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

Steven M. Goodman

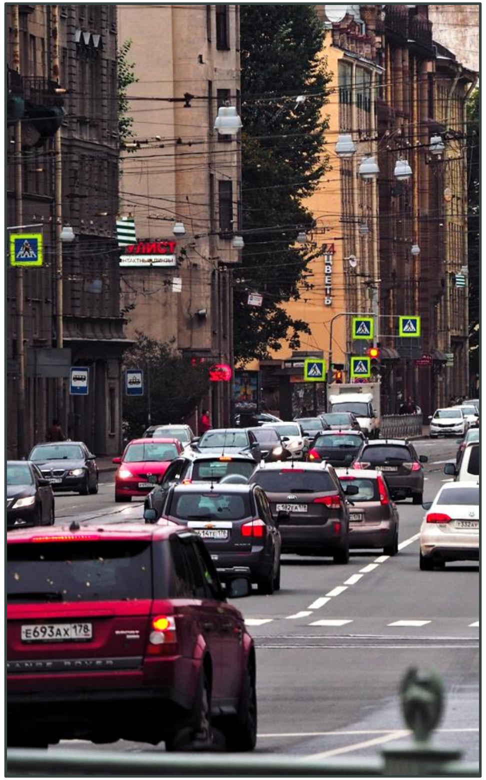
Dissertation Defense

November 7, 2024

inclusive design lab
HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON

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

Motivation

Desire for sound awareness

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

Survey of 201 DHH participants[1]:

Smartwatches are the preferred portable device for sound awareness

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

However, the effective combination of this feedback — particularly in busy environments—is an open question.

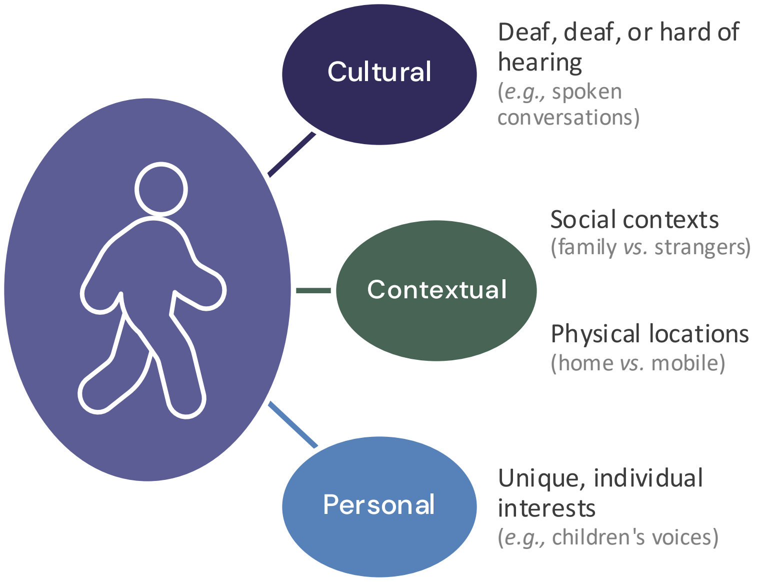
[1] *Findlater et al., CHI 2019; Bragg et al., ASSETS 2016; Jain et al., CHI 2019; Matthews et al., 2006*

Sound interests are diverse

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.



Bragg et al., ASSETS 2016; Findlater et al., CHI 2019; Jain et al., CHI 2019; Matthews et al., 2006

Introduction

Current technologies are limited

Android & iOS include automatic sound recognition models.

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

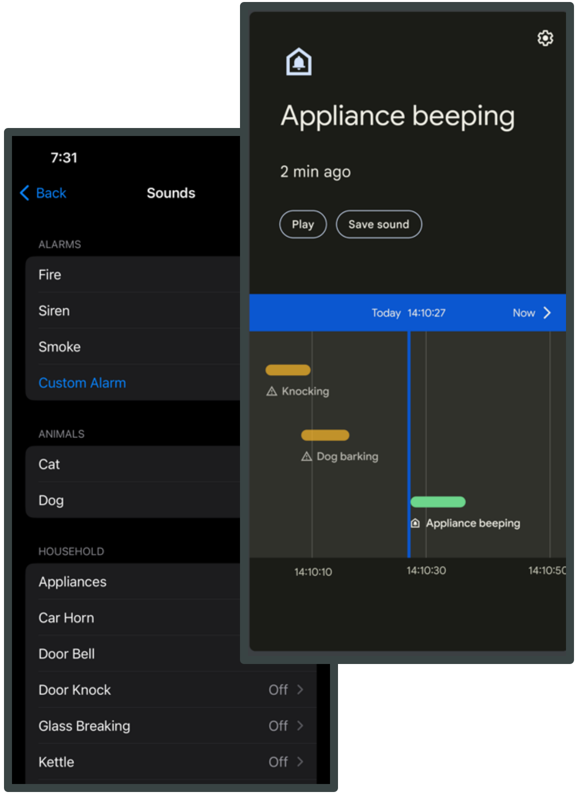
- *E.g.,* appliances, alarms, pets

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 … with accuracy and flexibility, and desire t… sound recognition model.

[Jain et al., CHI 2022]

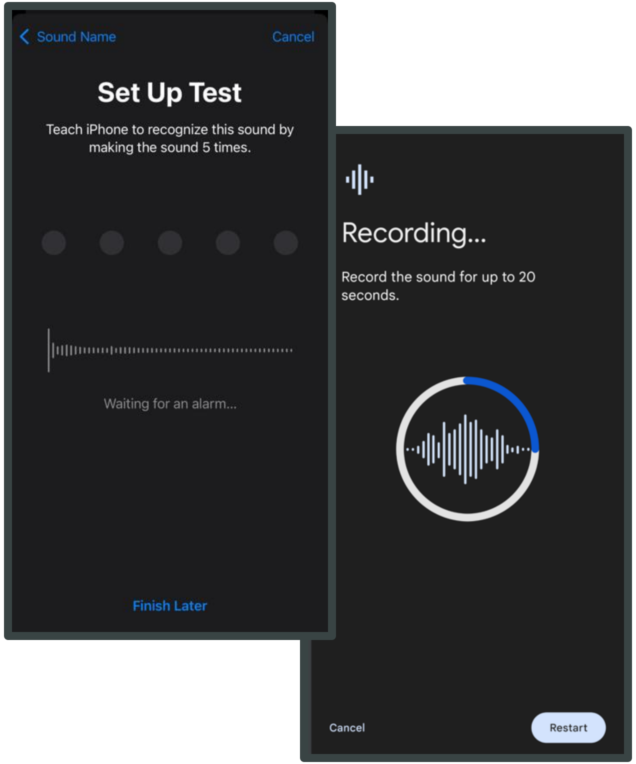


Introduction

Small steps toward personalization

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

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



This "AutoML" approach is fast and easy but lacks transparency and control—which could limit users' trust and long-term use.

