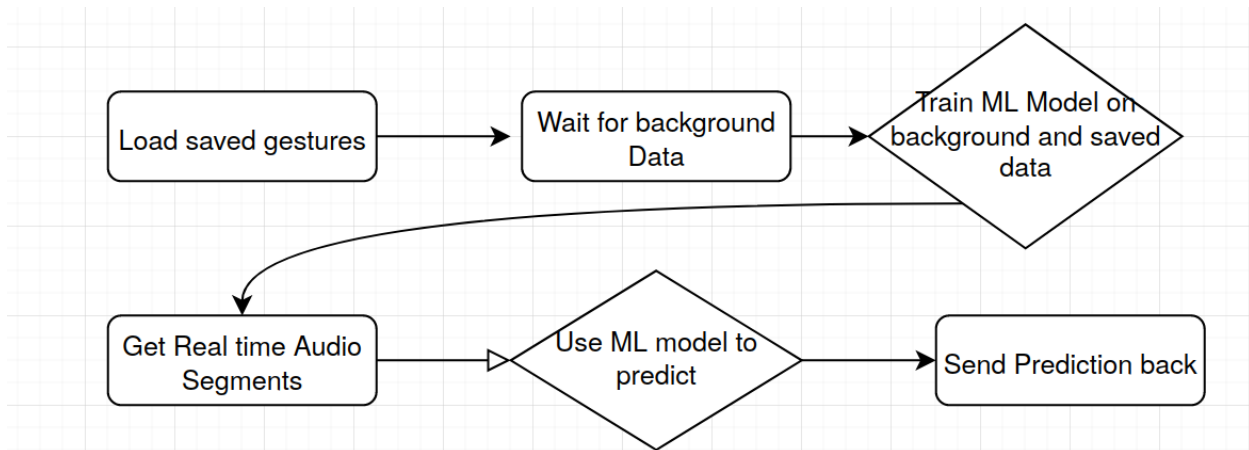
³¹ Available at: <https://scikit-learn.org/stable/>.

the model is trained on background data, gets acknowledgment code; 3) starts sending segments of audio (2.5 secs each); 4) gets prediction in reply; and 5) finally, it performs the action.



The pipeline for the application side.

The pipeline for the python backend: 1) loads saved gestures data; 2) waits for the background data; 3) trains machine learning model on the background and saved data; 4) gets real-time audio segments; 5) uses the machine learning model to predict the gesture; 6) sends prediction back.



The Python backend pipeline.

As a result, the system can classify the sounds generated by different hand gestures performed on the table surface, such as tap, knock, scratch in, and scratch out. When I implemented the scratch in/out approach, I realized that the sound of the music would also be captured which makes it impossible to use scratch to control the volume of the music. Instead, I created an image gallery app (through Processing) that uses scratch in/out sound to show the previous/next image and the tapping sound to add a filter to the displayed image.

From this experiment, I learned that everyday material is full of potential, but a lot of infrastructure work needs to be done to make the system applicable to any material. The intuitive gesture varies when people face different materials, such as scratch/tap/knock for some hard materials (a wooden table) and press/squeeze for some soft materials (a sponge). It would be really lovely if the system can connect the intuitive gesture with the material.

When working on this experiment, I came across a video³² that introduces what a sound designer cares about when creating sounds for a product. As the sound designer puts it:

"At a primitive level, we experience language as music. Contour is the music of the language. It's the melody, the rising and falling of the pitch, and the rhythm, the speeding up and slowing down..."

These words make me think about the relationship between sound and human perception, and how we resonate with the sound we hear.

I start to think about whether I can create a surface so that the sound of scratching/tapping/knocking it is like human verbal language with which a tone is associated. This thought leads to my next experiment.

Sound Activated Shade

(p5.js³³, Arduino³⁴, Teachable Machine, 3D Printed forms, 11/2021)

This experiment explores how to create a sound with which a tone is associated and the feasibility of using sound as a trigger to control a daily object.

Since it is easier for me to build surfaces and adjust them through Rhinoceros 3D, I decided to 3D print the surfaces and test their sound qualities. Besides, as a proof of concept, I chose to light up an LED when the sound is classified by the system.

The pipeline for the experiment: 1) Teachable Machine classifies the sound input; 2) p5.js receives the classified result and communicates

³² Available at: <https://www.youtube.com/watch?v=1sqqg63imHx0>.

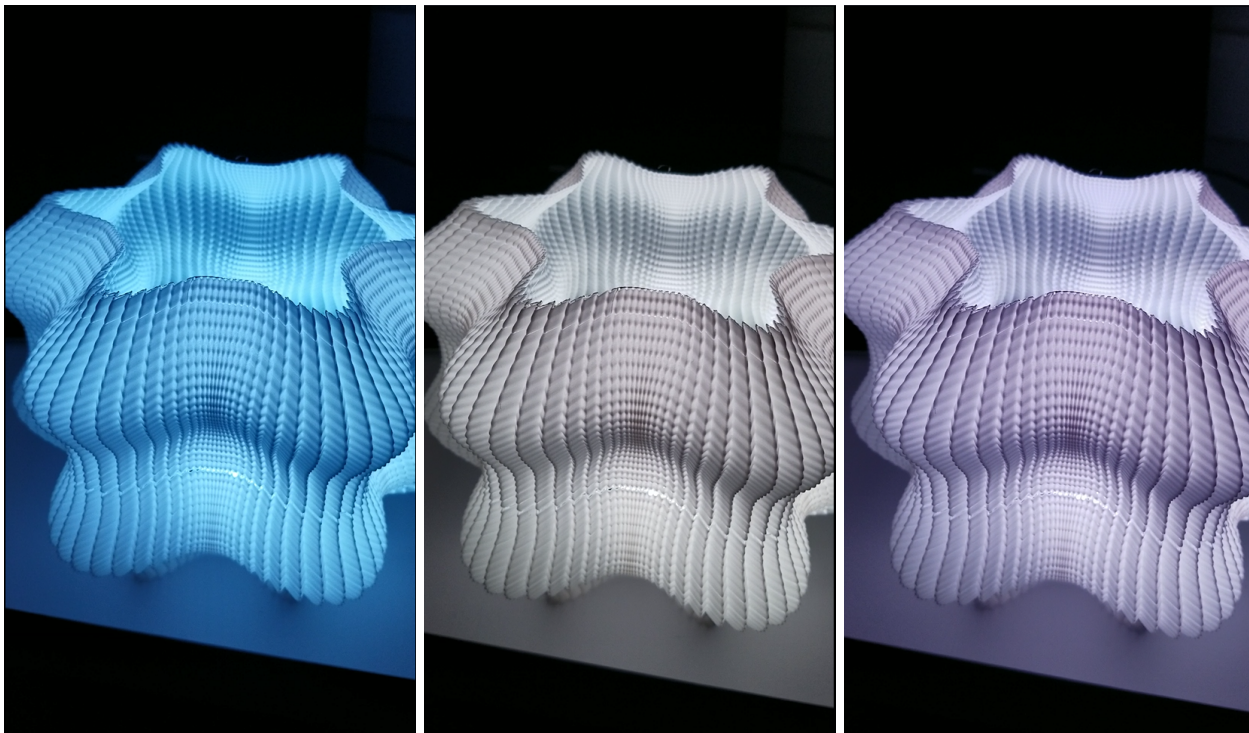
³³ Available at: <https://p5js.org/>.

³⁴ Available at: <https://www.arduino.cc/>.

with Arduino board (ITP has some good resources³⁵ for this step); 3) Arduino board lights up the LED.



Diagram of the connection from the serial port to p5.js through p5.serialcontrol.
Image Credit: ITP Physical Computing.



Left: Tap. Middle: Scratch down. Right: Scratch up.

