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

IMWUT 2021

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

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

Background

Sound Recognition

Automatic sound recognition can support communication and environmental awareness for d/Deaf and hard of hearing (DHH) people.
e.g., Bragg et al., ASSETS 2016; Findlater et al., CHI 2019

Prior implementations use pre-trained models with generic sound classes.
e.g., Sicong et al., IMWUT 2017; Jain et al., ASSETS 2020

These systems do not meet personalization requests from DHH users:
e.g., Bragg et al., ASSETS 2016; Jain et al., CHI 2020

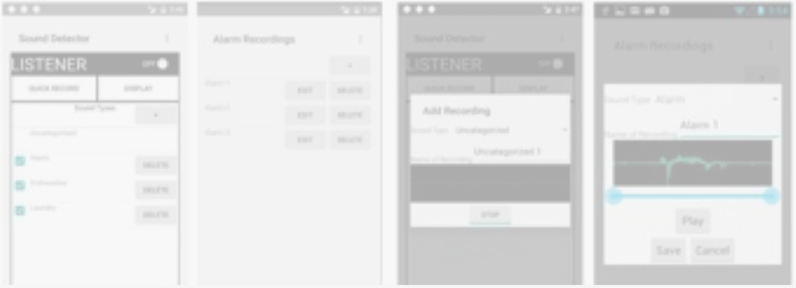
1. Do not allow for **custom sounds**

2. Do not account for **edge cases**

Related Work

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

I. How can a DHH person, who has difficulty hearing a sound themselves, **effectively record samples to train an ML system** to recognize that sound?

II. What considerations do DHH people make when recording in environments with **real-world acoustic variation?**

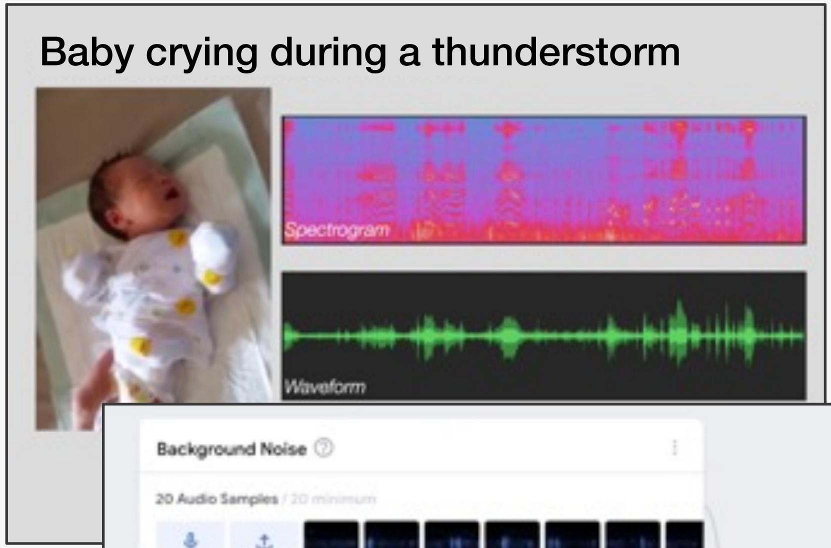
III. What kinds of features can aid DHH users in **assessing their recorded samples** as training data?