Drozdal et al., IUI 2020

Interactive personalization is possible

User-driven ML systems can yield automatic tools that are tunable to wide-ranging needs.

Further, interactive machine learning (IML) offers a framework for increased understanding of the model's strengths and limitations and fostering trust and transparency.

Sanchez, CSCW 2021

But most IML literature assumes end-users have domain knowledge and access to the model's underlying data.

Dudley et al., 2018

An open question lies in how to support a DHH user—who does not have full access to a sound themselves—in training an ML model to recognize that sound.

Dissertation outline

Part 1

Evaluation of **sound feedback** for DHH users **across contexts**

Part 2

Investigation of DHH users' **audio recording experience**

Part 3

Evaluation of an **accessible training pipeline** with DHH users

How do DHH users desire sound information to be delivered?

How do contextual factors impact these preferences?

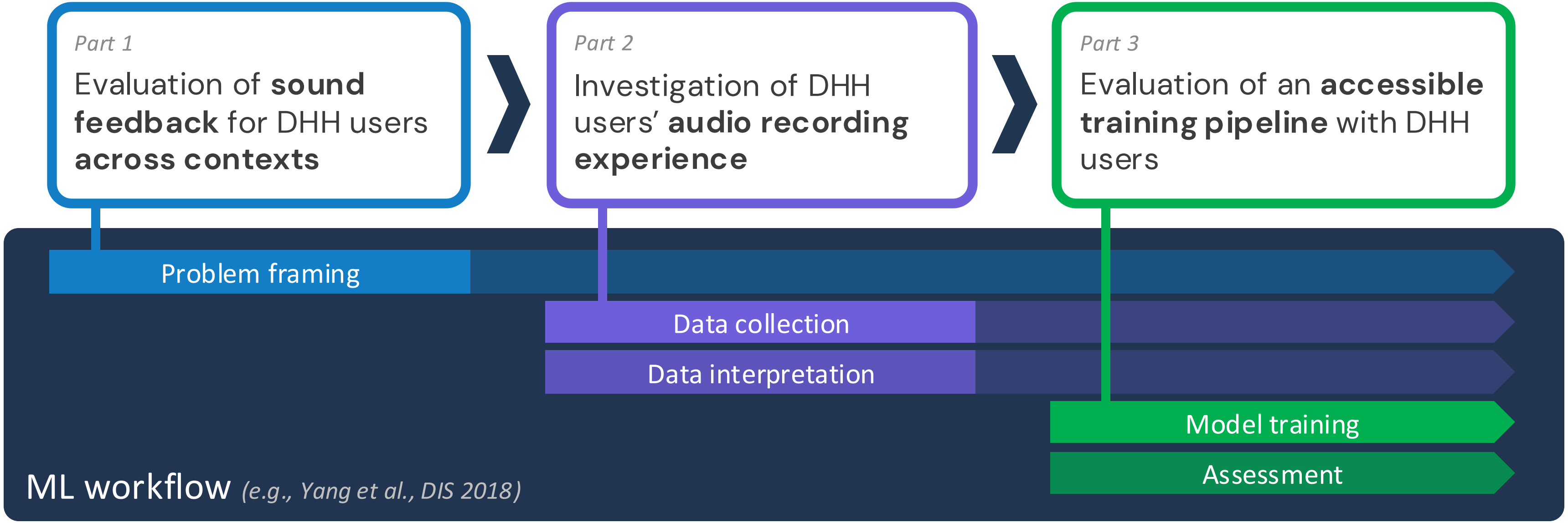
How do DHH users capture, interpret, and conceptualize audio data for automatic sound recognition?

What feedback and data representations can effectively support DHH users in personalizing a sound recognition model?

How does a complete training cycle affect DHH users' perspectives on sound recognition systems?

Human-Centered Sound Recognition Tools

Dissertation outline



GOAL: Understanding Deaf and hard of hearing individuals' needs and preferences around personalization in sound recognition tools.

Thesis Statement

For DHH people who desire greater access to sound information, ***technology should be designed for personalized and adaptable experiences****—providing relevant information, offering granular control, and promoting confidence and agency among users.*

Dissertation outline

Part 1
Evaluation of sound feedback for DHH users across contexts

Part 2
Investigation of DHH users' **audio recording experience**

Part 3
Evaluation of an **accessible training pipeline** with DHH users

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

CHI 2020

Steven Goodman
Susanne Kirchner
Rose Guttman
Dhruv Jain
Jon E. Froehlich
Leah Findlater

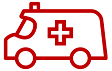
UNIVERSITY of WASHINGTON
HUMAN CENTERED DESIGN & ENGINEERING
inclusive design lab
DUB

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 

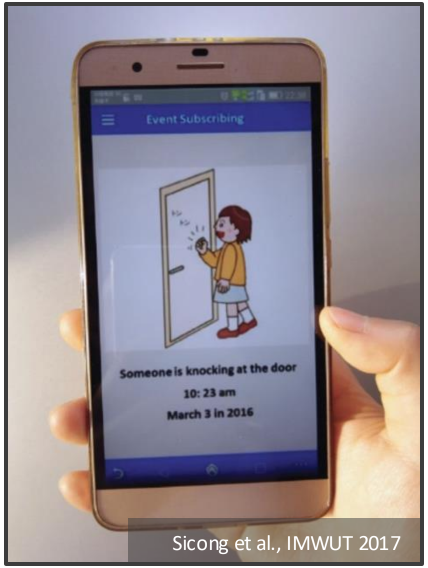
For sound awareness technology:

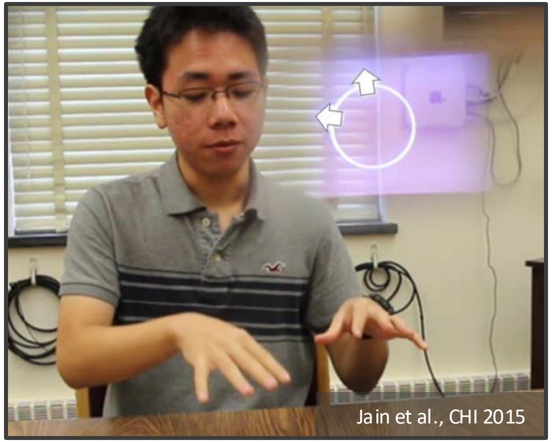
- Tools should be portable for use in a variety of contexts.
- Users desire sound feedback through visual and haptic modalities.
- Unimportant sounds should be filtered out from incoming feedback.