The 3D printed lampshade has exaggerated surface textures. Three gestures (tap, scratch up, and scratch down) are chosen to control the color of the LED inside the lampshade. A tap on the lampshade lights up the LED in blue color, a scratch-up on the lampshade surface turns the LED into white color, a scratch-down on the lampshade surface turns the LED into pink color.

³⁵ Available at:
<https://itp.nyu.edu/physcomp/labs/labs-serial-communication/lab-serial-output-from-p5-js/>.

Because of the tiny data set I have, the accuracy of the classifier is not very high, especially when classifying the sound of scratch up and scratch down.

From this experiment, I learned that based on my system, if I want to increase the accuracy of sound classification, my finger has to follow the same path and scratches the path at the same speed when creating the training data set. This may be quite interesting in some cases. It is like you have to remember the unique way of creating the sound. Although I want to make a sound with which a tone is associated, I do not think I achieve this goal since the speed of scratching plays an important role which I did not include in my experiments so far.

I come up with some questions after this experiment, what are the intuitive ways to control the light when people face a lamp/lampshade³⁶? Would people connect an object with a certain sound? What the sound will trigger?

Mismatch

(Video editing, 10/2021)

This experiment studies the relationship between sound, hand movements, and body perception.

A series of hand interactions with daily objects are recorded as video clips. Various materials seen in everyday life are chosen when planning this experiment. During the post-processing phase, the sound of a video clip is separated from its original picture and matched/paired with the picture of another video clip to create a mismatched visual and aural perception. The picture is key framed to match the sound pattern. The final edited video clip³⁷ consists of all the processed video clips in a random sequence.

When interacting with the objects, my eyes, hands, and ears are participating at the same time, while in the final edited video clip the hand perception is eliminated. Through this experiment, I want to explore how people perceive an interaction, first through visual organ or auditory organ, how people regard their hand performance with the associated sound, and what would people expect from a such a hand performance.

³⁶ Zheng, C., Oh, H., Devendorf, L., & Do, E.Y. (2019). Sensing Kirigami. Proceedings of the 2019 on Designing Interactive Systems Conference.

³⁷ Available at: <https://vimeo.com/617126807>.

To evaluate this experiment, I interviewed 6 people after they watching the final video clip. Some of them like the idea and want to see and experience some intended examples in real life, for example, what kind of sound and experience they would have when they close a door but hear some sound that is really different to their conventional associated memory. They would imagine that they are entering/leaving a new environment. The precedent “ListenTree”³⁸, in a different setting but shares some interesting insights, employs transducers which are attached to the underground base of a tree so that the tree is transformed into a living speaker. Some of them like the performative aesthetic of the hand gestures and one of them shared a precedent³⁹ with me. I like the way the filmmaker documented the hand gestures that people would perform daily, what are the intentions behind these gestures, and what comes after these gestures.

I like these thoughts and want to expand the study of the performative side of the hand movement and relate that to sound. If people realize that their performance would affect others in the same environment through sound, how would they perform and will this provide an opportunity for collaboration/co-create?

My Ears, My Sonic Memory (Clay modeling, Field recording, Sound mixing, 11/2021)

I always wonder what the associated sound was like when I recall a memory. I remember what happened in that memory and can recall some of the scenes as pictures. Seldom do I remember the sonic part of a memory, but some of the sonic memories haunt me all the time. This makes me start thinking about the ways of capturing a sonic memory or reconstructing one. People can draw some paintings or take pictures to document the visual part of what they are experiencing or what they have experienced, but what about the sonic part?

³⁸ Portocarrero, E., Dublon, G., Paradiso, J.A., & Bove, V.M. (2015). ListenTree: Audio-Haptic Display In The Natural Environment. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems.

³⁹ Available at: <https://vimeo.com/593396130>.

The Sound of Clouds and other Sonic Memories⁴⁰, by Joel Ong, the project covers various topics around listening, politics, personal history, etc.

This experiment explores sets to study the relationship between sound and memory. In this, I am inspired by Ong's work, but take a different approach to creating a sonic memory of my own.

I made a binaural recording system. To do this, I created two air-dry clay models that represent both of my ears. They contain two lavalier microphones (one per ear), connected to a sound recorder. For the recording, I first wrote down a description of a haunting sonic memory and focusing on the different layers and qualities of sound. I use the following text to describe the memory:

In the summer of 2018, I was visiting the Woodland Cemetery in Stockholm. Interestingly, the part that touches me the most is not the giant cross at the entrance, but the scene I experienced after I climbed up a mound. I heard the sound of a woman walking up the mound on my right side and later sitting on a bench inside the grove on top of the mound in front of where I was standing. It was the sound of her leather sandals treading on the grass and her cloth fabric touching the wooden bench. These soft sounds blended with the sound of a breeze crossing the grove that kept and left in my mind until now. I took several images at that time in that place. I do not need to view those images when thinking about the memory.



Making my ear model



The image of the memory I want to recreate.

⁴⁰ Available at:
<https://www.arkfrequencies.com/the-sound-of-clouds-and-other-sonic-memories/>.

Then, I placed the system in different landscapes and tried to capture sounds that match the sound from my memory. Once recorded, I mix them together layer by layer. Mixing the sound is what I call the reconstruction process. Within this, I try to fine-tune the sound sequence and curate the sound sample repeatedly to approach and resemble the original sonic memory. The whole process is similar to the one shown in the video⁴¹ which explores what a cochlear implant sounds like.

I shared the final audio clip⁴² with some people, and asked them about what kind of scenario they created based on it. Some of them told me that they imagined a similar sequence to my scenario but in some other locations with their own details. I really enjoyed listening to their scenarios and noticed that a piece of sound can make people imagine the story behind it.

It is like a human version of machine learning generating a scene based on sound input where people are trained on their entire life experiences.

After this experiment, I learned that sound is capable of bringing people from reality into imagination and each individual will have a unique experience of it. I believe sonic representation has the potential to show the other fold of memory or create an experience, the part that is not visible that softly touches us.



Field recording of different sounds

They Speak (Soundwalk, 11/2021)

This experiment sets out to study how people will perceive a landscape/environment when sound is accompanied when they traverse the landscape.

Without sound, I see landscape as a series of images, a series of information distributed around/within the environment. I have to watch, to see, to observe the environment to collect the information I want. Based on the collected information, I build a sense of the space in my mind with a little bit of narrative from my perspective. I am wondering how differently people would interpret the same landscape in their own narratives. And how differently people would perceive the landscape even when the same sound is accompanied.

⁴¹ Available at:
https://www.youtube.com/watch?v=1dhTWVMcpC4&ab_channel=CochlearImplantResearchLaboratory.

⁴² Available at: <https://vimeo.com/641464390>.

To construct the soundwalk, I picked a route and walked it several times at first to get a general understanding of the environment. Then I divided the route into 5 parts, each I think has a unique story to tell. I sat in each part and listened to its environment. I could hear very different sounds in each part and they revealed information about the environment. I conceived many stories and chose one for each part and connected them as a final piece.



Left: The route I picked for the soundwalk. Right: 6 people conducting my soundwalk.

During the process of creating the soundwalk, I encountered a lot of objects that I did not pay attention to before. I started to look at them, to touch them, and to closely listen to the sound around them. I was wondering what they wanted to tell us.

The result of this experiment is a 15 minutes long soundwalk⁴³, but what I want to say here is that instead of me thinking of and sonicating the spatial and sound quality of an environment, can we learn or hear anything from the environment we experience or dwell? Will these heard sounds create any design space?



I was listening to a sculpture.

⁴³ Available at: <https://soundcloud.com/user-806568602/they-speak>.

Thing Ethnography

(Sound recording from an object's perspective,
12/2021-2/2022)

I followed the Thing Ethnography⁴⁴ approach with some adaptations. Instead of conducting a visual analysis using wearable cameras placed on everyday objects, I used a lavalier microphone to capture and record ambient sounds that might be 'heard' by everyday objects. I began by conducting a four hour study with a lavalier microphone attached to my cup of coffee. According to the playback, the paper cup knew when I was working, when I was leaning on the table, and when I left the room. I thought there are some potential design spaces here and I decided to keep experimenting with this idea.

