“Not There Yet”: Feasibility and Challenges of Mobile Sound Recognition to Support Deaf and Hard-of-Hearing People

Jeremy Zhengqi Huang

University of Michigan, zjhuang@umich.edu

Hriday Chhabria

University of Michigan, hridayc@umich.edu

Dhruv Jain

University of Michigan, profdj@umich.edu

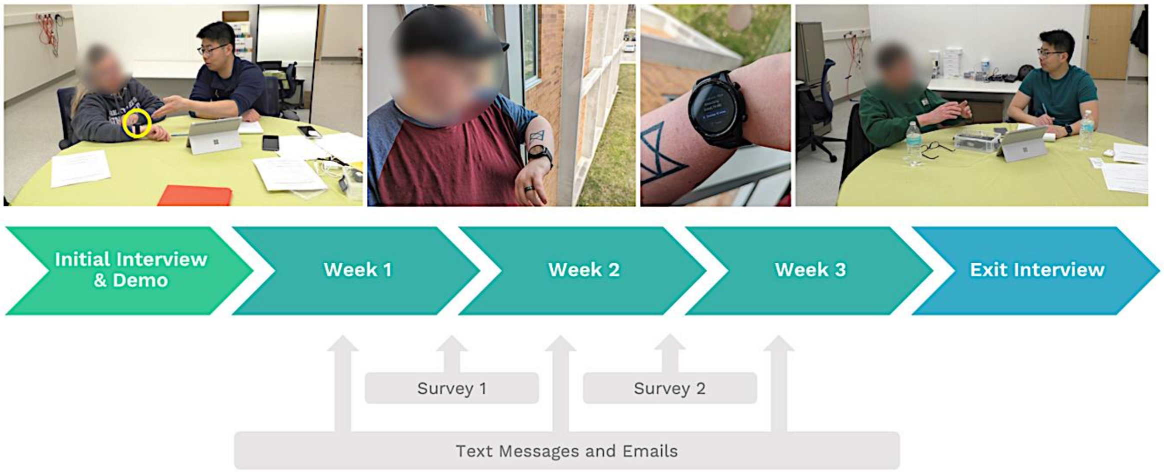


Figure 1. Pictures of participant interaction and the flowchart of our field study containing three parts: (1) an initial interview and demo, (2) three weeks of app usage, and (3) an exit interview. The yellow circle in the first picture marks the smartwatch with the SoundWatch app installed.

While recent advances have enabled mobile sound recognition tools for deaf and hard of hearing (DHH) people, these tools have only been studied in the lab or through short, controlled experiments. To assess the real-world feasibility and guide the future designs of mobile sound awareness systems, we conducted a three-week field study of *SoundWatch*, a smartwatch-based sound recognition app, with 10 DHH participants. Our findings suggest the app’s utility in increasing environmental awareness and facilitating everyday tasks for DHH users. However, several challenges, such as background noises, variability of real-world sounds, and confusion among similar sounding sounds, indicated that mobile sound recognition solutions are “not there yet” for adoption and use in daily life. We close by presenting HCI design opportunities to improve model reliability by increasing contextual awareness, supporting end-user customization, and fostering the collective improvement of sound recognition models.

CCS CONCEPTS • Human-centered computing ~ Accessibility ~ Accessibility technologies

Additional Keywords and Phrases: Accessibility, deaf, Deaf, hard of hearing, sound awareness, sound recognition, field study.

ACM Reference Format:

First Author’s Name, Initials, and Last Name, Second Author’s Name, Initials, and Last Name, and Third Author’s Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock ’18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

People who are deaf and hard of hearing (DHH) may have limited access to the sounds in their environment, potentially hindering them from effectively performing everyday tasks (*e.g.,* knowing when the washing machine is done) or being aware of their environment (*e.g.,* noticing sirens or water running). Motivated by this challenge, researchers have developed and studied systems to help DHH users interact with sound information. These systems range from early desktop-based solutions that visualized sound position [18], volume [8,14], or frequency [29] to mobile-based sound awareness tools that help DHH users localize sound [12,14], caption speech [14], or recognize sound source [3,8,12–14] in different contexts.

