

# Human-Centered Sound Recognition Tools for Deaf and Hard of Hearing People

Steven M. Goodman

*Dissertation Proposal*

April 18, 2023

UNIVERSITY *of*  
WASHINGTON

**W**  
HUMAN CENTERED  
DESIGN & ENGINEERING

## Introduction

# Sound carries rich information about the world around us



## Introduction

# Sound carries rich information about the world around us



**But it may not be accessible to people who are Deaf and hard of hearing**

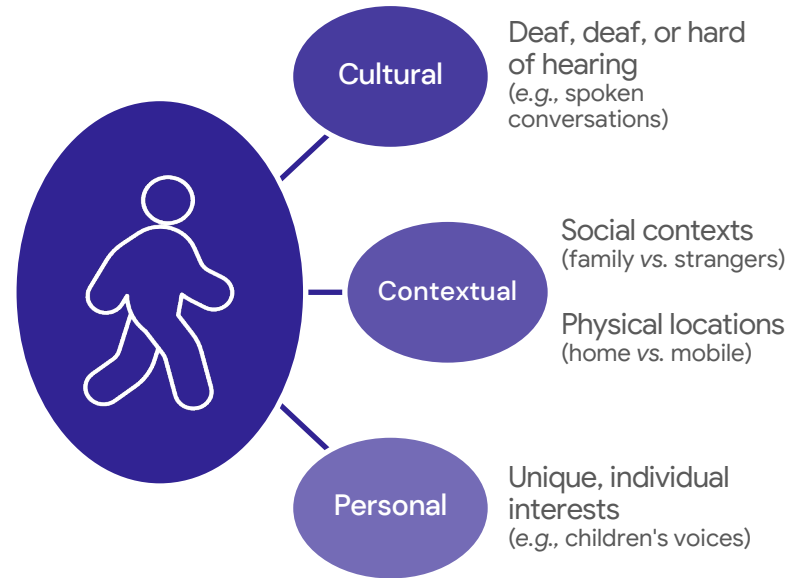
# Sound interests of DHH individuals

Deaf and hard of hearing (DHH) people are interested in sound awareness technologies.

But different factors influence sound preferences among DHH individuals.

A "one-size-fits-all" sound awareness solution is not tenable.

**Personalized tools are necessary to meet individual needs.**



## Introduction

# Current technologies

Android & iOS include automatic sound recognition models.

Both use a pre-trained model supporting ~15 sound classes.

- E.g., appliances, alarms, pets

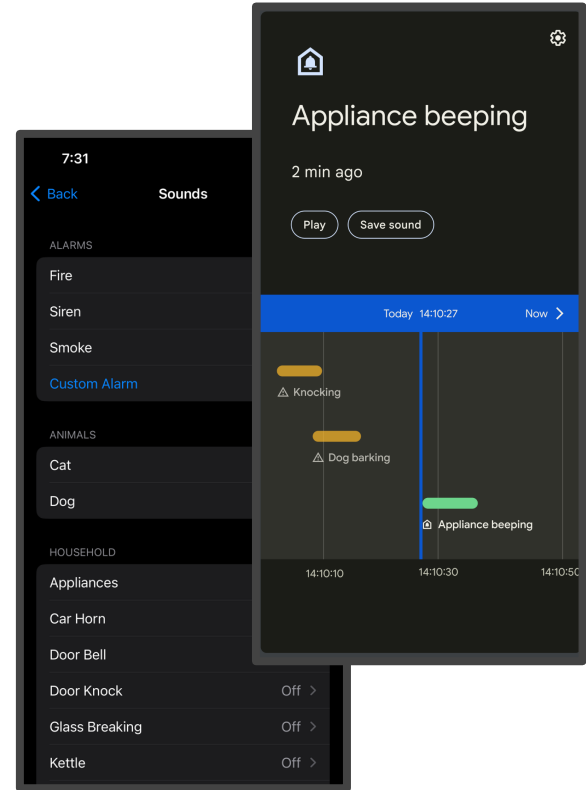
However, these sound categories are generic:

- They do not adapt to varied sound environments
- They do not account for **edge cases**.



A survey of DHH Android users revealed **dissatisfaction with accuracy and flexibility**, and desire to **personalize a sound recognition model**.

[Jain et al., CHI 2022]



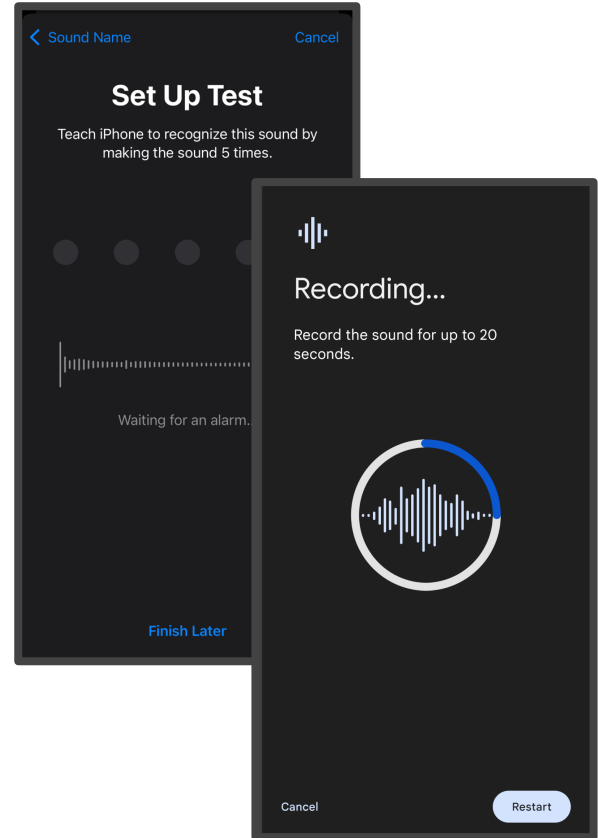
## Introduction

# Progress towards personalization?

Users can filter alerts and extend the pre-trained model with their own recordings.