To further study Thing Ethnography, I was thinking about what object travels with me, what object lives in my private space, and what object lives in shared spaces. I was curious about what sound they can hear and what design spaces I can get from the sound. This time, I put a recorder with them and did a 24-hour recording. It turned out to be difficult for me to imagine what happened during the recording time and to find out possible design space. I did another round of study. This time, I did several 4-hour studies of each object and place and then compared them to see what may be interesting to me. I was assuming that there should be some consistency of the sound quality of a space during certain times of day but there are some uncertainties in my experiment.

Although I did not have a valid idea or conclusion from this experiment, I found it interesting to ask my roommates to guess what was happening at that time when I shared a piece of the recording with them. We all had different ideas sometimes and we could give some audio evidence to support our opinions. This somehow made me think that a sound without any visual context could bring a sense of imagination to its listener.



The setting of the first experiment.

⁴⁴ Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016). Thing Ethnography: Doing Design Research with Non-Humans. Proceedings of the 2016 ACM Conference on Designing Interactive Systems.

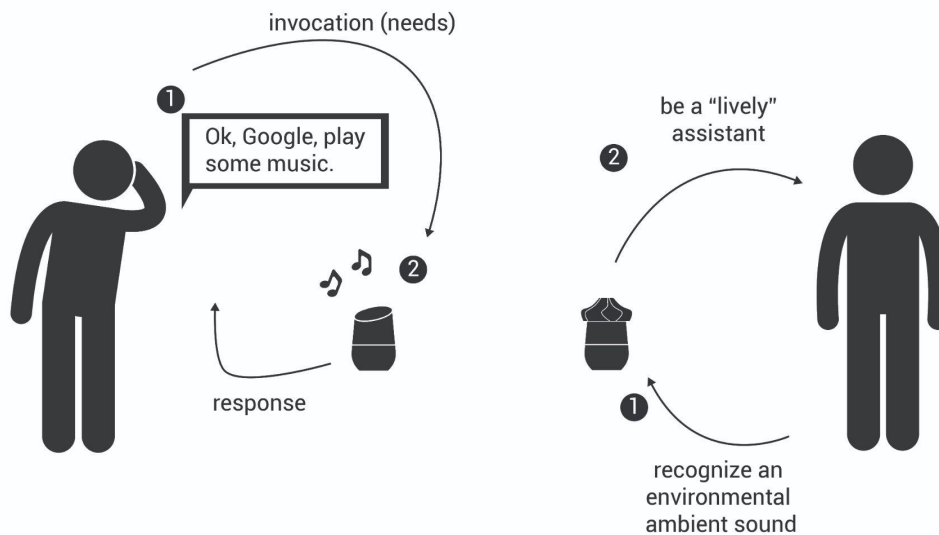
Summary of Speculative Experiments

The seven speculative experiments informed my understanding of sound as a design material. These can largely be organized into two aspects of non-verbal sounds and communication. Through the experiments, I 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, I developed a “listening” lampshade that when scratched or tapped, the color of the lamp changes. I 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. I 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, I explored sounds produced by everyday gestures and interactions with objects by creating short filmings, documenting interactions with materials, and rapidly prototyping interactive devices.

These speculative experiments contributed to the approach of my final prototype: a helper for a Google Home voice assistant. This is described in the next section.

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

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*.



Left: How a voice assistant normally works. Right: What I am trying to create.

Within this design research project, I 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 tabletop)? 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 I present – aims to create *prosthetic intelligences*; 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, I focus specifically on the design opportunities for intelligent understanding within non-verbal audio.

Context and Related Work

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^{45,46,47,48}. Porcheron et al conducted an ethnomethodologically informed study of conversations around the Amazon Alexa device. These 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 similarly study the 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 (Ammari et al⁴⁹). Not only do VUIs shape perceptions and experiences in the home, but work by Sutton et al⁵⁰ highlights that such systems also shape cognition of language and influence patterns of speech. They also suggest voice -- and by extension the sounds that VUIS interprets -- as a design material.

⁴⁵ Porcheron, M., Fischer, J.E., Reeves, S., & Sharples, S.C. (2018). Voice Interfaces in Everyday Life. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

⁴⁶ Beneteau, E., Richards, O.K., Zhang, M., Kientz, J.A., Yip, J.C., & Hiniker, A. (2019). Communication Breakdowns Between Families and Alexa. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

⁴⁷ Sciuto, A., Saini, A., Forlizzi, J., & 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.

⁴⁸ Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening? Proceedings of the ACM on Human-Computer Interaction, 2, 1 - 31.

⁴⁹ Ammari, T., Kaye, J., Tsai, J.Y., & Bentley, F. (2019). Music, Search, and IoT. ACM Transactions on Computer-Human Interaction (TOCHI), 26, 1 - 28.

⁵⁰ Sutton, S.J., Foulkes, P., Kirk, D., & Lawson, S.W. (2019). Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Recent work by Desjardins et al⁵¹ and Brand et al⁵² suggests how to design experimentation and research can investigate new opportunities, imaginaries and relationships with voice user interfaces and sound materials.

Within this work, I draw particular inspiration from the design speculations of Bjørn Karmann. Project Alias⁵³, 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. 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.

As VUIs are largely organized around structured verbal commands, there are many opportunities for augmenting 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 events⁵⁴. Similarly, Knocker⁵⁵ recognizes the acoustic characteristics of an object, such as a bottle, as it is knocked against a smartphone while SoundCraft⁵⁶ 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 al⁵⁷ used a 4-microphone array to augment wood and glass panels to detect and localize knocking and

⁵¹ Desjardins, A., Psarra, A., & Whiting, B.A. (2021). Voices and Voids: Subverting Voice Assistant Systems through Performative Experiments. *Creativity and Cognition*.

⁵² Brand, N., Odom, W., & Barnett, S. (2021). A Design Inquiry into Introspective AI: Surfacing Opportunities, Issues, and Paradoxes. *Designing Interactive Systems Conference 2021*.

⁵³ Available at: https://bjoernkarmann.dk/project_alias.

⁵⁴ Harrison, C., Xiao, R., & Hudson, S.E. (2012). Acoustic barcodes: passive, durable and inexpensive notched identification tags. *Proceedings of the 25th annual ACM symposium on User interface software and technology*.

⁵⁵ Gong, T., Cho, H., Lee, B., & Lee, S. (2018). Identifying Everyday Objects with a Smartphone Knock. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*.

⁵⁶ Han, T., Hasan, K., Nakamura, K., Gomez, R., & Irani, P. (2017). SoundCraft: Enabling Spatial Interactions on Smartwatches using Hand Generated Acoustics. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*.

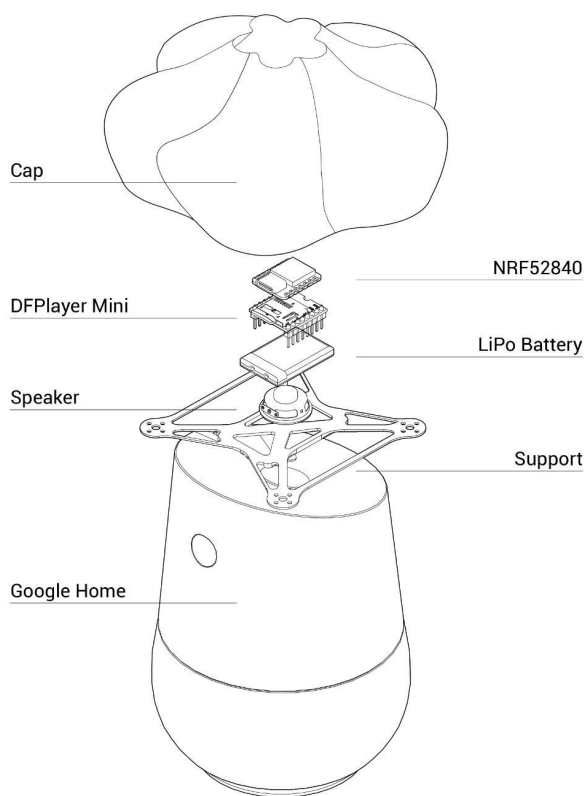
⁵⁷ Jeong, J., Kim, J., Yoon, H., & Jeong, J. (2021). Knock&Tap: Classification and Localization of Knock and Tap Gestures using Deep Sound Transfer Learning. *Companion Publication of the 2021 International Conference on Multimodal Interaction*.

tapping gestures. Kuo & Rawn⁵⁸ demonstrated everyday materials -- namely the sound of velcro ripping apart -- as an interactive input. Kurosawa et al⁵⁹ 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.