Evaluations of mobile sound awareness tools have been constrained to the lab or controlled environments [5,10,22]. For example, Bragg *et al.* [5] conducted a Wizard of Oz study lab study to evaluate a preliminary smartphone app that recognized two sounds (door knocks and alarm clock). Jain *et al.* [22] evaluated a smartwatch-based deep-learning sound recognition solution for a few minutes in selected contexts (an office, a building lounge, and a bus stop), quantifying performance on a dataset of sound recordings.

While these evaluations have provided valuable insights into the technical feasibility of mobile sound recognition, these short-term evaluations have limitations. First, the hand-selected “controlled” scenarios do not accurately represent the dynamic acoustic conditions of the real world (*e.g.,* a crowded restaurant *vs.* a quiet home). Second, the evaluations of system performance in these studies were primarily based on the classifications of high-quality sound effects or sound recordings and did not fully capture the variability of real-life sounds (*e.g.,* beeps from different microwaves) and ambient noise (*e.g.,* raining). Third, the controlled evaluations may overestimate the system’s robustness by overlooking user behaviors and edge scenarios with the device (*e.g.,* obstruction of the microphone by jacket sleeves). Finally, insights from short-term user studies do not reflect the long-term experience of the system, including the adoption, social implications, and the change of use over time.

To address this gap, we conducted a three-week field study to evaluate the real-world use of the mobile-based sound recognition system with 10 DHH participants. Among the possible mobile and wearable options, we selected smartwatches for our evaluation due to their availability on the wrist, which offers two important advantages: (1) obstruction-free microphone (compared to a smartphone, which may be kept in the pocket) and (2) ability to provide instant vibrational feedback for attention. Indeed, smartwatches were the most preferred sound feedback devices by DHH people in a recent large-scale survey with 201 participants [9]. Specifically, we used SoundWatch, a publicly available sound recognition app that classifies sound locally on a smartwatch and provides real-time visual and vibrational feedback about three key sound properties: sound identity, loudness, and time of occurrence (Figure 2) [22]. Since SoundWatch was released three years ago, we replaced the underlying model with the state-of-the-art sound classification model for portable devices and used the updated app for our field study.

During the study, participants provided regular feedback through emails, texts, and weekly surveys. At the end of the three weeks usage period, we additionally conducted a semi-structured interview to collect feedback on the overall experience of using a mobile sound recognition app, any observed change in sound awareness, thoughts on social and privacy implications of an always-sensing app, and any improvement suggestions.

Our findings demonstrate the mobile sound recognition solutions' capability to enhance DHH users' sound awareness and help them perform everyday tasks in diverse contexts, reaffirming the findings from previous work [9,10]. However, the app's reliability issues, especially the errors in varying acoustic contexts and ambient noises, hindered the effective adoption of this technology in everyday life. Participants indicated the need to improve the adaptability of sound recognition models and provided suggestions for interaction design ideas to mitigate the risk of AI errors. We conclude by outlining design guidelines to make future sound recognition technology more useable and adoptable.

Our work makes the following primary contributions: (1) the first longitudinal field study of a mobile sound recognition system to assess its feasibility for daily use by DHH users, and (2) design guidelines, especially for future human-AI interactive systems, to address potential errors from sound recognition technology.

2 RELATED WORK

We provide a background on Deaf culture and DHH people's sound awareness needs as well as situate our work within sound awareness technologies, other assistive technologies, and Human-AI design considerations.

2.1 Deaf Culture and Sound Awareness Needs of DHH People

The DHH community encompasses people who identify as Deaf (capital 'D'), deaf (lowercase 'd'), or hard of hearing [8,43]. Individuals who identify as Deaf embrace Deaf Culture [8,31,43] and follow an established set of norms and languages like American Sign Language (ASL). In comparison, deaf or hard-of-hearing individuals connect to hearing loss from a medical or audiological perspective and may not necessarily identify with the Deaf Culture [8,31]. These cultural differences may lead to different preferences regarding sound awareness, such as differing sounds of interest preferences [5]. For example, hard-of-hearing people may desire speech or other sounds related to human activity (*e.g.,* footsteps, doorbells) more than Deaf people [9].