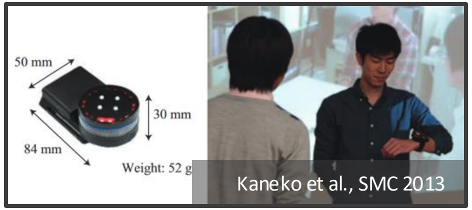
Bragg et al., ASSETS 2016; Findlater et al., CHI 2019; Matthews et al., 2006; Sicong et al., IMWUT 2017

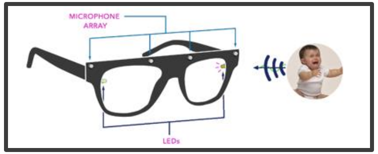
Motivation

Form factors of sound awareness tools









smartphones

head-mounted displays

wearable solutions

Motivation

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 also 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

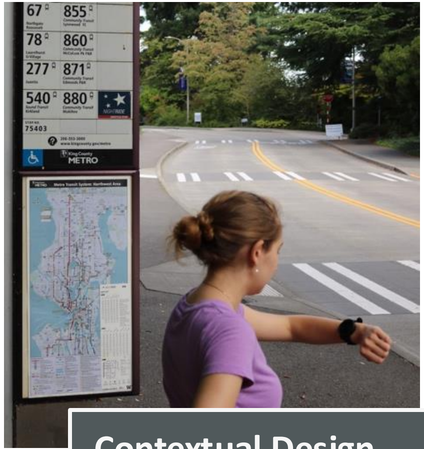
Lab-based Design Probe (30 min)

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



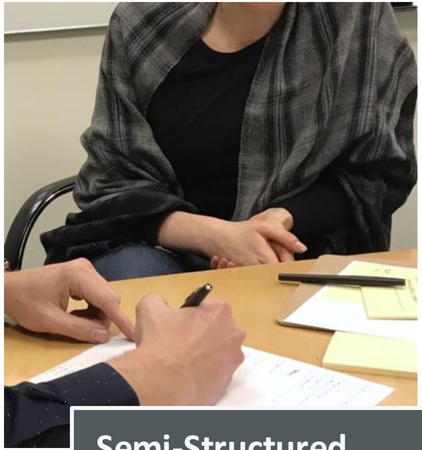
Contextual Design Probe (25 min)

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 #1:

Visual and haptic feedback have complementary roles in sound awareness.

Findings

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 <u>with</u> vibration were more useful than <u>without</u>:

- 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 #2:

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

Findings

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**

Findings

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**.”*

Findings

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.

Findings

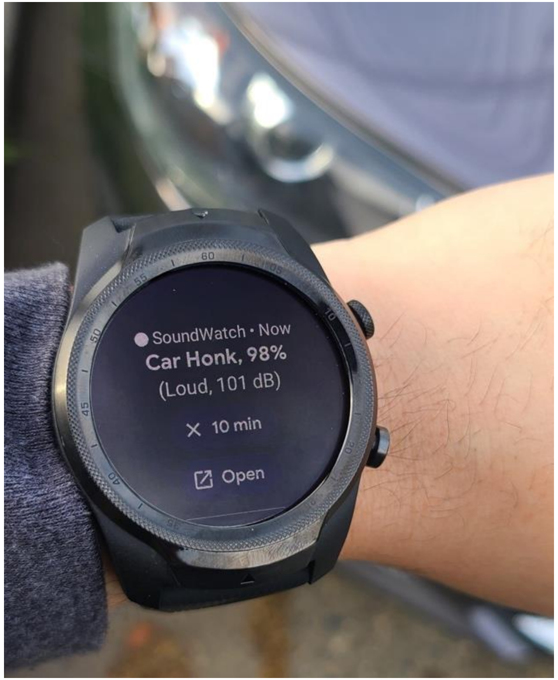
Soundscape Filtering

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...

Dissertation outline

Part 1

Evaluation of **sound feedback** for DHH users **across contexts**

Part 2

Investigation of DHH users' audio recording experience

Part 3

Evaluation of an **accessible training pipeline** with DHH users

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

IMWUT 2021

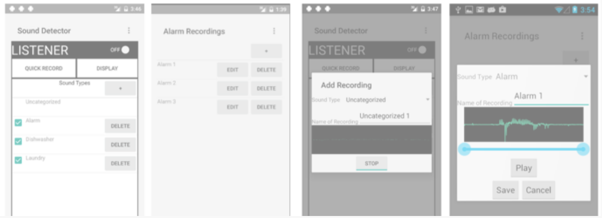
Steven Goodman
Ping Liu
Dhruv Jain
Emma J. McDonnell
Jon E. Froehlich
Leah Findlater

UNIVERSITY of WASHINGTON
HUMAN CENTERED DESIGN & ENGINEERING
inclusive design lab
DUB

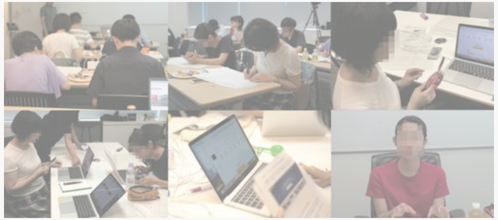
Motivation

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?

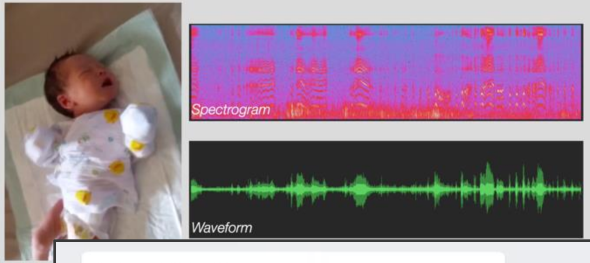
Study Method

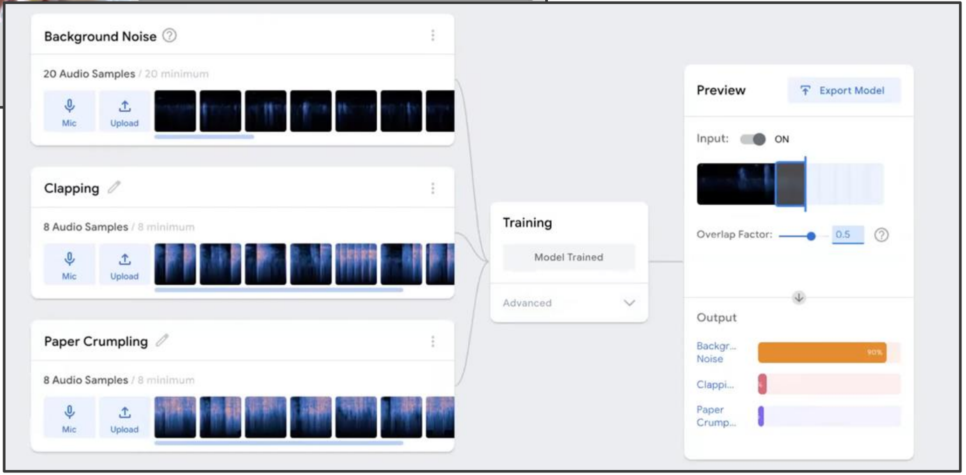
14 DHH participants
avg. 43.3 years old (*SD*=21.3, range=19-87)

- 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

Baby crying during a thunderstorm





Study Method

14 DHH participants

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

Introductory Session (75 min)

- Introduce recording for sound recognition

Field Study (1 week)

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

Study Method

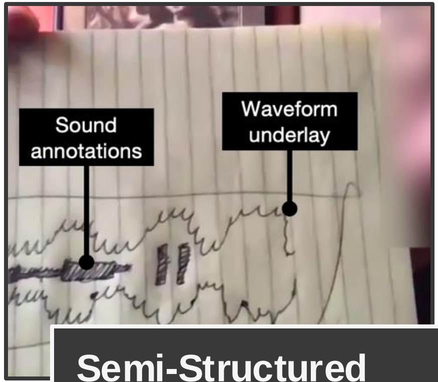
14 DHH participants

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

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.