Design and Engineering

Within this section, I describe my approach to a “helper” intelligence for the Google Home and how it enables recognition, interpretation, and response to everyday non-verbal audio.

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 the following image.



The physical structure of the “helper”.

⁵⁸ Kuo, T., & Rawn, E. (2020). Let It Rip! Using Velcro for Acoustic Labeling. Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology.

⁵⁹ Kurosawa, T., Shizuki, B., & Tanaka, J. (2013). Keyboard Clawing: Input Method by Clawing Key Tops. HCI.

The “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 TensorFlow⁶⁰, 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.

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.

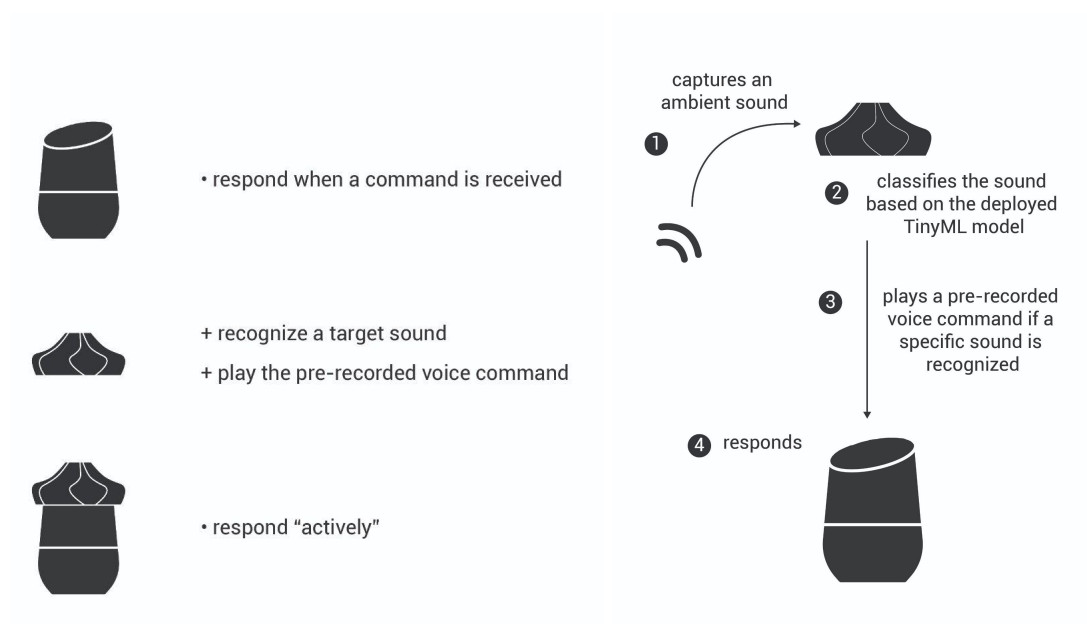
The idea of a trumpet shape leads to the design of the cap. The trumpet shape can be found in many sound art installations, such as Sonic Playground⁶¹ by Yuri Suzuki and Hypothetisches Gebilde⁶² by Alicja Kwade.

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.

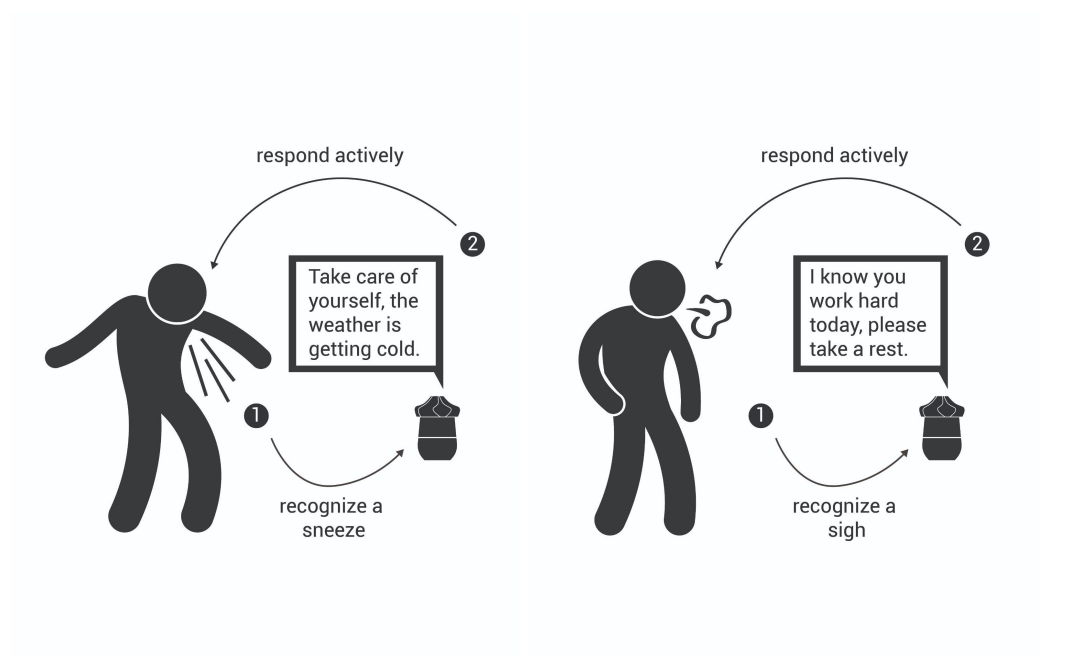
⁶⁰ Available at: <https://www.tensorflow.org/>.

⁶¹ Available at: <https://www.yurisuzuki.com/artist/sonicplayground>.

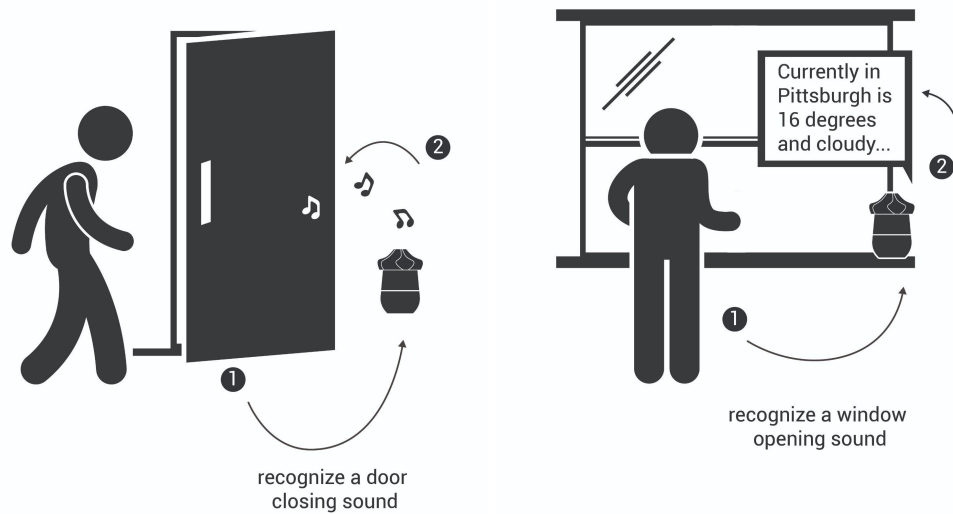
⁶² Available at:
https://alicjakwade.com/works/hypothetisches_gebilde.