- iOS: fine-tuning existing categories
- Android: adding custom sound categories

The “AutoML” approach is fast and easy but **lacks transparency and control**—and could reduce trust and long-term use among users.



# ***Interactive Machine Learning for DHH users***

DHH users interacting with machine learning can lead to automatic tools specialized to their wide-ranging needs.

**Interactive machine learning** (IML) systems provide an understanding of the model's strengths and limitations, fostering trust and transparency.

*[Sanchez, CSCW 2021]*

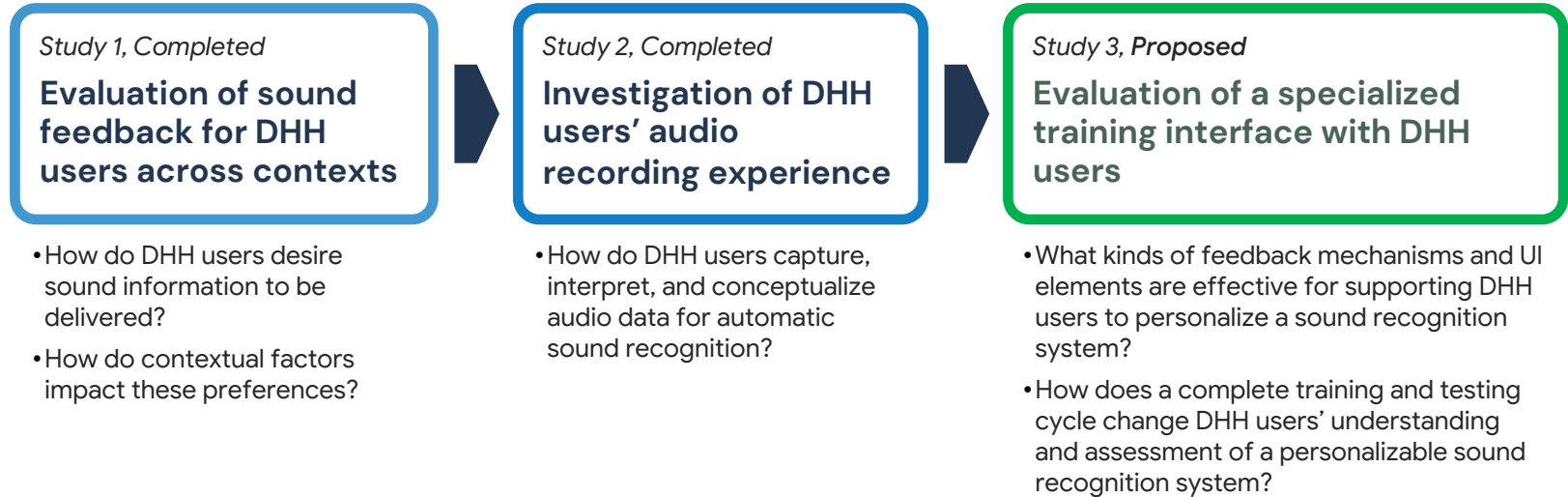
But IML literature assumes end-users have **domain knowledge** and can **access a model's underlying data**.

*[Dudley et al., 2018]*

*How can a DHH user, who has difficulty hearing a sound themselves, train a machine learning model to recognize that sound?*



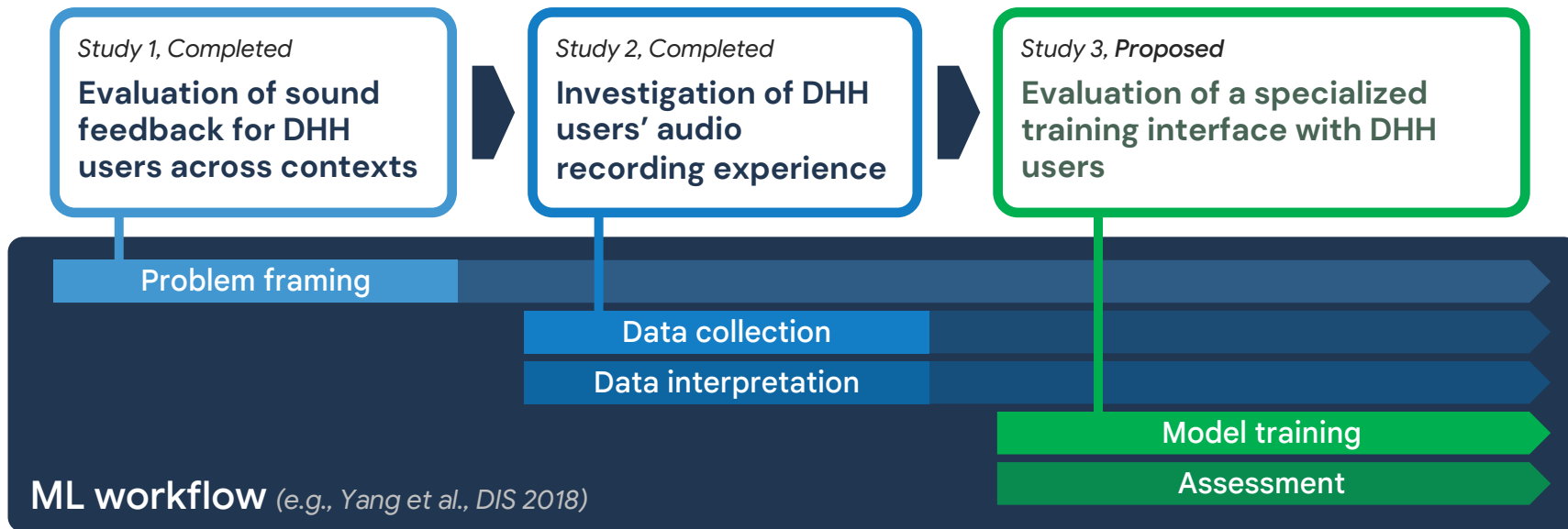
# My dissertation work



**GOAL:** A framework for supporting Deaf and hard of hearing individuals to personalize sound recognition tools that meet their everyday needs.