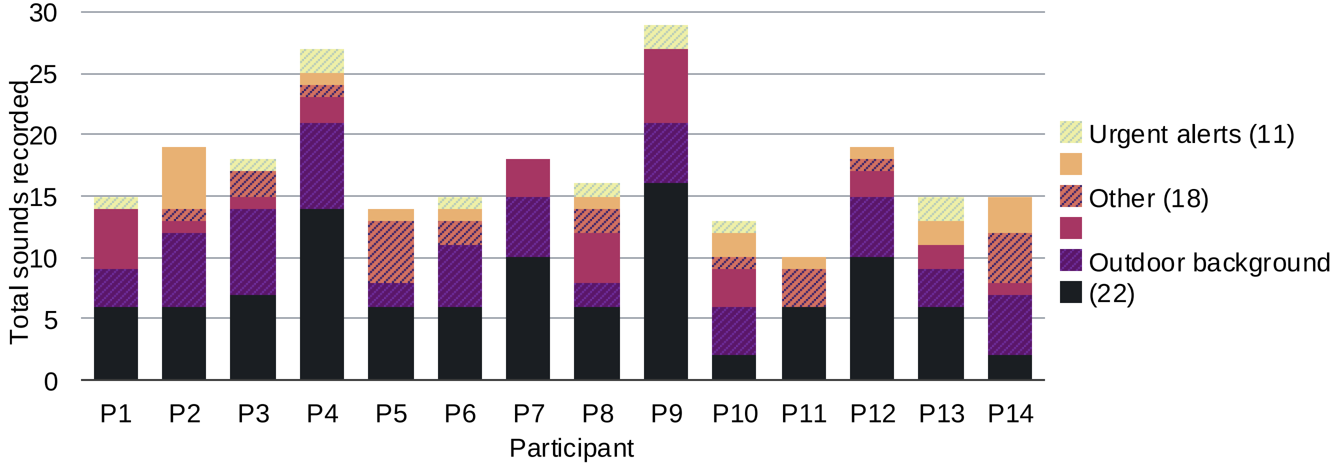


Semi-Structured Interview (60 min)
- Reflect on the experience
- Design probe activity

Findings

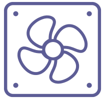
All 14 participants were **enthusiastic about recording sounds** and described the experience as “easy” (*N*=9), “interesting” (7), and “fun” (P4, P10).

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

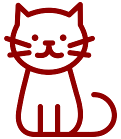
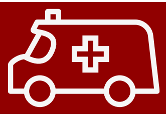


Successful & Challenging Sounds

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: 

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”* **produces 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.

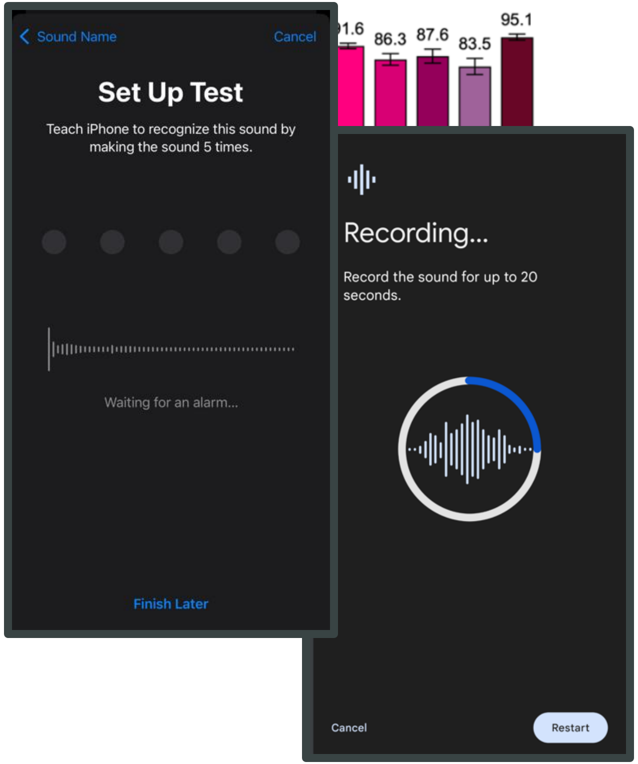
Outcomes

Follow-up analysis of the audio samples collected revealed the potential for ranging sounds.

ProtoSound - 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 indicate the quality of samples
- Do not convey how the model "learned" the sound



Dissertation Outline

Part 1

Evaluation of **sound feedback** for DHH users **across contexts**

Part 2

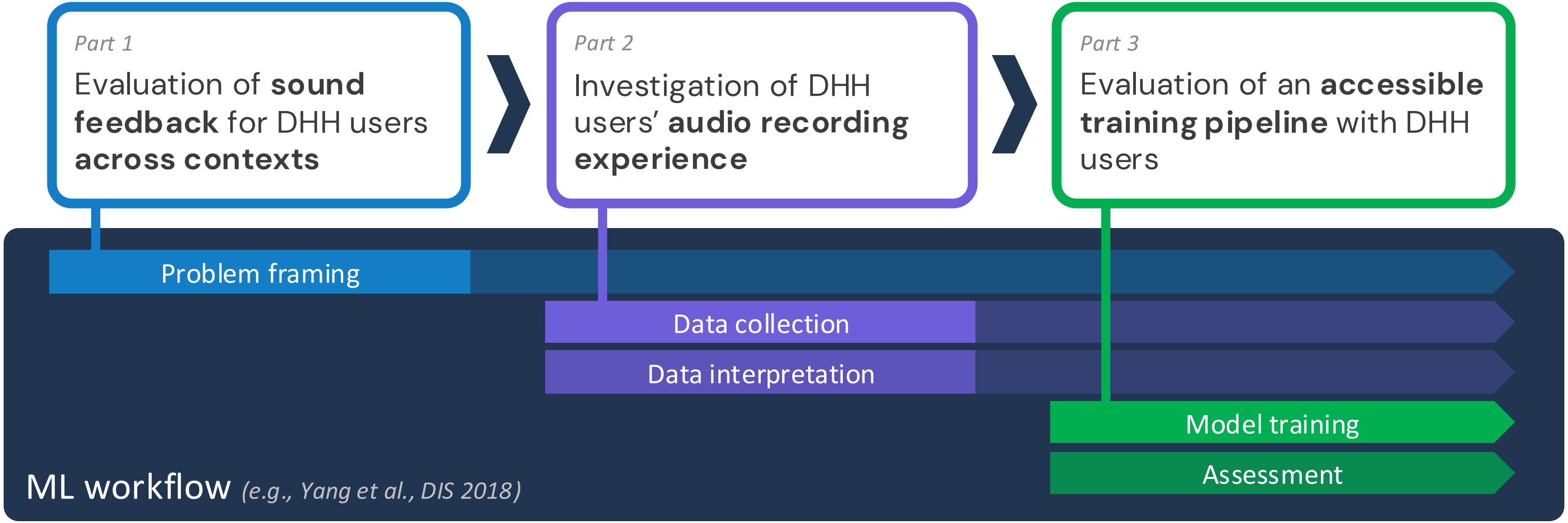
Investigation of DHH users' **audio recording experience**