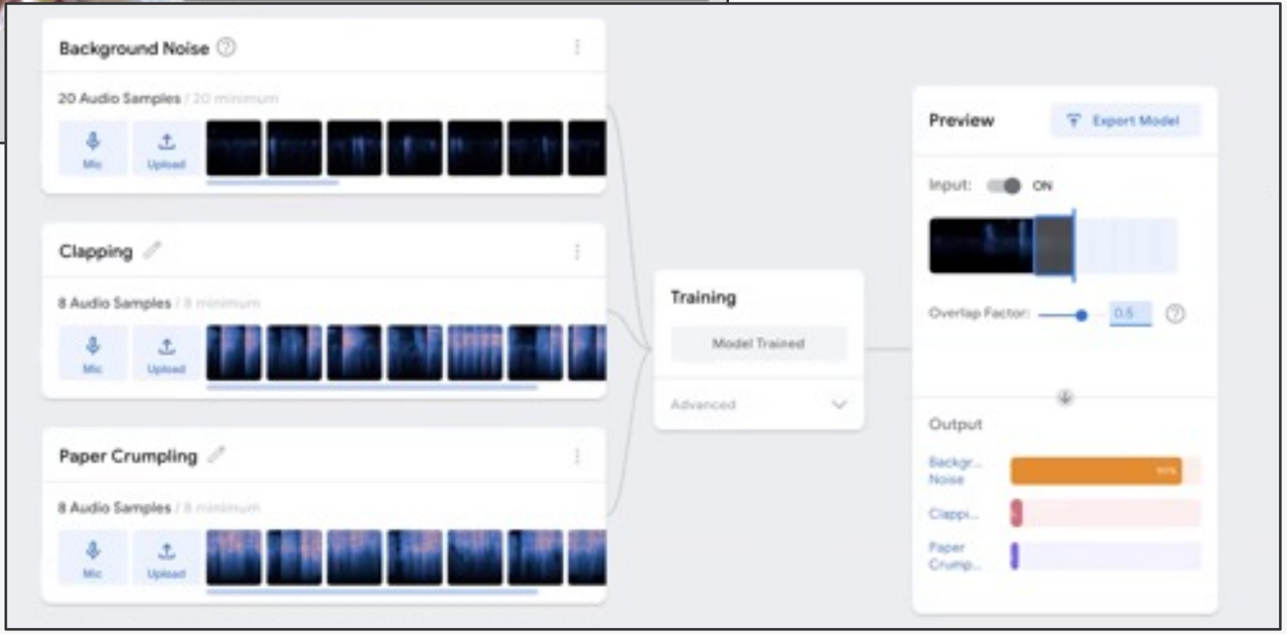
Study Method

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





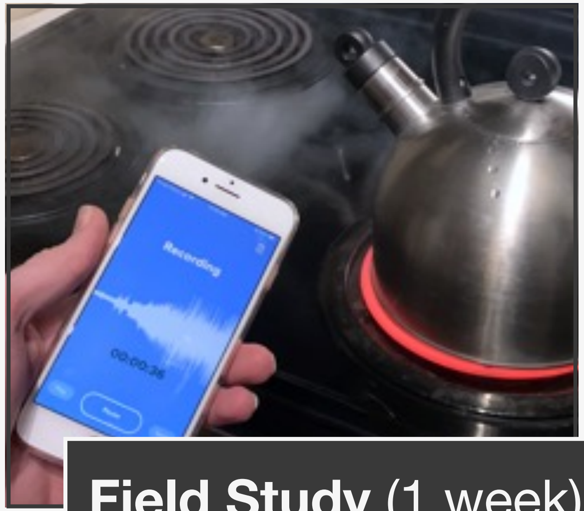
Study Method

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)

P3 tests his recognizer by crumpling paper.

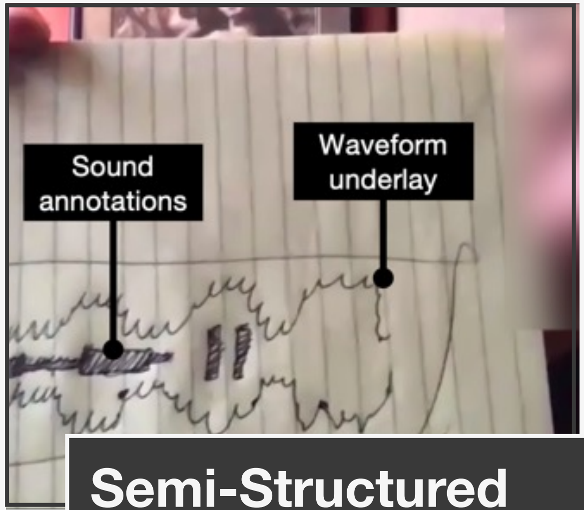
Introductory Session (75 min)

- 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.