## Overview

# My dissertation work



**GOAL:** A framework for supporting Deaf and hard of hearing individuals to personalize sound recognition tools that meet their everyday needs.

## Overview

# My dissertation work



**GOAL:** A framework for supporting Deaf and hard of hearing individuals to personalize sound recognition tools that meet their everyday needs.

# Evaluating Smartwatch-based Sound Feedback for Deaf and Hard-of-Hearing Users Across Contexts

CHI 2020

75403

204-552-3000  
www.kingcounty.gov/metro



King County  
METRO

**Steven Goodman**

Susanne Kirchner-Adelhardt

Rose Guttman

Dhruv Jain

Jon Froehlich

Leah Findlater

UNIVERSITY of  
WASHINGTON

**W**  
HUMAN CENTERED  
DESIGN & ENGINEERING






## Motivation

# Sound awareness preferences

Prior work highlights general preferences among DHH users.

The most important sounds are:

1. Safety-related 
2. Indicators of others' presence 
3. Contextual alerts 

For sound awareness technology:

- Tools should be **portable** for use in a variety of contexts.
- Users desire sound feedback through **visual and haptic** modalities.
- Unimportant sounds should be **filtered out** from incoming feedback.

Motivation

## Findlater et al., CHI 2019

Survey of 201 DHH participants:

**Smartwatches** are the preferred portable device for sound awareness

- Useful, socially acceptable, and glanceable
- Provides both **haptic and visual feedback**



# Unknowns for feedback & filtering

Research on using smartwatches for sound awareness is limited to **brief lab-based** study with six participants (Mielke & Brück, 2015).

- The best method for **combining visual and haptic feedback** on a smartwatch remains an open question.

The importance of sounds can vary based on one's social context and physical location.

- Portable tools need to be adaptable to these changes, as **filtering preferences** might change as users move through different contexts.

# Research Questions

How do DHH users desire sound information to be delivered, and how do contextual factors impact these preferences?

- What are effective methods of combining visual and haptic sound feedback on a smartwatch?
- How should sound filtering be designed, and what are the implications for filtering when both visual and haptic feedback is present?



# Method

Single-session study employing **design probe** methodology with 16 Deaf and Hard of Hearing participants

- Average age: 56 years old (SD=17.7, range=19-85)
- Choice of ASL interpreter (n=6) or real-time captioner (n=2)

## Method

# Study Procedure, Part 1



### Lab-based Design Probe (30 min)

- Wizard-of-Oz evaluation

- A quiet lab setting to demonstrate how a watch could sense and convey sounds.
- Three sounds produced: door knock, phone ring, name call
- Visual feedback designed with high-contrast, glanceable aesthetic to convey sound direction, identity, and loudness
- Two haptic designs used: single vibration to notify sound occurrence, and tacton to convey sound direction, loudness, or identity.

## Method

# Study Procedure, Part 2



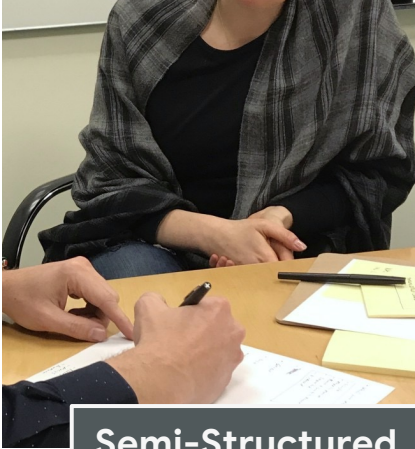
### Contextual Design Probe (25 min)

- Exploration of campus locations

- Contextual exploration of sound feedback and filtering options at three locations on campus
- Participants used an iPad map to orient themselves to a preset sound scene at each location
- Wizard triggered the watch to give visual feedback for a sequence of 10 sounds
- To demonstrate filtering sounds based on different criteria, to three of the sounds based on direction, identity, and loudness

## Method

# Study Procedure, Part 3



### **Semi-Structured Interview (20 min)**

- Reflection on overall experience

- I asked about participants' experience in the lab and around campus
- Other questions probed for insight on:
  - contextual factors
  - filtering criteria
  - social acceptability
  - privacy concerns

Key finding:

**Visual and haptic feedback  
have complementary roles in  
sound awareness.**

# Complementary Modalities

Overall, **participants responded positively** to the idea of smartwatch-based sound feedback.

Participants desired visual feedback across all conditions:

- *“It's nice to have visual and the sensory input as well [but] I mean **without the visual, I feel like there's not really a point.**” (P10)*

Designs with vibration were more useful than without:

- For example, most participants ( $n=13/16$ ) were concerned **they would miss sounds** without vibration

# Complementary Modalities

Past work shows deaf and hard of hearing people make strong use of **visual cues for environmental awareness** [Matthews 2006]

Haptic feedback (simple or tactons) gets a DHH user's attention **without interfering** with visual awareness strategies:

- The user can **respond to the environment** immediately
- Or turn to the watch's screen for **more information**

Key finding:

**Complex soundscapes present awareness issues that may be mitigated through sound filtering.**



# Soundscape Filtering

Following our visits to different locations on campus, most participants ( $n=11/16$ ) mentioned **new use cases or increased interest** in watch-based sound awareness

This often pertained to use **complex soundscapes**: areas with frequent, overlapping sound events

- **Experienced in the café and bus stop**

## Soundscape Filtering

P14 returned feeling far more positive about the idea:

*“ [The café’s] the thing that really **gives people anxiety.***

*“Are they going to hear me? Am I going to hear them?”*

*There's so much ambient noise.*

*In a place like [the student lounge] or in your house with the microwave and whatever, okay, it's quiet.*

*But when you go to a place outside, bus stop, [café], outside your home, and again in your car, **this is just incredible.**”*

## Soundscape Filtering

Quotes like P14's highlight **the challenges, and necessity**, of sound awareness in complex soundscapes.