Part 3

Evaluation of an accessible training pipeline with DHH users

Human-Centered Sound Recognition Tools

Dissertation outline



GOAL: Understanding Deaf and hard of hearing individuals' needs and preferences around personalization in sound recognition tools.

UNIVERSITY of WASHINGTON

HUMAN CENTERED DESIGN & ENGINEERING

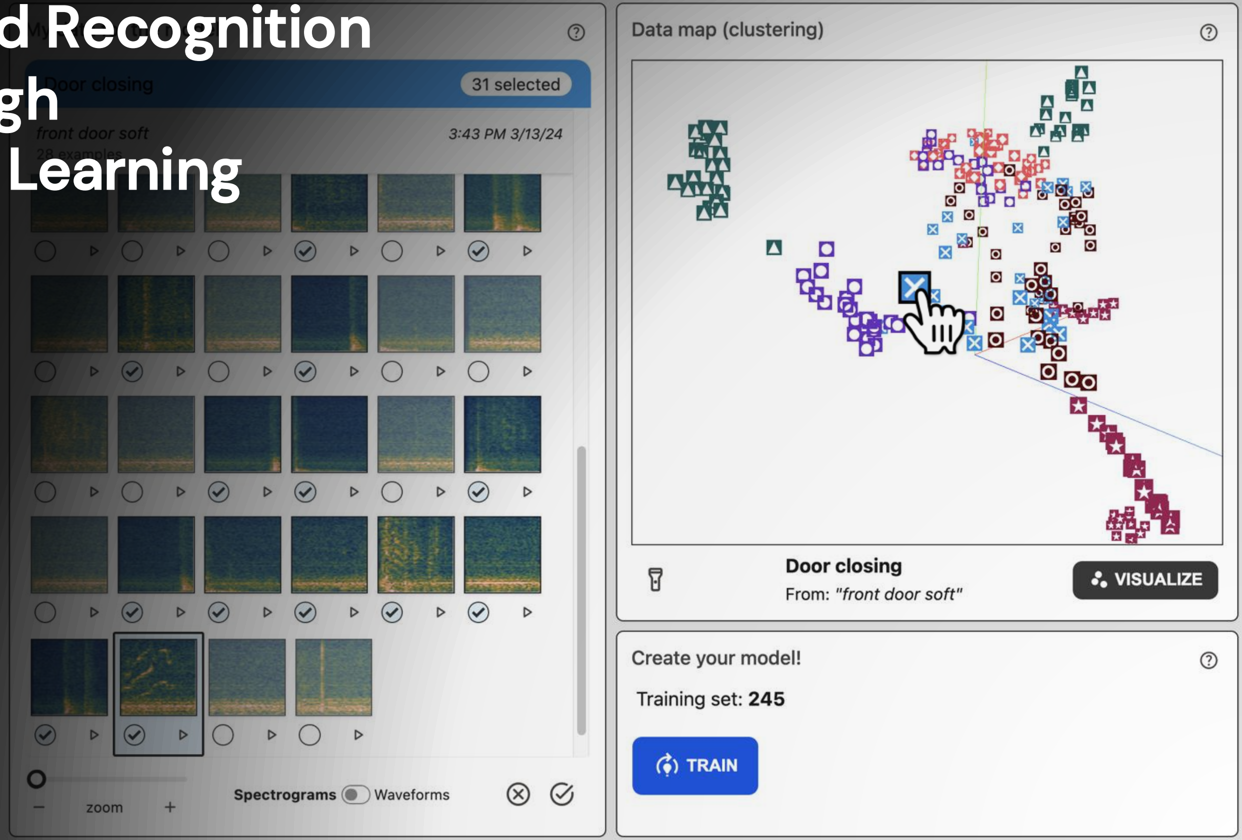
inclusive design lab

DUB

SPECTRA:

Personalizable Sound Recognition for DHH Users through Interactive Machine Learning

In submission



Steven Goodman
Emma McDonnell
Jon E. Froehlich
Leah Findlater

Motivation

Can IML systems benefit DHH users?

Interactive ML is promising for accessibility applications.

- Personalized assistive technology can meet individual needs.
- Users learn system's strengths and weaknesses.

However, IML within sound recognition tools has not been explored with DHH users.

- Nakao *et al.* found greater understanding w/ an AutoML system (lacked audio visualizations or insight into the model).

Aim #1: Investigate IML's impact on DHH users' perspectives around sound recognition.

Nakao et al., NordiCHI 2020

Kacorri et al., CHI 2017; Nakao et al., NordCHI 2020; Sosa-Garcia & Odone; TACCESS 2017

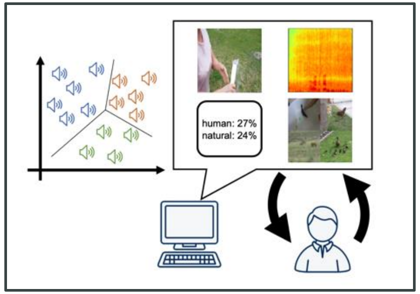
Motivation

Models depend on training data

A model's accuracy is determined by the similarity of its training data to real-world inputs.

IML work often seeks to improve models by refining the training data (adding, removing, correcting examples).

- *E.g.,* Ishibashi *et al.* explored different visualization options for selecting from large audio datasets.