The working principle of the "helper".



The "helper" can recognize some human related sounds and responds accordingly.



Other than human centered scenarios, the “helper” can also recognize an environmental ambient sound.
Left: It can recognize a door closing sound by playing some music to welcome anyone entering the space.
Right: It recognize a window opening sound by telling the local weather.

Evaluation and Design Reflections

Design Opportunities

Adding an intelligent helper with broader forms of a sound recognition model to Google Home opens up many opportunities. It encourages us to reconsider our relationship with voice assistants and think of novel and creative encounters with these devices. This proposition – and the main contribution of this final 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. This 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.

Pilot Evaluation

To assess the design concept, I conducted a preliminary evaluation of these opportunities while developing the project. I recruited 4 study participants with different backgrounds from Carnegie Mellon University with an average age 28 years old, two identified as females and two identified as male. I interviewed them and gathered anecdotal evidence for the value of these design opportunities. I introduced them to the project, and how it functions. I also solicited ideas for the non-verbal sound gestures they would like to be recognized and the corresponding responses they would expect. For example, because one participant likes cooking, they explained that they wanted the prototype to recognize the sound of knocking a pot so that Google Home would create a 15-minute timer. Generally, participants enjoyed the playfulness of this project. It suggests many opportunities to gather and co-design interactions with this platform. These suggestions helped to validate the concept and to generate additional scenarios of use presented in the formal evaluation.

Formal Evaluation

After finalizing the “helper” prototype, I evaluated it through a generative workshop.

I conducted a generative workshop with 10 students from the undergraduate program in Carnegie Mellon’s School of Design. All the students are juniors in the Environments track, which focuses on designing interactions in hybrid or physical environments. Participants identified as five females and five males. They were approximately 20 years old. I mainly solicited their responses to the design possibilities for non-verbal sound with the “helper”. Questions and answers were delivered and collected via a Google Form⁶³ survey.

The generative workshop was intended both to get participants’ feedback to evaluate the work and generate other scenarios that the participants could imagine and would like to see enacted based on their interpretation of and response to the “helper.” The survey asked participants about how they would want to use the “helper”, their opinion about using sound as an input for interaction design, and what they would like learn from this project. The survey took about 10 minutes to complete.



The 10 students are watching the video.

⁶³ Available at: <https://forms.gle/6ZqSmRAVGneZQXfE8>.

Process

The workshop was scheduled as a 30-minute long session and its steps were as follows:

1) Everyone watched a one-minute long video⁶⁴ on their screens. It introduced the physical structure of the “helper”, its working principle. Two simple scenarios were shared with the participants to help them get acquainted with the “helper”. These scenarios included the “helper” listening and responding to a table knock sound and to the sound of me sneezing.

2) A Q & A session for participants allowed them to voice any questions they had about what was shown in the video as well as any general questions about the project and the research. I also showed them the helper in person.

3) The Google Form was shared with the participants to collect feedback and ideas about the project (see below).

Survey

The Google Form consists of the following 6 qualitative questions:

Q1: Imagine this helper is in your home. What would you train this helper to do and why? Fill in up to 5 scenarios using the prompts below.

A helper to hear ____ in the situation ____ / when I do ____ / when ____ happens and respond by ____.

e.g. A helper to hear (my sneeze sound) when I (sneeze) and respond by (some comfort words).

Q2: Comparing this to standard voice assistants, what was different and what (if anything) is valuable to this approach?

Q3: Spend a moment listening to the sounds around you right now. Close your eyes and hear the space. What opportunities for designing interactions do you hear in these sounds?

Q4: Thinking about the prototype you have just seen, what affordances of everyday sounds?

⁶⁴ Available at: <https://youtu.be/HqgNliaO3ms>.

Q5: Thinking about the prototype you have just seen, what does this make you think about as a designer?

Q6: Thinking about the prototype you have just seen, how does this change your perspective on designing voice interactions?

The Google Form has the option for allowing response editing, which means that participants can still edit and resubmit the form after their first entry if they come up with some other new ideas or encounter something later in everyday life that helps polish the project.

I did not introduce quantitative questions to evaluate the project since I am not aiming to make a product or assess its utility. score (on a 5 point scale) would not help me to examine the idea of using ambient sound as an input for interaction design. Instead, qualitative questions helped me to better understand what the participants' think of the project, the opportunities they could see for this project, and, importantly, what aspects they would be concerned about when using ambient sound as an input in an everyday phenomena.

Analysis

For Q1 (*Imagine this helper is in your home. What would you train this helper to do and why?*), the 10 participants created a total of 35 scenarios. This greatly expanded the scope of scenarios that I set for the project.

I divided the 35 scenarios into two categories, *user-centered sound listening* and *environmental ambient sound listening*. The following are examples for the two categories respectively:

"When the helper hears me snuffle and cough, the helper can let me know if it's going to be cold out today or tell me to close the window. The helper could also adjust the temperature and make it warmer."

"When the helper hears the laundry machine turn on or start watching, it automatically sets a timer for the amount of time the laundry usually runs."

Some other scenarios seem kind of funny to me but they help me to realize what other people are expecting from the project, and what experiences they want to have with the "helper".

"A helper to hear a drop sound when I dropped something and respond by "oOps"."

"A helper to hear "uhgggg that is so stupid, why did I do that" when I regret sth I did and respond by telling me if it is pattern of mistake and how to improve."

Five years ago, people would think it is funny to ask a voice assistant like a Google Home to order pizza for them. But right now, people are already accustomed to this kind of interaction and even want it to do more for them, such as home automation. I think these two "funny" scenarios will likely become a reality when people want to build a deeper relationship with their voice assistants and realize the "local culture" that they build with their family is valuable.

For Q2 (*Comparing this to standard voice assistants, what was different and what (if anything) is valuable to this approach?*), the 10 participants contributed 8 responses.

Most of the responses mentioned that the "helper" builds a playful or an intimate relationship with the space/environment. Here is an example:

"I liked that this interaction is driven by sounds that naturally occur in the space as opposed to voice prompts; it feels more integrated into the space."

On the contrary, one participant addressed the concern of user's privacy and the possible tensions in everyday sound being monitored by the helper:

"It seems both more and less intrusive at the same time. Less of an intrusive presence within the home since you aren't talking to it all the time, so that you aren't being constantly reminded of the device. It might be easier to forget about the presence of something like this at home. However, it also seems somewhat subversive in the way that it picks up on other verbal cues, and may make me feel as if I am being watched in ways I don't want to be watched."

Although, this project does not focus explicitly on issues of privacy or domestic surveillance, this response makes me think about some larger questions: How can we shape our relationships with AI? Except or other than the cybernetic loop⁶⁵ (sensor, controller, another system), what should we add or reduce to build a "reliable" ecosystem with AI? Do we have to lose or give up anything when living with the system? We have to teach the machine or the AI (feed them with a super large amounts of data) so that they can learn and understand human needs and perform properly. Can we live with a "malfunctioning" machine or AI if it is trained by ourselves and it

⁶⁵ Available at: <https://en.wikipedia.org/wiki/Cybernetics>.

learns by mistakes? I do not have any solid ideas for these questions and I do not know if a machine or AI could be cultivated in the same way as plants and animals that humans have built relationships with since several thousand years ago. But I think the way we teach the machine reflects what the relationship we want to shape with it.

For Q3 (*Spend a moment listening to the sounds around you right now. Close your eyes and hear the space. What opportunities for designing interactions do you hear in these sounds?*), the 10 participants contributed 9 responses.