- **All participants in the study desired filtering** due to exposure to realistically complex soundscapes.

Filtering sounds, rather than showing more, may lead to enhanced awareness in these contexts.

## Soundscape Filtering

Questions arose over whether to trust the system making automatic filtering decisions.

*“[It] might be filtering out other awareness that you have built up over the years in favor of, ‘Well, this thing knows, and in fact this thing might know better than me, so I’m just gonna ignore my instinct, I’m not going to bother looking because this will tell me.’ [...] I want to hear it all, and I want my own, I want to be able to choose what’s more important.” (P4)*

Most participants desired choosing sounds themselves over automatic filtering.

# Outcomes



Jain et al. (2020) built a smartwatch-based sound recognition app for DHH people.

- Trained for 20 sounds
- Included filtering for individual sounds.

Evaluation: DHH participants found the app useful but enabled only a fraction of the sounds at different locations.

- They also requested custom sound categories.

Filtering notifications within pre-trained models is a nice step towards personalization, but...

# Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing

*IMWUT 2021*

**Steven Goodman**

Ping Liu

Dhruv Jain

Emma J. McDonnell

Jon E. Froehlich

Leah Findlater

UNIVERSITY of  
WASHINGTON

**W**  
HUMAN CENTERED  
DESIGN & ENGINEERING

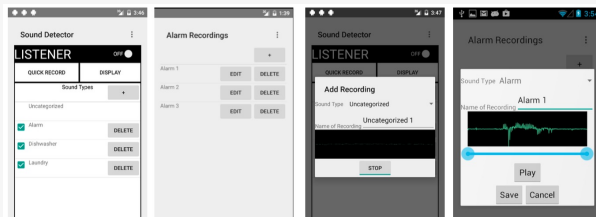


# Motivation

Enabling personalization would benefit DHH users, but systems that augment sensory abilities present challenges for users with sensory disabilities.

*Kacorri et al., SIGACCESS 2017*

Two studies explored personalizable sound recognition tools with DHH participants:



*Bragg et al., ASSETS 2016*



*Nakao et al., NordiCHI 2020*

**How DHH users record and engage with audio data is absent**—despite this data predicating the effectiveness of a sound recognizer.

# Research Questions

**How do DHH users capture, interpret, and conceptualize audio data for the purpose of automatic sound recognition?**

- What considerations do DHH people make when recording in environments with **real-world acoustic variation**?
- What kinds of features can aid DHH users in **assessing their recorded samples** as training data?



# Study Method

## 14 DHH participants

avg. 43.3 years old ( $SD=21.3$ , range=19-87)

- Demonstrate spectrograms and waveforms
- Introduce ML workflow via **Google's Teachable Machine**
  - Record claps, paper, background noise
  - Train and test
- Discuss quality of audio data

### Introductory Session (75 min)

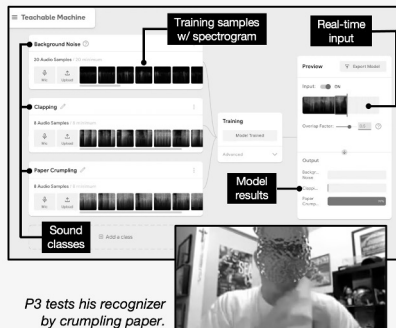
- Introduce recording for sound recognition

The screenshot displays the Teachable Machine interface. At the top, a video titled "Baby crying during a thunderstorm" shows a baby in a white onesie with a yellow duck pattern. To the right of the video are two audio analysis visualizations: a red spectrogram and a green waveform. Below the video, the interface is divided into three sections for audio data collection: "Background Noise" (20 Audio Samples / 20 minimum), "Clapping" (8 Audio Samples / 8 minimum), and "Paper Crumpling" (8 Audio Samples / 8 minimum). Each section has "Mic" and "Upload" buttons and a row of spectrogram thumbnails. On the right side, the "Training" section shows a "Model Trained" button and an "Advanced" dropdown. Below that is a "Preview" section with an "Export Model" button, an "Input" toggle set to "ON", and an "Overlap Factor" slider set to 0.5. At the bottom right, the "Output" section shows three colored bars representing the model's output for "Backgr... Noise" (orange), "Clappi..." (red), and "Paper Crump..." (purple).

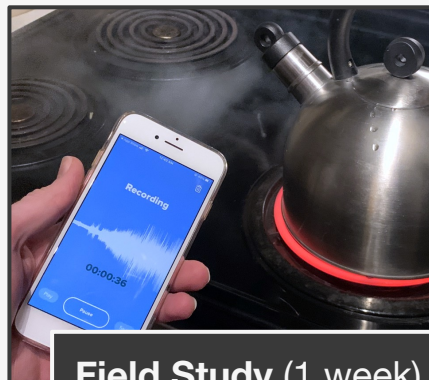
# Study Method

## 14 DHH participants

avg. 43.3 years old ( $SD=21.3$ , range=19-87)

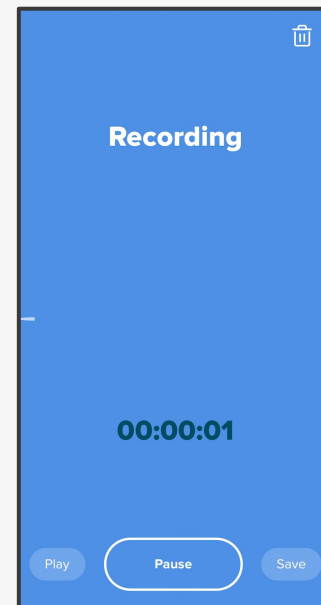


P3 tests his recognizer by crumpling paper.