Ishibashi et al., IUI 2020

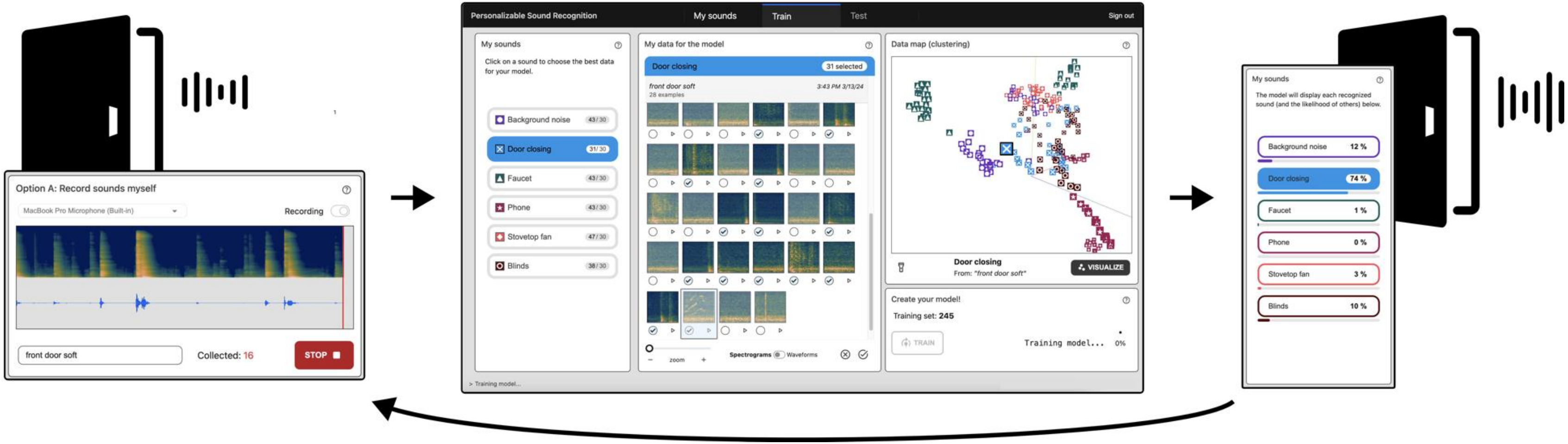
However, DHH users may struggle to interpret their data to…

- Identify appropriate training examples.
- Understand the data's impact on a model's learning and performance.

<u>Aim #2:</u> Identify effective support for DHH users to engage with IML.

Dudley et al., Trans. Interact. Intell. Sys. 2018; Ishibashi et al., IUI 2020

Overview of the SPECTRA pipeline



1. Planning &
Data collection

2. Data curation &
Model training

3. Testing &
Assessment

Personalizable Sound Recognition | My sounds | Train | Test | Sign out

My sounds

You must have at least two sounds to continue.

1. *Background noise*
2. Phone
3. Microwave
4. Siren
5. Door closing
6. Sound name

+ ADD ANOTHER SOUND

Option A: Record sounds myself

Default - MacBook Pro Microphone (Built-in)

Start listening

Add a description...

RECORD

Option B: Find a recording from the library

- Human sounds
- Animal
- Music
- Sounds of things
- Source-ambiguous sounds
- Channel, environment and background

Click on a category to find subcategories or browse available audio clips.

My sound recordings

No sound selected

> Editing sound names...

Research Questions

- How do DHH users engage with SPECTRA to train a personalized sound recognition model?
    - Specifically, how do waveform and spectrogram visualizations, interactive data clustering, and data annotating impact their participation in an interactive ML process?
- How does interaction with SPECTRA affect DHH users' perspectives on sound models and their confidence with custom training?

User evaluation (120 min)

12 DHH participants

- Any tech. experience level
- Moderate confidence in ML concepts
  (7-point scale, avg.=4.8, range=3-6)
- Five w/ hands-on experience

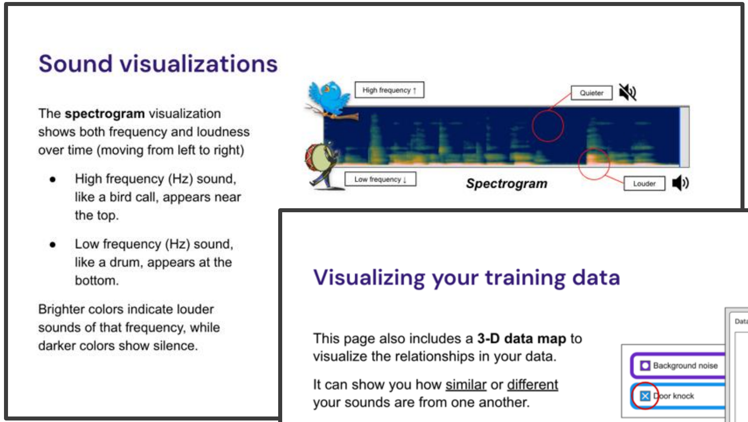
User evaluation (120 min)

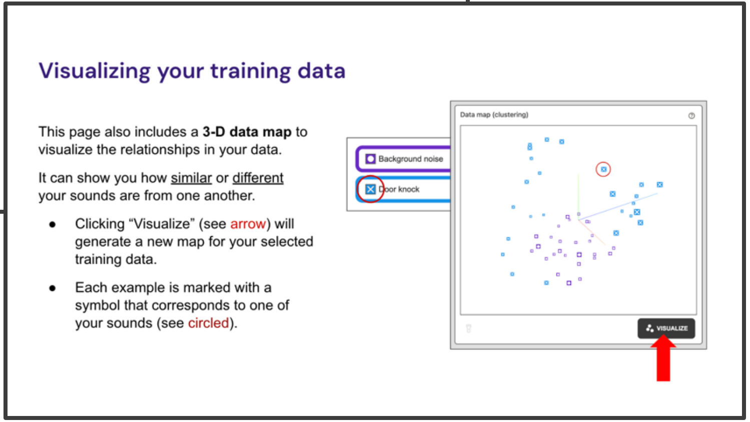
12 DHH participants

- Any tech. experience level
- Moderate confidence in ML concepts
  (7-point scale, avg.=4.8, range=3-6)
- Five w/ hands-on experience

Tutorial & Interview
(30 min)

- Introduce SPECTRA and sound recognition concepts
- Capture pre-use expectations

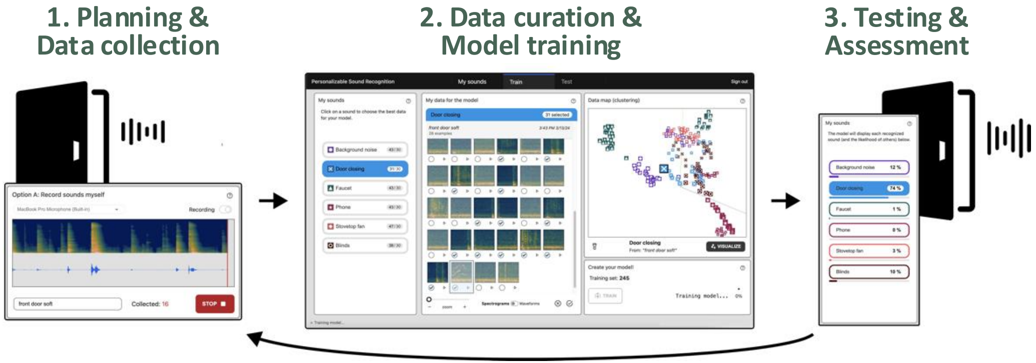




User evaluation (120 min)