They all found some sounds that existed in the environment and some of them mentioned that there are a lot of opportunities for them to be aware of their habits/behaviors and the space via sound. Here is an example of that:

“Sound holds a lot of potential for recognizing the presence of different people and how they interact with things/other people in their spaces. There is a lot of opportunity for heightening people's own awareness of their own habits/behaviors and becoming more mindful within their space.”

For Q4 (*Thinking about the prototype you have just seen, what affordances of everyday sounds?*), the 10 participants contributed 4 responses.

One response is “no”, and the other three responses mention the meaning of a sound and the experience that associates with a sound. Here are two examples of that:

“I think oftentimes we think of sound as a way of having feedback from the world around us. We listen to understand what is happening around us without having to place our full attention on a particular scene. so these affordances of everyday sounds could be soft nudges for the users that is meaningful to them.”

“I think these affordances can take advantage of everyday sounds and give subtle responses that heighten user's experience.”

Don Norman defines the word affordance⁶⁶ as follows:

“The term affordance refers to the relationship between a physical object and a person (or for that matter, any interacting agent, whether animal or human, or even

⁶⁶ Norman, D.A. (2013). The Design of Everyday Things: Revised and Expanded Edition.

machines and robots). An affordance is a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used.”

Patrick Thornton further explains the concept of affordance in Medium⁶⁷:

“Affordances can both be affordances (to help us do something) or anti-affordances (to prevent us from doing something). They can also be visible and invisible, but for an affordance to be effective, people have to be able to perceive it.”

I would assume that participants are either not familiar with the concept of affordance or it takes a longer time for them to think about the answer since this question has the least entries out of 6 questions. To me, sound affords a piece of information that people can retrieve from it. It enables me to know what is happening and what I should base on the sound I hear. From my perspective, I have been trained on a given sound over time and understand the information behind it directly, but other people may get totally different experiences from the same sound since we each have a unique life experience. If this is the case, I cannot help thinking of some questions: How can we define a “global” or a “local” affordance of sound in general? Can we simplify or complicate the information behind a sound?

For Q5 (*Thinking about the prototype you have just seen, what does this make you think about as a designer?*), the 10 participants contributed 8 responses.

One participant questions about how to opt-out the ecosystem:

“It is a unique and valuable approach to designing for smart homes, but makes me question the opt-out factor of using the product, especially if there are times where I want to do certain actions but not have the google home respond.”

I have not thought about this kind of question before, but I would assume there will be a designated sound that users can make or get triggered from the environment to close the system if this project develops further.

Some other responses pick up some points that I also never think of. Here are two examples:

⁶⁷ Available at:
<https://uxdesign.cc/what-is-an-affordance-6b60f2de79f2>.

"It makes me think about how we can design interactions that is helpful to the user but not overwhelming. I think often times when I interact with UI designs there are a lot of notifications coming from an app in hopes of being useful to the user but end up becoming distracting. I hope the helper is actually needed and memorable."

"It makes me think about the arbitrary/normal actions that we may not notice right away, but that this prototype is able to pick up on. I think this is great for building a relationship with spaces, but also starting to notice our own tendencies and habits that we might not have noticed before. I wonder how this cause behavior change?"

These two responses reminds me of calm technology⁶⁸, and what Mark Weiser said about designing calm technology⁶⁹:

"As we learn to design calm technology, we will enrich not only our space of artifacts, but also our opportunities for being with other people."

When developing this project, sometimes I was thinking about the "uncanny" side of a sound and what the "uncanniness" of a sound would bring to people. Some of the scenarios that I imagined are intentional and sometimes there is a sense of surprise behind them. I do think a more natural setting for the design and interaction would help this kind of project be accepted in a wider scope of an environment. A question that comes after this is: Will people benefit from a more generic system or a more personalized one?

For Q6 (*Thinking about the prototype you have just seen, how does this change your perspective on designing voice interactions?*), the 10 participants contributed 8 responses.

Most of the responses share a similar idea that it is good to see a fluid integration of a certain technology in everyday life in an unobtrusive way. Here is an example of that:

"It makes me think more about all noises, not necessarily a voice or command that is often associated with Alexa, Siri, etc. This prototype opens up the possibilities of how sound can transform our relationship with spaces and ourselves."

One participant shows a sense of concern in the response:

⁶⁸ Case, A. (2015). Calm Technology: Principles and Patterns for Non-Intrusive Design. China: O'Reilly Media.

⁶⁹ Weiser, M.D. (2004). Designing Calm Technology.

"It opens up a lot more possibilities about designing voice interactions that could be more natural than using a wake word (ex. "Hey Google"), and make for more natural conversation starters. However, getting too comfortable with only using these learned sounds to trigger interactions might seem creepy and intrusive to a person's daily life, and might turn a person off from the idea of interacting with the voice assistant."

I agree with the participant that this project makes a more natural conversation starter with a voice assistant in some situations. But I do not think this will prevent people from interacting with a voice assistant since a conversation should be consequent with a flow not separated prompts. The intention behind this project is to enrich certain experiences via ambient sound but not about how to create different sounds. I do not think people will stand up from their seat and go to close the door to ask a Google Home to play some music for them (as one example of my scenario is that when people come to a room and close the door, the Google Home plays some music. The intention is different here.) instead of directly asking it to play some music for them.

The survey provided me a good opportunity to revise my idea about the project and understand how other people think of it and want from it.

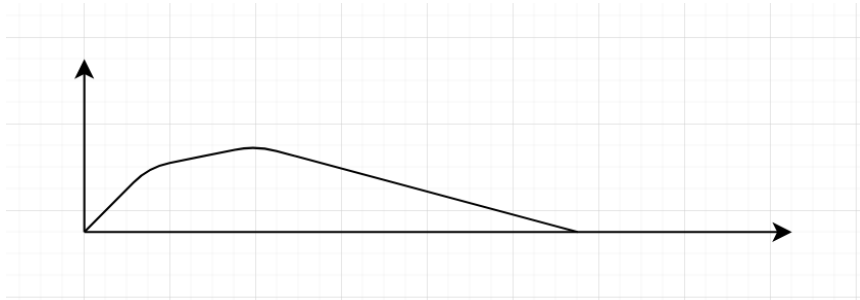
Discussion

Based on conducting the generative workshop and analyzing the responses from the participants, below are some takeaways for designers and researchers who may share the similar research interests with me in non-verbal/ambient sound.

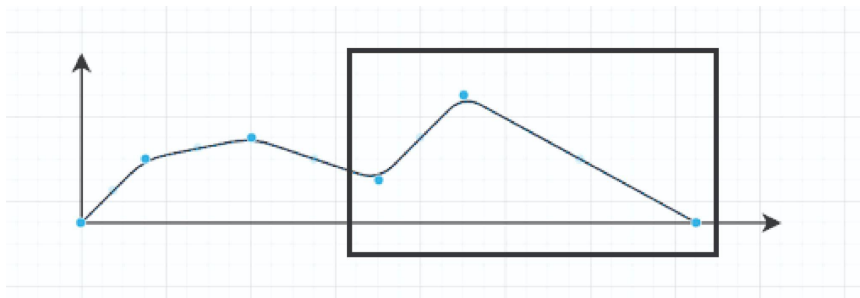
Ambient sound could be leveraged much more as an input for interaction design. It has great potential when integrated with TinyML, especially if the sound samples have been carefully selected and processed.

Ambient sound is not obtrusive, it blends well with the environment. It does not require new efforts to put on the environment or to add new elements to the environment. It is there, perceivable but oblivious. I recommend other designers, especially when processing the sound samples, pay attention to carefully crafting out any parts that might interfere with the recognition process. I'll illustrate why this is important next. As an example, if the target sound is a window opening sound, the other environmental sound should be eliminated from the samples collected, otherwise, it will increase the possibility

of misrecognition of the system later in the process. I am still thinking about better ways to classify sound samples if two sounds are quite similar in their waveform. For example, a tap sound will often be mis-recognized as a scratch-in sound in my experience within the speculative experiments (e.g. the lampshade). As shown in the following images, the amplitude of tap sound against time is very close to the second half of the amplitude of scratch-in sound against time. I am not sure this is the case, but accuracy drops when classifying these two types of sounds.