While we gracefully acknowledge these differences, we also point out that, in several past studies [5,9,29]—including a recent large-scale survey with 201 DHH participants [9]—DHH people from *all* cultural groups have expressed the need for greater access to sound information in their daily life. Among the possible types of sound information, DHH people see sound identity (*e.g.,* "dog barks," "emergency vehicles") as the most desired characteristic among others (*e.g.,* sound volume and frequency) [9]. Furthermore, smartwatches were the most preferred form factor, and the combination of haptic and visual feedback was the most popular sound feedback modality [9].

Our field study builds on the above findings and extends our understanding of DHH people's sound awareness needs by gathering feedback from real-world usage of a smartwatch sound recognition system.

2.2 Sound Awareness Technologies

Early work studied stationary visualization systems for sound awareness of DHH people [18,29,30,38], including desktop monitor- [18,29] and projector-based [38] visualizations. For example, Ho-Ching *et al.* [18] designed desktop interfaces that visualize sounds with spectrographs and "positional ripples" to represent the loudness and location of the sound. Tomitsch *et al.* [38] proposed sound visualizations that utilize large-scale projections on the ceilings. Matthew *et al.* [29] evaluated several prototypes of peripheral visualizations of non-speech sounds (*e.g.,* icons, maps, and spectrograms). These solutions focused on the basic ambient sound properties (*e.g.,* loudness, pitch) but were not able to recognize and distinguish individual sound sources.

In terms of sound recognition pipelines, early methods used silence ratio [15], variations of zero-crossing rates [13,15], and shallow learning (*e.g.*, support vector machine [25] and decision tree [28]) to classify audio data. However, these limited methods used very few sound classes (*e.g.*, speech vs. music), worked only clean sound files or struggled to maintain accuracy in diverse environments. Laput *et al.* [26] developed a sound-based activity recognition system that leveraged a deep learning approach to classify sound events, demonstrating the feasibility and flexibility of deep learning-based sound recognition across physical contexts. However, this system utilized a large memory-intensive model, which was not conducive for deployment on portable devices.

More recently, researchers have begun studying mobile sound recognition systems. For example, Bragg *et al.* [5] designed a smartphone app that allowed users to record sounds and train the model on the go. Their preliminary Wizard-of-Oz study found that the user interface effectively facilitated the training process and could be potentially successful in alerting users to the surrounding sounds. Goodman *et al.* [10] conducted Wizard-of-Oz and controlled studies with 16 DHH participants to evaluate smartwatches-based sound feedback designs and demonstrated the promise of using a combination of vibrational and visual feedback for sound awareness. Specifically, visual feedback provides intuitive sound information, while haptic feedback help capture users' attention without disrupting current tasks. However, both approaches studied formative sound feedback designs that are not implemented as working prototypes; thus, they did not offer insights into the technical feasibility of recognizing sounds.

Towards working prototypes, Liu *et al.* [37] developed a smartphone-based acoustic sensing and notification app that used a lightweight deep convolutional neural network (CNN) model to enable context-independent event recognition. Similarly, Jain *et al.* [22] developed a smartwatch-based sound recognition application powered by a transfer learning-adapted deep-CNN classification model (*i.e.,* VGG). Evaluations of both prototypes demonstrated promising results, such as high responsiveness and accuracy in recognizing and notifying sound events. However, those evaluations were either short-term (*e.g.,* two-day discontinuous user tests) or controlled (*e.g.,* designated contexts), therefore offering limited insights into the real-life, longitudinal use of mobile sound recognition systems. In contrast, Jain *et al.* [21] conducted a three-week field study by deploying a sound recognition system at participants' homes. However, this approach used stationary displays and studied a single context—homes. We extend this work to mobile technologies and evaluate their real-life feasibility and utility across diverse contexts.