## Field Study (1 week)

- Record three non-speech sounds each day
- Complete daily reflection



# Study Method

## 14 DHH participants

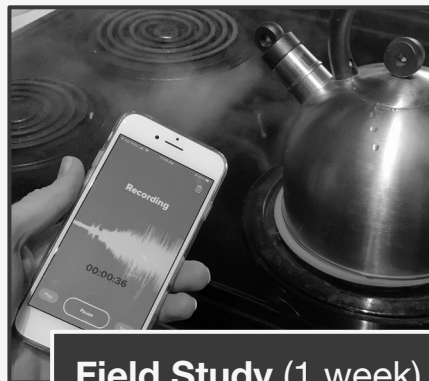
avg. 43.3 years old ( $SD=21.3$ , range=19-87)



P3 tests his recognizer by crumpling paper.

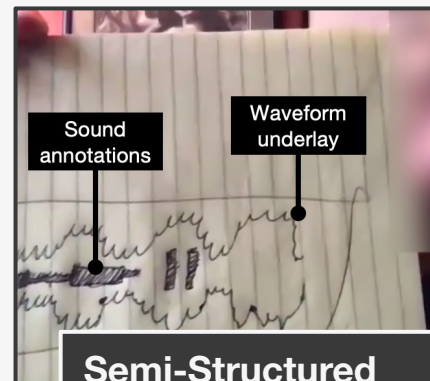
### Introductory Session (75 min)

- Introduce recording for sound recognition



### Field Study (1 week)

- Record three non-speech sounds each day
- Complete daily reflection



P9 suggests an enhanced waveform with individual sounds accentuated.

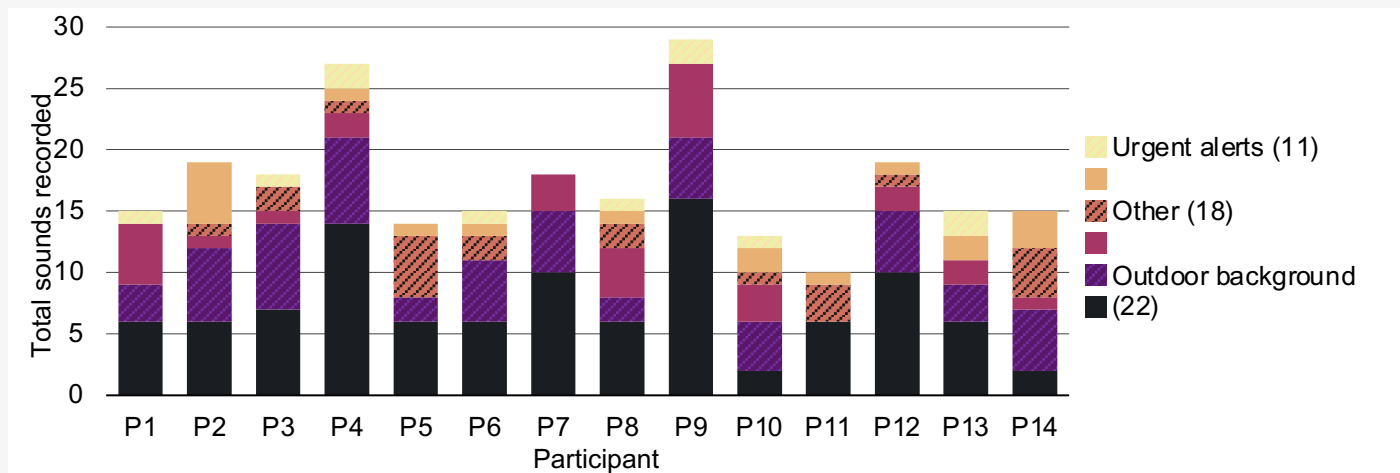
### Semi-Structured Interview (60 min)

- Reflect on the experience
- Design probe activity

# Findings

All 14 participants were **enthusiastic about recording sounds** and described the experience as “easy” ( $N=9$ ), “interesting” (7), and “fun” (P4, P10).

**243 sounds in total** (avg.=17.4/participant,  $SD=5.1$ ), avg. **2.8 samples per sound** ( $SD=1.2$ )



## Successful & Challenging Sounds

Participants reported **success** in recording sounds that were:

- **Continuous** P12:



- **Prominent** P14:



- **Controllable** P13:



They reported **challenges** in recording sounds that were:

- **Uncontrollable**

P3:



- **Complex-to-produce**

- **Delayed**

- **Hidden**

P7:



Toward Sound Recognizer Personalization with DHH Users

## Key Challenge 1: Waveform Interpretation

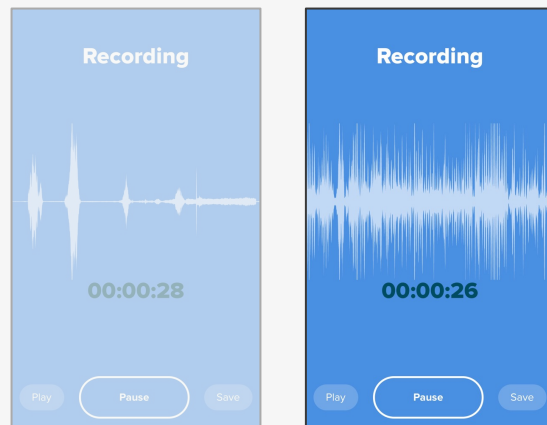
Though they were unable to hear aspects of the sound being recorded, Rev's **waveform was crucial for interpreting** the contents of samples.