12 DHH participants

- Any tech. experience level
- Moderate confidence in ML concepts
  (7-point scale, avg.=4.8, range=3-6)
- Five w/ hands-on experience



Tutorial & Interview (30 min)

- Introduce SPECTRA and sound recognition concepts
- Capture pre-use expectations

System Use (60 min)

- Train model for six sounds
- Collect data, build training dataset, test

SPECTRA

User evaluation (120 min)

12 DHH participants

- Any tech. experience level
- Moderate confidence in ML concepts
  (7-point scale, avg.=4.8, range=3-6)
- Five w/ hands-on experience

Tutorial & Interview (30 min)
- Introduce SPECTRA and sound recognition concepts
- Capture pre-use expectations

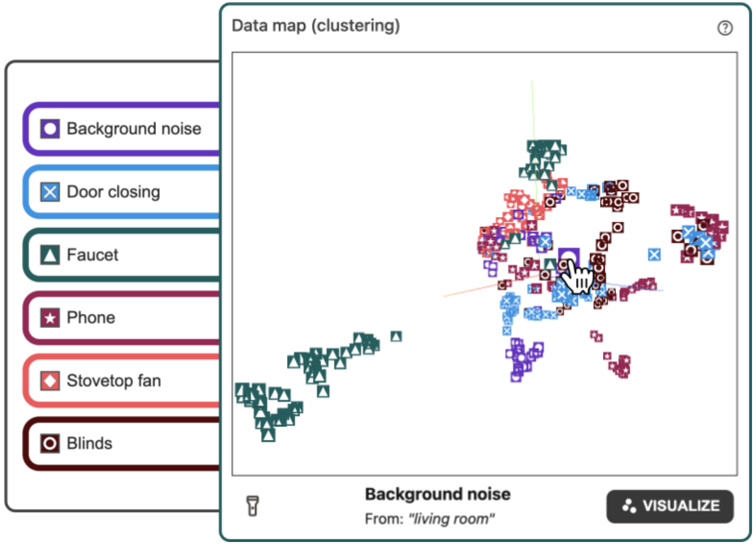
System Use (60 min)
- Train model for six sounds
- Collect data, build training dataset, test

Semi-structured Interview (30 min)
- Reflect on the experience
- Capture post-use perspectives

Findings

#1: Insights through clustering

P10's unfiltered data



> *"An understanding of what's happening under the hood"* – P8

Clustering was deemed critical to the IML process.

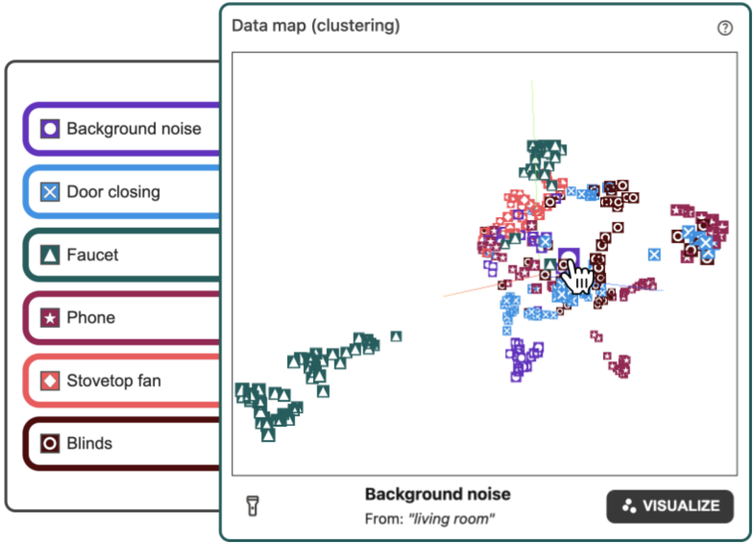
It offered an accessible way for users to:

- Understand similarities among sounds
- Troubleshoot sources of misclassification (via overlap)
- Highlight the most distinct sounds
- Identify outliers and iteratively refine training dataset

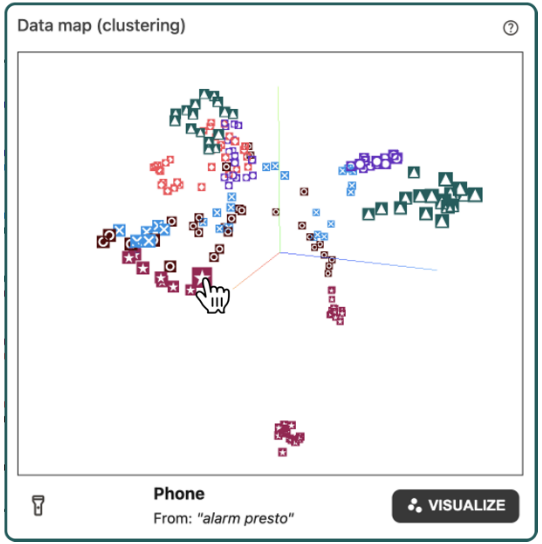
Findings

#1: Insights through clustering

P10 - unfiltered data



P10 – final training dataset



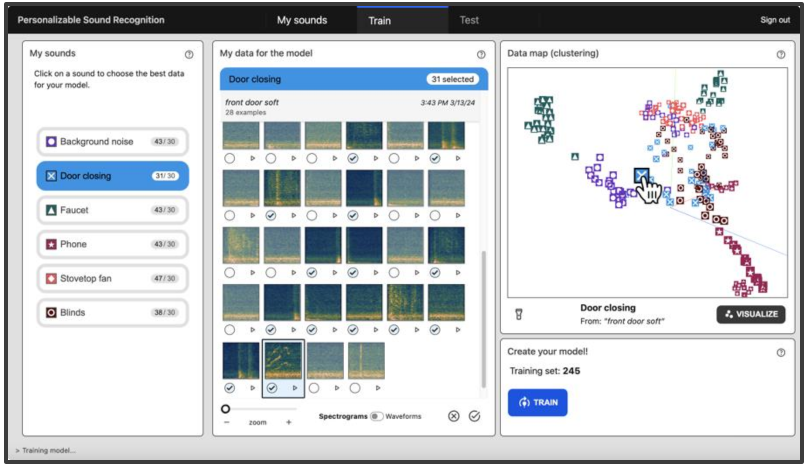
"[It] was satisfying to see, 'Okay, like it's actually working; what I'm doing.'" - P1

Findings

#2: Value of multimodal information

Participants combined multiple information streams to make decisions about their models.