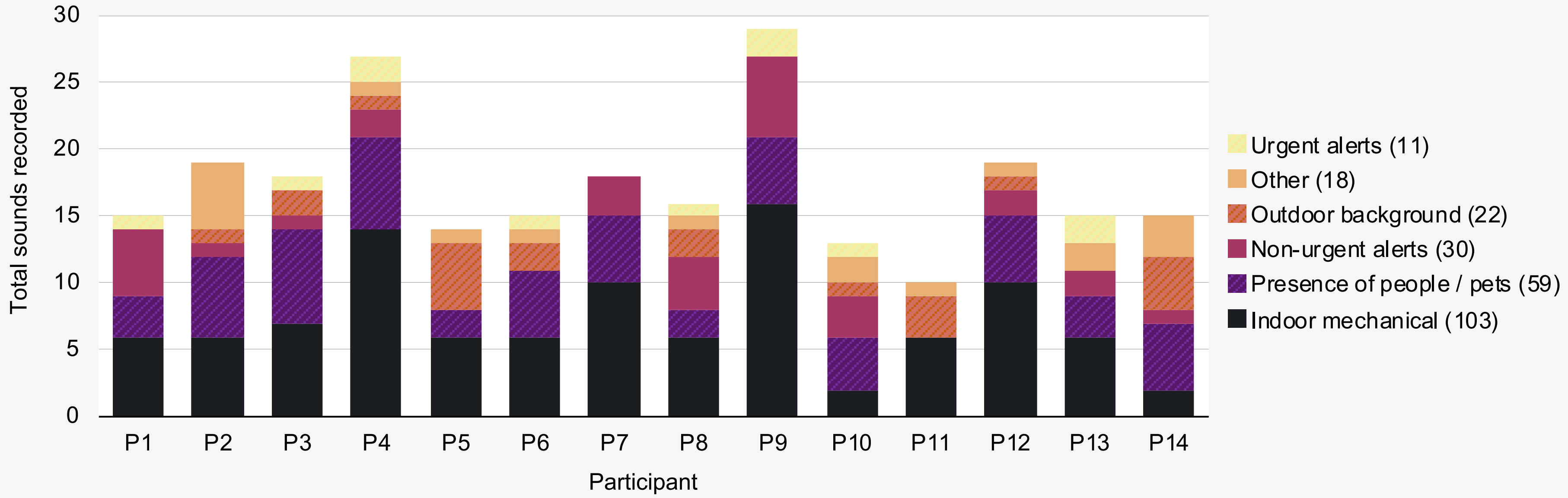
Semi-Structured Interview (60 min)

- 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: 

Key Challenge 1: Waveform Interpretation

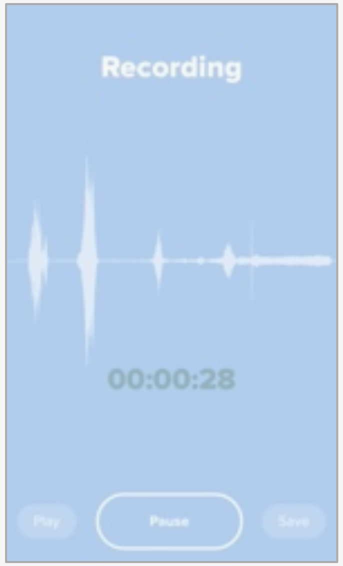
Though they were unable to hear aspects of the sound being recorded, Rev's **waveform was crucial for interpreting** the contents of samples.

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.

Thank You

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

Steven M. Goodman

smgoodmn@uw.edu
PhD Student
Human Centered Design and Engineering
UNIVERSITY *of* WASHINGTON

Co-authors:

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