In summary, while prior work has studied mobile and wearable sound recognition tools for DHH users, the evaluations were restricted to lab or controlled settings and did not offer comprehensive insights into the utility of these systems for highly variable real-life use (*e.g.,* in the presence of ambient noises, differing user behaviors, and changing acoustic contexts) – a gap we address in our work.

2.3 Privacy and Social Implications of Assistive Technologies

DHH users generally considered sensing activities from assistive devices acceptable [21]. However, an interview with older adults experiencing pointing difficulties showed that users of intelligent assistive technologies might be concerned about the handling of their personal data [16,17]—a concern exacerbated by the lack of transparency in data policies [17]. This research highlighted the importance of openly communicating how data is handled and keeping the data anonymous. At the same time, some users were willing to share their personal data for the goodwill of improving the technologies, indicating a potential tension with disclosure [16]. Akter *et al.* conducted online surveys on people with visual impairments' opinions towards camera-based assistive technologies showing that, if not designed carefully, assistive technologies could be prone to improper usage that violates bystanders' privacy [1]. In the present study, we probe these privacy concerns with DHH users while evaluating a mobile sound sensing and recognition device.

Assistive technology use could also introduce social tensions [6,33–36]. For example, in a recent study, blind employees' use of screen readers during synchronous meetings was perceived as "disruptive" [6]. A large-scale crowdsourced survey with 1200 individuals argued that the use of head-mounted displays would be considered socially acceptable only if used by people with disabilities, which could cause burden regarding disclosure of disability status [33]. DHH people have also reported feeling self-conscious or sensitive to other people's perceptions when wearing hearing aids [36]. Shinohara and Tenenberg showed that these social tensions can be reduced by adopting more "socially acceptable" designs [35]. We explore these tensions in our work.

2.4 Human-AI Design Challenges and Strategies

Our work is also informed by prior work in human-centered AI, including Amershi *et al.*'s guidelines for human-AI interaction [2] and Google PAIR's guidebook for designing AI products [47]. Google PAIR guidebook specifies several types of AI errors, which include the prediction error that is closely related to our current study. Since assistive technologies are essential for some users with disability, the stakes of AI errors can be high, eliciting the need to gracefully address these errors [47]. First, AI systems should support efficient correction and teaching from end users [2,47]. For example, ImageExplorer encourages blind users to be skeptical about AI-generated captions and helps them determine their correctness by providing additional information [27]. Similarly, ProtoSound enables users to train the sound recognition system to detect specific sounds by recording their sound samples [20]. One caveat of this strategy is that some disabled/Deaf users may be unable to provide feedback [32]. For example, DHH users may not be able to effectively determine the correctness of a sound recognition model if they are unable to hear the sounds, suggesting the need for visualizations to assist them in the task [20]. Second, users should be able to invoke the AI system or disregard the AI output efficiently [2]. SoundWatch implements this guideline by enabling users to "snooze" notifications for certain sounds. Finally, AI systems should acknowledge and signal uncertainty when in low confidence [2,47]. To our knowledge, little prior work has explored this strategy in working AI systems. We explore these considerations by obtaining DHH users' feedback on two interactive design prototypes aimed at mitigating AI sound recognition errors.

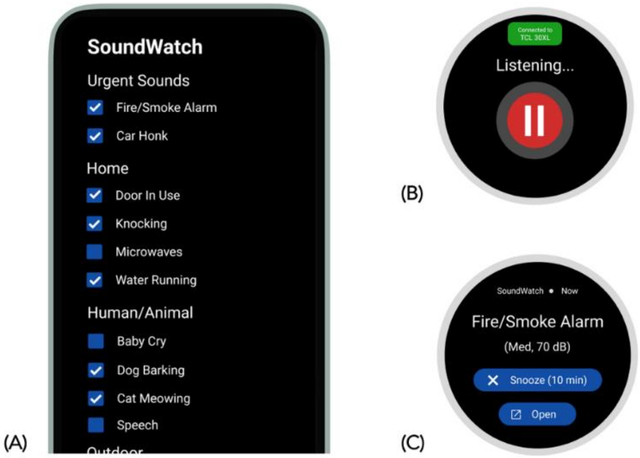
3 THE MOBILE SOUND RECOGNITION SYSTEM

Our system preference was informed by four goals: (1) portable support for diverse contexts, (2) availability of obstruction-free microphones, (3) support of both visual and haptic feedback to timely capture users' attention, and (4) glanceable and readily accessible visual display. These goals were informed by prior surveys with DHH participants [9], who desired a glanceable, always-available, and unobstructed portable sound recognition system to support their sound awareness needs across multiple contexts.