The amplitude of tap sound against time.



The amplitude of scratch-in sound against time.

Nevertheless, TinyML is an efficient and powerful tool for rapidly producing edge applications for microcontrollers. Without the help of TinyML, the final project would not have been as small in size as the “helper”. It is so energy efficient that it allows edge ML applications to work for a long time without recharging the battery. Although the final project does not focus on privacy, TinyML reduces the chance of information leakage since all the sound is collected, processed and managed locally for recognition, in my case.

Besides these two empirical takeaways I learned from conducting my thesis, there are some unmatured thoughts I come up with when reviewing the final project and the whole thesis.

First of all, I think there is a long history of humans cultivating “things”, such as a plant, a poultry animal, or a pet, the relationship between the human and the cultivated “thing” greatly depends on the mode of cultivation. For poultry or livestock, humans feed it but not

live with it. For a pet, humans even take care of every aspect of it. How do we and how should we “cultivate” a voice assistant (an AI in general)? What do we want from them? What is the relationship we want to shape with it? I have no answer for any of these questions, but I think any perspective will lead to a unique situation, a design opportunity, and a relationship.

Beyond this, I also can not help but think about the question: *what is the cultural identity of a sound?* To me, there are different ways to answer this question. On the “global” scale, I think a cultural identity is acknowledged by a large group of people living in the same region. One simple example is a sneeze. In some Western cultures, when hearing a sneeze, people would say “bless you” without a second thought. While in some Eastern (Oriental) cultures, when sneeze is heard, people imagine that some beloved ones are missing them. Another example is the sound of opening a window. In some Western cultures, people are used to lifting the window to make it open. While in some Eastern (Oriental) cultures, people are used to pushing the window to make it open. The resulting sound can be very different from these two actions, not to mention the materials used to make the window frame which also lead to different sounds. This example can also be used to involve a “local” culture in designed interactions. The different living conditions, from my perspective, are different “local” cultures. Even people can create a very own “local” culture with their partner and thus the meaning and identity behind a sound can vary.

It is interesting to regard a voice assistant as a universal device, a Western mass-produced device. What can we do to add our own “local” culture to it? I think the customization process is a way of enhancing a ritual, a routine, and a “local” culture.

One idea behind the final project, during the time when I was making it, was to make me aware of the “uncanniness” of an everyday sound. I believe it is the “uncanniness” of a thing (a sound in my case) that bring people from reality into imagination. It can be surprising sometimes. While one response from the survey values Weiser’s calm technology that seamlessly blends with and embeds into the environment. I think these two attitudes can be two kind of rules that we set for our own “local” cultures.

Finally, this thesis began with a discussion of the cortical homunculus as a metaphor of showing what body parts are capable of and how great they are. To me, most people are born with the ability to hear, and a bunch of information can be retrieved from a sound. I think, not limited to sound, if we observe more carefully (use different sensory organs), we can find a larger and wider world for designed interactions. We are not lacking information in the world, but merely ways to retrieve and respond to information.

Conclusion

This thesis presents an investigation into leveraging ambient sound in interaction design by using it as an input for voice assistant devices instead of voice prompts. It focuses on how to shape our relationships with the space we inhabit and the voice assistant devices that occupy our homes

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 through sound.

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, I 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 (and thereby the smart home too). 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 is programmed by a group of people.

Limitations, and Future Work

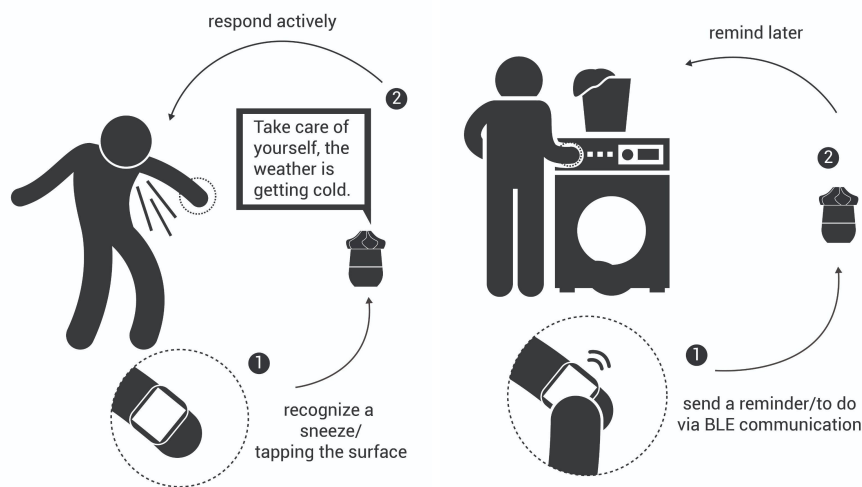
As such, I envision *Google Home, Listen* as a resource for future design inquiry that will explore the affordances of everyday sounds in augmenting the capacities of currently voice-enabled intelligences.

As part of future work, I plan to refine and scale the generative workshop format to better explore these possibilities. I envision a series of generative workshops that use the helper as a generative design tool to investigate our relationship to smart home automation. I also plan to deploy the platform to homes to examine its affordances in rethinking how we encounter devices that occupy our domestic spaces. It could help enabling designers explore, understand, and imagine the potential of environmental and contextual cues in future scenarios of interactive experiences.

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 to improve the 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 Knocker⁷⁰, SoundCraft⁷¹, and Knock&Tap⁷², 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.



There are many opportunities for further implementations.

⁷⁰ Gong, T., Cho, H., Lee, B., & Lee, S. (2018). Identifying Everyday Objects with a Smartphone Knock. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.

⁷¹ Han, T., Hasan, K., Nakamura, K., Gomez, R., & Irani, P. (2017). SoundCraft: Enabling Spatial Interactions on Smartwatches using Hand Generated Acoustics. Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.

⁷² Jeong, J., Kim, J., Yoon, H., & Jeong, J. (2021). Knock&Tap: Classification and Localization of Knock and Tap Gestures using Deep Sound Transfer Learning. Companion Publication of the 2021 International Conference on Multimodal Interaction.

Bibliography

1. Crary, J. (2001). *Suspensions of Perception: Attention, Spectacle, and Modern Culture*.
2. Cortical homunculus. (2021). In Wikipedia.
https://en.wikipedia.org/w/index.php?title=Cortical_homunculus&oldid=1040682309
3. Weiser, M.D. (1991). The computer for the 21st century. *Human-Computer Interaction*, 933-940.
4. Lenzi, S., & Ciuccarelli, P. (2020). Intentionality and design in the data sonification of social issues. *Big Data & Society*, 7.
5. The Quietus | Features | Tome On The Range | The Ear Of The Other: Colonialism & Decolonial Listening. (n.d.). The Quietus. Retrieved February 10, 2022, from <https://thequietus.com/articles/29445-sound-colonialism-and-decolonial-listening-focus-on-sound-emily-hansell-clark>
6. Bell, G., & Kaye, J. (2002). Designing Technology for Domestic Spaces: A Kitchen Manifesto. *Gastronomica*, 2, 46-62.
7. Suchman, L.A. (2011). Anthropological Relocations and the Limits of Design. *Annual Review of Anthropology*, 40, 1-18.
8. अहमद अंसारी, (2020) احمد انصاري, August 11). White supremacy & epistemic colonialism in design discourse, scholarship, & practice: A basic primer (thread) [Tweet]. @aansari86.
<https://twitter.com/aansari86/status/1293262346761179137>
9. Leimma. Retrieved May 11, 2022, from <https://isartum.net/leimma>
10. Apotome. Retrieved May 11, 2022, from <https://isartum.net/apotome>
11. Harrison, C., & Hudson, S.E. (2008). Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. *UIST '08*.
12. Laput, G., Brockmeyer, E., Hudson, S.E., & Harrison, C. (2015). Acoustruments: Passive, Acoustically-Driven, Interactive Controls for Handheld Devices. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*.
13. Harrison, C., Xiao, R., & Hudson, S.E. (2012). Acoustic barcodes: passive, durable and inexpensive notched identification

tags. Proceedings of the 25th annual ACM symposium on User interface software and technology.