But **breakdowns** occurred when participants' **intuition of the sound did not align** with the displayed visualization.

*Example:* P6 expected peaks during a thunderstorm.

Instead found a *“jumble of noise”* and *“blob of information”*.

She disregarded the waveform during the rest of the week.



## Key Challenge 2: Replicating Sounds

Participants' limited frame of auditory reference led to **uncertainty over how closely their samples replicated the real-world population.**

Those with residual hearing tried playback to determine whether the recording reflected the real-world version, but this was unreliable.

Many did not have this ability:

*“As a deaf person, [...] I’m just relying on my vision and my [other] senses [...] **there are visual indicators, but it’s hard to emulate [realistically].**” (P12)*

## Key Challenge 2.5: Replicating Variation

When recording samples of the same sound, limited perception of audible differences caused **further uncertainty about capturing realistic variation.**

*Example:* P2 recognized the benefit of diversity in samples of the same sound but incorporating this into her data was left to guesswork.

*“I **suspect** the doors and [blinds] sound differently when they are pulled or pushed in different speeds.”*



## Key Challenge 3: Uncertain Boundaries

Limited ability to hear audible differences between sounds also contributed to **uncertainty toward possible decision boundaries** within the model.

*Example:* P9 desired separate sound classes for the faucets in his home.

But he was unsure whether “*a stainless steel rectangular sink*” and “*a rounded porcelain sink*” **produce an audible difference.**

## Findings Summary

Participants reported a positive subjective experience, but their limited auditory expertise led to unique challenges with:

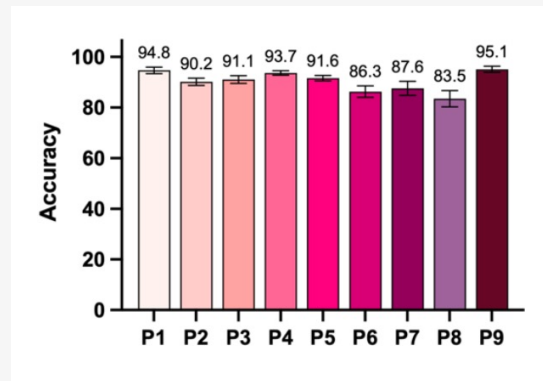
1. Assessing a sample's contents via playback or waveform.
2. Replicating a sound's real-world occurrence and range of variation.
3. Estimating decision boundaries via audible differences.

# Outcomes

I conducted a follow-up analysis of the audio samples collected by participants (*Jain et al., Sec. 6, CHI 2022*).

iOS and Android allow users to record samples to extend pre-trained models, but...

- Use low-fidelity audio visualization.
- Do not offer transparency for the quality of samples, or convey how well the model "learned" that sound



## Overview

# My dissertation work



**GOAL:** A framework for supporting Deaf and hard of hearing individuals to personalize sound recognition tools that meet their everyday needs.

## Motivation

# Models depend on their training data

Classification model accuracy depends on the similarity of training data to real-world data.

User-centered IML research often involves improving models by refining the training data

- *E.g.*, corrections, removal of erroneous samples, generating of new samples

Ishibashi et al. explored visualization options (spectrograms, thumbnails, and 2D embeddings) for browsing large sets of unlabelled audio samples.

DHH users may have difficulty interpreting audio samples...

- To identify appropriate training data.
- To understand how the model makes its decisions.

**This research aims to identify effective mechanisms to help DHH users build a high-quality training dataset.**

## Motivation

# IML can enhance users' understanding

- IML is promising for accessibility applications.
  - Disabled users can personalize assistive technology to meet their individual needs.
  - They can build an understanding of the system's strengths and weaknesses.
- Limited research has looked at how HCML applications impact DHH users' understanding of sound recognition tools.
  - Nakao et al.'s study found that users had a better understanding of ML after engaging with an AutoML training process, but their study had some limitations.
- The impact of an interactive machine learning process on DHH users' understanding and assessment of a sound recognition system is a second focus of my proposed research.

## Motivation

# IML can enhance users' understanding

IML is promising for accessibility applications.

- Disabled users can personalize assistive technology to meet their individual needs.
- They can build an understanding of the system's strengths and weaknesses.

Limited research has looked at how HCML applications impact DHH users' understanding of sound recognition tools.

- Nakao et al.'s study found that users had a better understanding of ML after engaging with an AutoML training process, but their study had some limitations.

The impact of an interactive machine learning process on DHH users' understanding and assessment of a sound recognition system is a second focus of my proposed research.

# Research questions

- 1. What kinds of feedback mechanisms, and UI elements are effective for supporting DHH users to personalize a sound recognition system?**
  - How do elements impact users' experience, understanding, and confidence during use?
  - What UI additions or improvements could provide further support?
- 2. What is the effect of a complete training and testing cycle on DHH users' understanding and assessments of a personalizable sound recognition system?**
  - How does the full cycle impact users' comprehension of the system (e.g., performance expectations and tolerances) compared to their pre-usage comprehension?
  - How confident are users in their own ability to assess the quality of a model that they have trained within a familiar, meaningful soundscape?