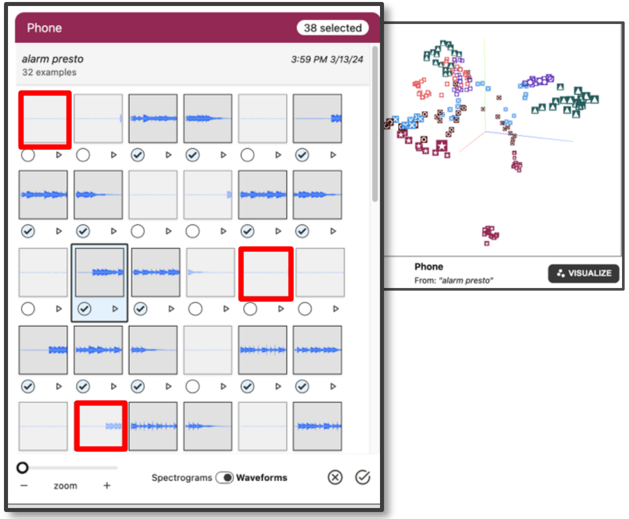
- Clustering: High-level view of data structure & relationships
- Waveform: Intuitive, glanceable, good for quick assessment
- Spectrograms: Less intuitive, but useful for in-depth analysis (for some)
- Annotations: Provide context, aid in recall, support deeper understanding



#2: Value of multimodal information

Training strategy A *(Example-centric)*

Analysis via example icon, clustering for monitoring



P4 on the waveform's glanceability:

"The background noise [vs.] whenever I was talking,

Being able to figure out which [example] was which—I think that was really helpful."

#2: Value of multimodal information

Training strategy A *(Example-centric)*

Holistic analysis via example icon, clustering for monitoring

"I was driven by what I was seeing in the chart [...] to eliminate some edge cases and anomalies.

Everything is [shown] together.

In [the selection panel], I have to compare one by one" – P11

Training strategy B *(Clustering-centric)*

Clustering as interactive flagging tool, example icon for targeted analysis

Findings

#3: Balancing engagement & efficiency

Interactive ML process promoted understanding and confidence, but the process was time-consuming.

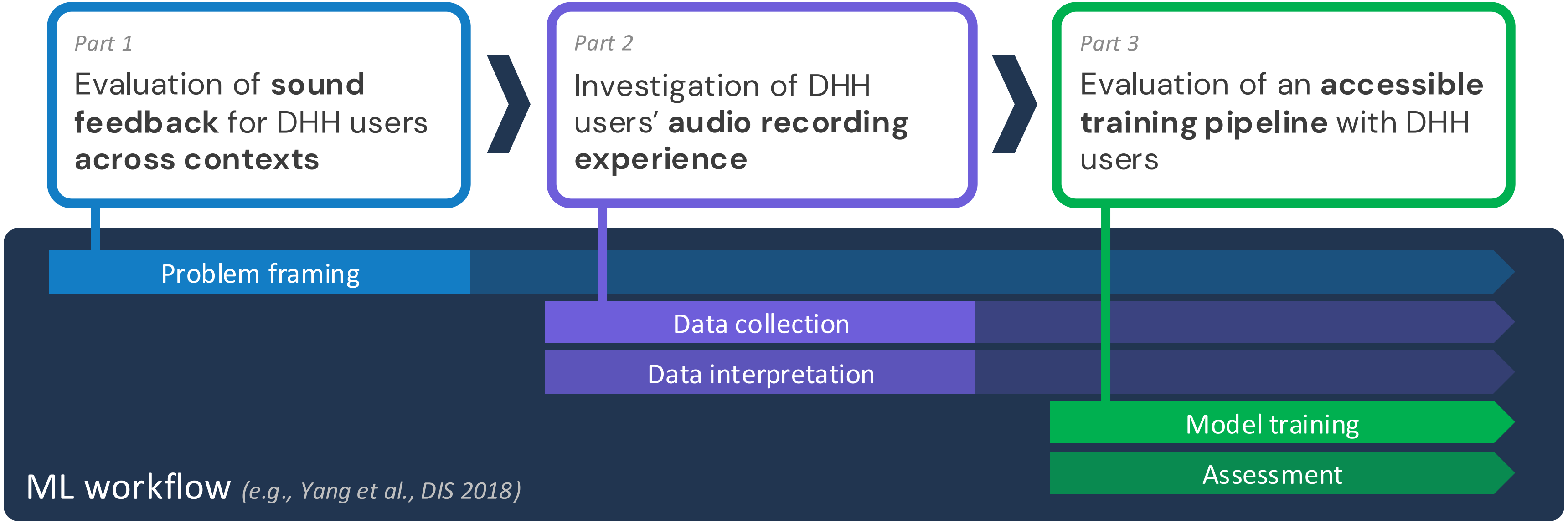
All participants trained just one model due to time limits or fatigue.

- Workflow requires too many interactions to produce a useful result
- P3: *"The unchecking: not my favorite; [...] it just ate up time."*

How to optimize this process without sacrificing meaningful engagement with the model?

- Automation (*e.g.,* background noise removal)
- *In situ* help (*e.g.,* P7: *"text reminders"* suggesting problematic examples)
- Starting from pre-trained model and adding custom classes (P11)

Dissertation outline



GOAL: Understanding Deaf and hard of hearing individuals’ needs and preferences around personalization in sound recognition tools.

Thesis Statement

For DHH people who desire greater access to sound information, ***technology should be designed for personalized and adaptable experiences****—providing relevant information, offering granular control, and promoting confidence and agency among users.*

Contributions

Empirical contributions

- Utility of different sound feedbacks and how contextual factors modulate the relevance of this feedback
- Considerations and sense-making strategies that DHH people use in recording and interpreting real-world audio data
- Insight on training strategies and their conceptualization of ML when creating sound models

Design contributions

Contributions

Empirical contributions

- Utility of different sound feedbacks and how contextual factors modulate the relevance of this feedback
- Considerations and sense-making strategies that DHH people use in recording and interpreting real-world audio data
- Insight on training strategies and their conceptualization of ML when creating sound models

Design contributions

- Characterization of visual and vibrational feedback's roles in sound awareness devices
- Considerations for specialized recording tools to aid in capturing an audio dataset
- An end-to-end prototype for interactive training of a sound recognition model
- UI recommendations to support interpreting audio data, building a training dataset, and evaluating a model

Discussion

Open Questions

How do DHH users integrate personalizable sound recognition tools into their daily lives? How do their attitudes towards such tools change over time?

- Model deployment and continued refinement are additional steps of the ML process.

Can privacy-preserving techniques ensure that personalizable sound recognition tools are safe for DHH users and bystanders?

- Custom models need recordings, and DHH users and bystanders alike may not be aware when private conversations are being recorded.

Thank you!

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

Committee:

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

Collaborators:

Rose Guttman
Dhruv Jain
Susanne Kirchner
Raja Kushalnagar
Ping Liu
Emma McDonnell

Lab members:

Gina Clepper
Lucy Jiang
Avery Mack
Abigale Stangl
Lotus Zhang

Steven M. Goodman

PhD Candidate
Human Centered Design and Engineering

UNIVERSITY *of* WASHINGTON

Questions?