14. Hypersurfaces. Hypersurfaces. Retrieved May 11, 2022, from <https://www.hypersurfaces.com>

15. Das, S., Laput, G., Harrison, C., & Hong, J.I. (2017). Thumprint: Socially-Inclusive Local Group Authentication Through Shared Secret Knocks. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.

16. Asakawa, S., Guerreiro, J.P., Sato, D., Takagi, H., Ahmetovic, D., Gonzalez, D., Kitani, K.M., & Asakawa, C. (2019). An Independent and Interactive Museum Experience for Blind People. Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web.

17. Zhao, Y., Bennett, C.L., Benko, H., Cutrell, E., Holz, C., Morris, M.R., & Sinclair, M. (2018). Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.

18. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R., Miller, R., Tatarowicz, A., White, B.A., White, S., & Yeh, T. (2010). VizWiz: nearly real-time answers to visual questions. Proceedings of the 23rd annual ACM symposium on User interface software and technology.

19. Portocarrero, E., Dublon, G., Paradiso, J.A., & Bove, V.M. (2015). ListenTree: Audio-Haptic Display In The Natural Environment. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems

20. Perini, A. (2017, January 17). Sounding canebrake. Alessandro Perini. <https://alessandroperini.com/portfolio/sounding-canebrake/>

21. Soundwalk. (2021). In Wikipedia. <https://en.wikipedia.org/w/index.php?title=Soundwalk&oldid=1045141226>

22. Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016). Thing Ethnography: Doing Design Research with Non-Humans. Proceedings of the 2016 ACM Conference on Designing Interactive Systems.

23. LAUREN — Lauren Lee McCarthy. Retrieved May 11, 2022, from <https://lauren-mccarthy.com/LAUREN>

24. Bjørn Karmann › project_alias. Retrieved May 11, 2022, from https://bjoernkarmann.dk/project_alias

25. Nabaztag. (2022). In Wikipedia.
<https://en.wikipedia.org/w/index.php?title=Nabaztag&oldid=1070029178>
26. Allen Pan. (2017, March 17). Zelda Ocarina Controlled Home Automation - Zelda: Ocarina of Time | Sufficiently Advanced.
<https://www.youtube.com/watch?v=glZnkpIDWSE>
27. Karana, E., Giaccardi, E., Stamhuis, N., & Goossensen, J. (2016). The Tuning of Materials: A Designer's Journey. Proceedings of the 2016 ACM Conference on Designing Interactive Systems.
28. Teachable Machine. Retrieved May 11, 2022, from
<https://teachablemachine.withgoogle.com/>
29. Processing. Retrieved May 11, 2022, from
<https://processing.org/>
30. Scikit-learn: Machine learning in Python—Scikit-learn 1.0.2 documentation. Retrieved May 11, 2022, from
<https://scikit-learn.org/stable/>
31. Microsoft. (2018, August 28). Matthew Bennett: How I composed the Windows 10 calendar alert | Microsoft Story Labs.
<https://www.youtube.com/watch?v=1sqg63imHx0>
32. Home | p5.js. Retrieved May 11, 2022, from <https://p5js.org/>
33. Arduino—Home. Retrieved May 11, 2022, from
<https://www.arduino.cc/>
34. Lab: Serial output from P5.js – ITP Physical Computing. (n.d.). Retrieved May 11, 2022, from
<https://itp.nyu.edu/physcomp/labs/labs-serial-communication/lab-serial-output-from-p5-js/>
35. Zheng, C., Oh, H., Devendorf, L., & Do, E.Y. (2019). Sensing Kirigami. Proceedings of the 2019 on Designing Interactive Systems Conference.
36. Strickland, R. (2021, August 27). FINGER FILM (1976).
<https://vimeo.com/593396130>
37. The Sound of Clouds and other Sonic Memories |. Retrieved December 13, 2021, from
<https://www.arkfrequencies.com/the-sound-of-clouds-and-other-sonic-memories/>

38. Cochlear Implant Research Laboratory. (2014, October 20). What a Cochlear Implant sounds like.
<https://www.youtube.com/watch?v=1dhTWVMcpC4>
39. Porcheron, M., Fischer, J.E., Reeves, S., & Sharples, S.C. (2018). Voice Interfaces in Everyday Life. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
40. Beneteau, E., Richards, O.K., Zhang, M., Kientz, J.A., Yip, J.C., & Hiniker, A. (2019). Communication Breakdowns Between Families and Alexa. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
41. Sciuto, A., Saini, A., Forlizzi, J., & 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.
42. Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening? Proceedings of the ACM on Human-Computer Interaction, 2, 1 - 31.
43. Ammari, T., Kaye, J., Tsai, J.Y., & Bentley, F. (2019). Music, Search, and IoT. ACM Transactions on Computer-Human Interaction (TOCHI), 26, 1 - 28.
44. Sutton, S.J., Foulkes, P., Kirk, D., & Lawson, S.W. (2019). Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
45. Desjardins, A., Psarra, A., & Whiting, B.A. (2021). Voices and Voids: Subverting Voice Assistant Systems through Performative Experiments. Creativity and Cognition.
46. Brand, N., Odom, W., & Barnett, S. (2021). A Design Inquiry into Introspective AI: Surfacing Opportunities, Issues, and Paradoxes. Designing Interactive Systems Conference 2021.
47. Gong, T., Cho, H., Lee, B., & Lee, S. (2018). Identifying Everyday Objects with a Smartphone Knock. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.
48. Han, T., Hasan, K., Nakamura, K., Gomez, R., & Irani, P. (2017). SoundCraft: Enabling Spatial Interactions on Smartwatches using Hand Generated Acoustics. Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.
49. Jeong, J., Kim, J., Yoon, H., & Jeong, J. (2021). Knock&Tap: Classification and Localization of Knock and Tap Gestures using

Deep Sound Transfer Learning. Companion Publication of the 2021 International Conference on Multimodal Interaction.

50. Kuo, T., & Rawn, E. (2020). Let It Rip! Using Velcro for Acoustic Labeling. Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology.

51. Kurosawa, T., Shizuki, B., & Tanaka, J. (2013). Keyboard Clawing: Input Method by Clawing Key Tops. HCl.

52. TensorFlow. TensorFlow. Retrieved May 11, 2022, from <https://www.tensorflow.org/>

53. Yuri Suzuki | Sonic Playground. Retrieved May 11, 2022, from <https://www.yurisuzuki.com/artist/sonicplayground>

54. Alicja Kwade | Hypothetisches Gebilde. Retrieved May 11, 2022, from https://alicjakwade.com/works/hypothetisches_gebilde

55. Cybernetics. (2022). In Wikipedia.
<https://en.wikipedia.org/w/index.php?title=Cybernetics&oldid=1087235475>

56. Norman, D.A. (2013). The Design of Everyday Things: Revised and Expanded Edition.

57. What is an affordance?. Understanding what affordances are and... | by Patrick Thornton | UX Collective. Retrieved May 11, 2022, from <https://uxdesign.cc/what-is-an-affordance-6b60f2de79f2>

58. Case, A. (2015). Calm Technology: Principles and Patterns for Non-Intrusive Design. China: O'Reilly Media.

59. Weiser, M.D. (2004). Designing Calm Technology.