Based on the above goals, we chose the *SoundWatch* app for our study, a publicly-available Wear OS/Android sound recognition app for DHH users designed by *researchers* from University of Washington [22]. SoundWatch uses an on-device deep CNN-based model to sense and classify sound among 20 categories (*e.g.,* dog bark, door open) on a conventional smartwatch in real-time. After processing the sound, the app conveys key sound characteristics desired by DHH people: the type or sound (or the sound identity), the loudness (or volume) of the sound, and the time of occurrence through the visual display (see Figure 2). Additionally, the app informs users of the occurrence of a sound through a push vibratory notification.

SoundWatch also includes other features to support customization. For example, users can choose to snooze to sound category for a certain period in cases where they may not desire repeated notification (*e.g.,* snoozing "speech" alerts while talking to somebody). As well, a companion smartphone app allows users to: (1) disable notifications for undesired sounds and (2) set the base minimum loudness threshold (called the microphone sensitivity) for sensing sounds.

Importantly, SoundWatch's codebase is fully open-source and well-documented in the repository (github.com/AccessibilityLab/SoundWatch), which allowed us to readily extend the app to support requisite features for our field study (*e.g.,* data logging and model updates).


SoundWatch

Urgent Sounds

☑ Fire/Smoke Alarm
☑ Car Honk

Home

☑ Door In Use
☑ Knocking
□ Microwaves
☑ Water Running

Human/Animal

□ Baby Cry
☑ Dog Barking
☑ Cat Meowing
□ Speech

Outdoor

Connected to TCL 30XL

Listening...

SoundWatch • Now

Fire/Smoke Alarm

(Med, 70 dB)

Figure 2. The SoundWatch user interface showing: (a) a partial view of the paired phone app to choose the list of enabled sounds, (b) the watch app background screen when sound sensing is enabled, (b) and the notification screen with an in-built "10-minute" snooze button. Additional snooze durations are supported (*e.g.,* 5 minutes, 20 minutes, 1 hour, and forever) but are not shown above.

3.1 Selecting the Sound Recognition Model

SoundWatch was released in Nov 2020. Since then, many advancements have occurred in the field of machine learning, leading to smaller and more reliable deep-learning models. To ensure we are using the state-of-the-art model, we compared the performance of SoundWatch with two recently released mobile sound recognition systems–Google's Sound Notification App [46] and Apple iOS's built-in sound recognition feature [48]. Both Google and Apple have released the sound recognition model and the companion code publicly, which allowed us to compare performance with SoundWatch's native *MobileNetV2* CNN model.

We evaluated performance on the dataset from SoundWatch's original paper [22], which contains samples for 20 sound classes recorded from nine real-world locations (three homes, three offices, and three outdoor locations). The total dataset spans 1.5 hours and contains 540 recorded sound clips. We did not use common ML benchmarks since they contain synthetically generated high-quality sounds and do not accurately represent real-world characteristics.

For our experiment, we classified each sound using the three approaches and calculated the classification accuracy using a clip-level prediction. Our results show that the performance of the three approaches was comparable, with the Google's (*mean accuracy*=83.6%, *SD*=4.9%) fairing slightly better than SoundWatch (*mean accuracy*=81.2%, *SD*=5.8%) and Apple's (*avg accuracy*=80.3%, *SD*=6.1%). Consequently, we replaced SoundWatch's MobileNetV2 model with Google's *YAMNet* architecture [49] and used the resultant app for our field study.