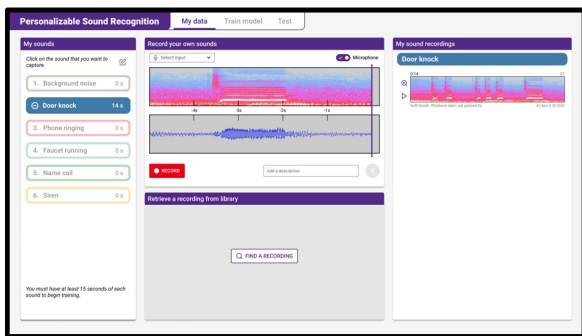


# Study Overview

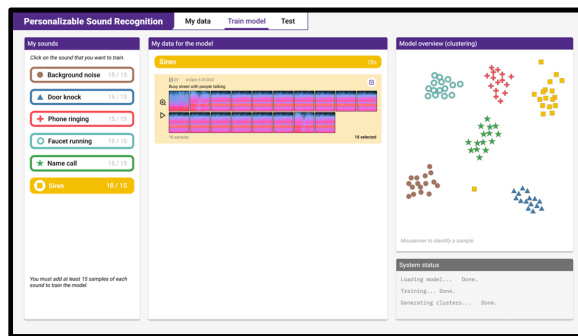
I plan to conduct a three-stage study:

1. **Develop an end-to-end sound recognition system** with a specialized interface to support DHH users in training a sound recognition model.
2. **Conduct usability testing with 4-6 expert and non-expert DHH participants** to improve the system's accessibility and the user experience.
3. **Evaluate 16-20 DHH participants' experience using the system** to create a personalized model for sounds within their homes.

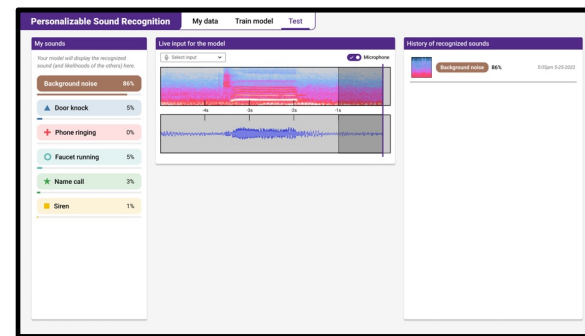
# Part 1: Specialized Interface



Planning &  
Data Collection



Data Curation &  
Training



Testing &  
Assessment

## Part 1: Specialized Interface

# Planning & Data Collection

**GOAL:** Users plan sound categories, then collect data

- Recording new samples
- Choosing from a library

**Challenge:** Identifying unintended sounds in recordings

- Requested a large, high-fidelity **waveform** to monitor ambient sound

**Idea: Spectrogram** visualizations add frequency information and match model's input

The screenshot displays a web application titled "Personalizable Sound Recognition" with three main tabs: "My data", "Train model", and "Test".

- My sounds:** A list of sound categories with durations:
  - 1. Background noise (0 s)
  - 2. Door knock (14 s)
  - 3. Phone ringing (0 s)
  - 4. Faucet running (0 s)
  - 5. Name call (0 s)
  - 6. Siren (0 s)
- Record your own sounds:** A section for capturing new audio. It includes a "Select input" dropdown, a "Microphone" toggle (checked), a spectrogram visualization, a waveform, a "RECORD" button, and a text input field for "Add a description".
- Retrieve a recording from library:** A section with a "FIND A RECORDING" search button.
- My sound recordings:** A section showing a specific recording titled "Door knock" with a spectrogram and a description: "Soft knock. Windows open, car passed by." The recording is dated "4:23pm 5-25-2022".

A note at the bottom of the "My sounds" section states: "You must have at least 15 seconds of each sound to begin training."

## Part 1: Specialized Interface

# Data Curation & Training

**GOAL:** Users “steer” the model by adding/removing data from the training set.

*Challenge:* How do audible differences impact the model's understanding?

- Requested comparison between samples and across sound classes

**Idea:** A clustering visualization (2D embedding) to reveal relative similarities

The screenshot displays a web application interface for 'Personalizable Sound Recognition' with three main tabs: 'My data', 'Train model', and 'Test'. The 'My data' tab is active and is divided into three panels:

- My sounds:** A list of sound classes with their respective sample counts. 'Siren' is highlighted in yellow and has 18 / 15 samples. Other classes include Background noise (15 / 15), Door knock (15 / 15), Phone ringing (15 / 15), Faucet running (15 / 15), and Name call (15 / 15). A note at the bottom states: 'You must add at least 15 samples of each sound to train the model.'
- My data for the model:** Shows a selected audio sample titled 'Siren' with a duration of 18s. Below it is a spectrogram of the audio, with 18 samples selected and highlighted in yellow.
- Model overview (clustering):** A 2D embedding visualization showing clusters of data points for different sound classes. The clusters are color-coded: cyan circles, red crosses, yellow squares, green stars, brown dots, and blue triangles. A tooltip提示 'Mouseover to identify a sample' is visible.