4 METHODS

4.1 Participants

We recruited 10 DHH participants through study ads, snowball sampling, and emails (Table 1). The average age for the participants was 48.6 years old (*SD*=19.78, *range*=21—75). In terms of onset ages of hearing loss, three participants reported congenital hearing loss, two reported onsets at 2 years old, and the remaining five reported 6 months, 7 years, 40 years, 44 years, and 61 years. In terms of the technologies participants currently use to support sound awareness, seven reported using captioning (*e.g.,* Live Transcribe [50], Zoom's audio transcription [51]), five reported mobile apps (*e.g.,* ReSound 3D [52], InnoCaption [53]), and one reported using Google Nest Aware [54]. In terms of hearing devices, seven reported using hearing aids, and two reported cochlear implants.

PID	Gender	Age	Identity	Hearing Loss	PMoC w/Deaf/deaf	PMoC w/Hearing
P1	M	24	Deaf	Profound	Sign Language	Interpreter
P2	M	51	deaf, HoH	Severe	Verbal, Writing, Text	Verbal
P3	M	64	Deaf	Severe	Verbal, Writing	Verbal
P4	M	65	HoH	Severe	Writing	Verbal
P5	M	75	HoH	Severe	Verbal	Verbal
P6	F	60	HoH	Severe	Verbal, Sign Language	Verbal
P7	M	23	HoH	Severe (L), No (R)	N/A	Verbal
P8	M	43	deaf, HoH	Profound	Verbal	Verbal
P9	M	21	Deaf	Profound	Sign Language	Sign Language
P10	M	60	HoH	Moderate	Verbal	Verbal

Table 1. Participants' background information. PMoC w/Deaf/deaf stands for "Preferred Mode of Communication with Deaf/deaf people"; PMoC w/Hearing stands for "Preferred Mode of Communication with Hearing people."

4.2 Procedure

The study contained three parts: (1) an initial interview and system demonstration, (2) three-week use of the SoundWatch systems, and (3) an exit interview (see Figure 1). Both initial and exit interviews were conducted in the Accessibility Lab at the University of Michigan campus. All sessions were audio and video recorded. We provided the participants an option to request any disability accommodations: one participant (P3) asked for a real-time captioner, and another (P1) attended the interview with their own Personal Care Attendant and ASL interpreter. To help participants accurately understand the questions, we presented all the interview questions in a slide deck on a Microsoft Surface Go tablet.

4.2.1 Initial Session

The initial sessions were conducted by the first author. We proceeded with the sessions once we received our IRB-approved written consent form from the participants. We first asked the participants to complete a short background form to collect demographic information such as age, gender, and level of hearing loss. Then, we interviewed participants about their experiences with sounds in daily life, including their desired sounds of interest, current tools and strategies to experience sounds, and any remaining challenges in accessing sounds. Finally, we explained SoundWatch, the smartwatch-based sound awareness system, gave a detailed demonstration of the system, and handed participants: (1) a smartwatch with the SoundWatch app installed (TicWatch Pro 3 Ultra), (2) the paired Android phone (TCL 30XL), and (3) a paper manual detailing the SoundWatch user-interface and frequently asked questions. We also encourage the participants to experience SoundWatch on their own for a few minutes in our lab by producing sounds (*e.g.,* by knocking on the door or tapping) and ask any follow-up questions. We gave the participants mid-range devices for the study (TicWatch Pro 3 Ultra and TCL 30XL) because they offered real-world applicability. To detail specifically, the TicWatch Pro watch has 1GB RAM and a quad-core (four 1.7 GHz Cortex-A53) processor and lasts for 18.0 hours and 14.2 hours with and without the SoundWatch app. The TCL 30XL phone contains 6GB RAM with an octa-core (four 2.0 GHz Cortex-A53 and four 1.5 GHz Cortex-A53) CPU.

The duration of the initial session ranged from 45 to 90 minutes. The completion time differed across participants because 1) they used different communication strategies, from sign language to verbal to real-time captioning, and 2) some participants took more time to familiarize themselves with the smartwatch and the SoundWatch app. The entire session was recorded with Google Recorder app and the video camera.