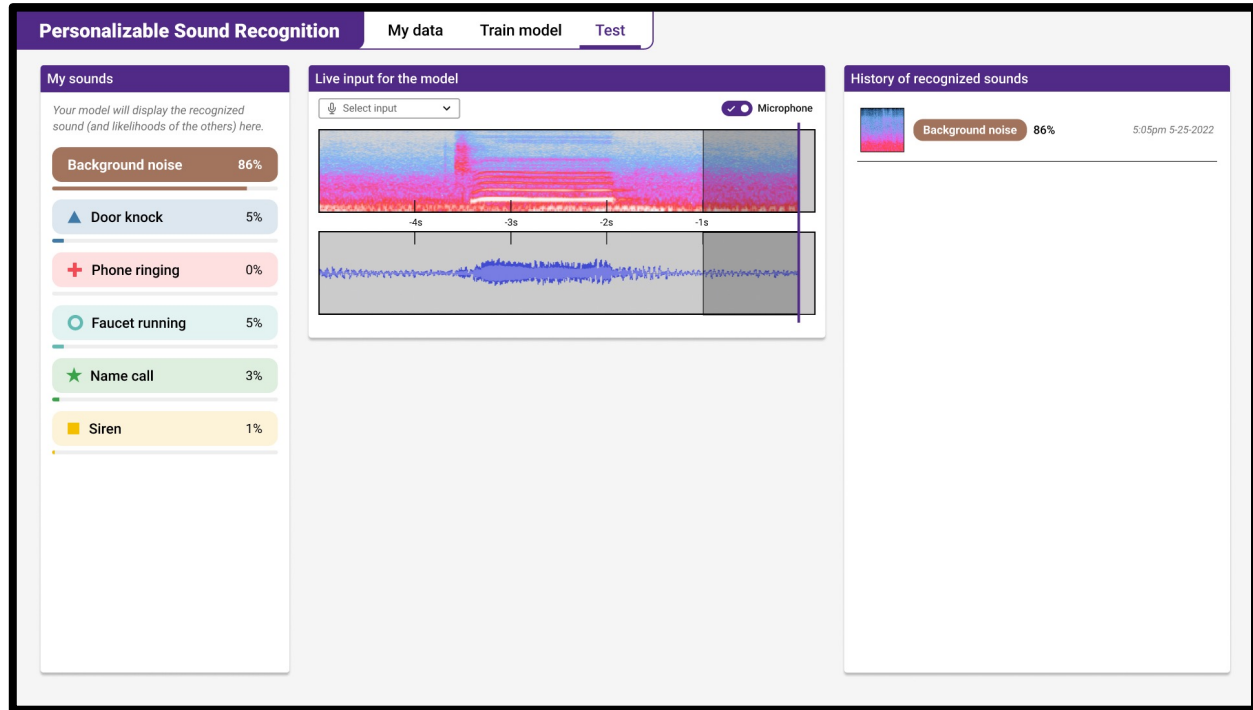
At the bottom right, a 'System status' panel shows the progress of the training process:

- Loading model... Done.
- Training... Done.
- Generating clusters... Done.

## Part 1: Specialized Interface

# Testing & Assessment

**GOAL:** Users assess the last iteration of their model by producing sounds and observing the output.



# Part 2: Usability Testing

Two rounds of usability testing with 4-6 DHH participants **to improve accessibility and user experience**

- 2-3 expert DHH machine learning and accessibility practitioners to identify potential technical issues
- 2-3 non-experts to provide insights as typical users
- 60 min sessions

Participants will use the system to train a model for two sounds and complete structured tasks on each interface tab .

- Ask for immediate thoughts, concerns, and suggestions for improving the associated interface element(s) after each task

Responses will be analyzed after each round of testing and integrated into a revised prototype.

## Part 3: User Evaluation

I will recruit 16-20 DHH participants for a **full evaluation of the system** for sounds in their homes.

- 120 min sessions over videoconferencing

Initial questions will **capture their immediate thoughts** on recording audio and use cases for a personalized sound recognizer.

Next, they will complete **a tutorial to provide scaffolding** for the technical aspects of the study.

- Explain how a sound recognition model works
- Define high-quality training and testing data
- Demonstrate the functionality of each interface element

## Part 3: User Evaluation

After the tutorial, participants will **use the system to train a model**.

1. Define sound classes for the model and plan how to capture these sounds.
2. Collect data by recording sounds or finding them in the video library.
3. Construct a high-quality training set based on clustering feedback.
4. Test the model's performance by producing the range of sounds.

The study will conclude with a **semi-structured interview** to debrief participants on the experience.

- *Satisfaction with their model, what they learned during the experience, and their confidence if performed again*



# Potential Findings

## **Design guidelines** for accessible machine learning interfaces

- The informational value of spectrograms and waveforms in the context of sound recognition
- Potential for clustering feedback to enhance DHH users' understanding of audio data
- Opportunities for other feedback mechanisms

## **Characterization of DHH users' understanding** of ML following an interactive training process

- Criteria for a personalizable sound recognition system to satisfy DHH users
- How interaction with training a model can change these criteria
- Whether the model's deployment context impacts understanding and confidence.

## Overview

# My dissertation work



**GOAL:** A framework for supporting Deaf and hard of hearing individuals to personalize sound recognition tools that meet their everyday needs.

# Dissertation Contributions

A **comprehensive empirical understanding** of DHH individuals' needs and preferences for personalization in sound awareness tools, including:

- The utility of different forms of sound feedback for DHH users and how contextual factors can modulate the relevance of that feedback (*Study 1*)
- The practical considerations and sense-making strategies that DHH people use in recording and interpreting real-world audio data to train a sound recognition model (*Study 2*)
- A deeper understanding of DHH peoples' training strategies and conceptualization of ML when creating a sound recognition model (*Study 3*)

**Guidance for the design** of personalizable sound awareness technology, such as:

- Characterization of the complementary roles of visual and vibrational feedback in sound awareness devices (*Study 1*)
- Implications and considerations for designing specialized recording tools to aid DHH users (*Study 2*)
- An end-to-end prototype system to sample, train, and test a sound recognition model (*Study 3*)
- Recommendations for UI elements that can facilitate DHH users in interpreting audio data, and training and evaluating a sound recognition model (*Study 3*).



## Discussion

# Open Questions

*How do DHH users integrate personalized sound recognition tools into their daily lives, and how do their perceptions and attitudes towards such tools change over time?*

- Model deployment and continued refinement are final steps in the ML process that are not included in my work.

*What other methods can enable DHH users to steer the training of sound recognition model?*

- My work only explores IML by adjusting the training dataset.

*Should privacy-preserving techniques be implemented to ensure that personalizable sound recognition tools are safe for DHH users and bystanders?*

- My work did not explore the concerns of DHH users toward personal audio data.

# Thank You

Dissertation title:

**Human-Centered Sound Recognition Tools for Deaf and Hard of Hearing People**

**Steven M. Goodman**

PhD Candidate

Human Centered Design and Engineering

UNIVERSITY *of* WASHINGTON

Committee:

Leah Findlater (chair), Julie Kientz,  
Jon Froehlich, Mark Harniss (GSR)