4.2.2 Three-week Use of the SoundWatch System

After completing the initial interview, participants used the SoundWatch application on the watch and phone we gave for three weeks in their daily lives. At the end of weeks 1 and 2, we asked participants to complete a short online survey containing three open-ended questions on the overall experience, usage contexts, and any particularly helpful or unhelpful incidences. We emailed a link to the online survey a day before it was due. If a participant failed to complete a survey, we sent a single reminder after 24 hours. We also offered participants the option to contact us through text or email anytime for any questions, concerns, or feedback.

Within the SoundWatch app, we curated an automatic logging system that collected information about the sounds recognized, timestamps, and participants' interactions with the app. The individual logs were stored locally on participants' phones, while the aggregated logs were uploaded periodically to the Firebase server. The logging system was designed to be privacy-preserving, meaning that under no circumstances were any continuous sound information (*e.g.,* speech or sound activities) collected.

4.2.3 Exit Interview

After three weeks of app use, we invited the participants back to our lab at the University of Michigan campus and conducted another interview on their experience using the app in their daily lives. During the interview, we asked 25 questions in 10 categories (see Table 2).

Overall usage and experience. This category aimed to understand the usage pattern and overall experience of the SoundWatch app. We asked about the usage time per week, usage patterns over time, and the overall experience using the app.

Contexts of use. We asked about the contexts and scenarios where they used the app and how their experiences varied across different contexts.

Helpfulness of the app. We sought to understand the situations where SoundWatch had been helpful and the remaining challenges not addressed by the app usage.

Privacy and social implications. This category of questions aimed to understand the privacy concerns from SoundWatch's continuous sound sensing and other people's perceptions of it.

Information overload. We asked participants about the amount of information offered through push notifications.

User interface. We obtained feedback on the design of the visual display and vibration interface.

Other suggestions. We prompted participants to provide additional design suggestions to improve the app.

Misattributions. After obtaining overall feedback, we asked participants about their reactions to inaccurate sound recognition.

Evaluating alternative designs. We invited participants to share their thoughts on potential designs for future mobile sound recognition solutions. Our aim was twofold: (1) understanding DHH people's preferences for user-programmable sound recognition systems and (2) exploring interaction designs that address AI errors, including showing uncertainty and user-initiated error correction.

Overall sentiment. Near the end of the interviews, participants shared any remaining thoughts and open-ended comments about the app.

After the interview, we collected the devices back and handed participants $150 cash as research payments. The sessions lasted 55 to 70 minutes and were audio and video recorded.

Category	Num. Questions	Examples
Overall usage and experience	3	"Could you describe your overall experience with the [SoundWatch] app?"
Contexts of use	3	"In what contexts or scenarios did you use the app?"
Helpfulness of the app	2	"You mentioned [sounds] to be particularly challenging to experience. Which of those sounds did SoundWatch help with?"
Privacy and social implications	3	"How do you feel about the app continually sensing sound information around you?"
Information overload	1	"How do you feel about the amount of information the app's sound notifications provided you with?"
User interface	3	"Overall, what do you think of the interface of the app?"
Other suggestions	1	"Do you have other suggestions for improving the app?"
Misattributions	2	"What did you do when you encounter inaccurate predictions?"
Evaluating alternative designs	5	"What do you think if the app notifies you about sound like this instead [presenting designs]?"
Overall sentiment	3	"What do you think of wearables as sound awareness devices in general?"

Table 2. Exit interview questions and categories.

4.3 Analysis

We analyzed the survey responses, interview transcripts, and texts/emails through a combined thematic analysis by treating each question or text/email thread as a separate unit. Specifically, we used Braun and Clarke's six-phase approach [7]. The first author skimmed the transcripts to familiarize with the data (step 1) and discussed with the research team to generate an initial codebook (step 2). The researcher then iteratively applied codes to the data while refining the codebook. The final codebook had a 3-level hierarchy: 8 first-level, 29 second-level, and 80 third-level codes (step 3). Another researcher used this final codebook to independently code all data (step 4). We then calculated IRR (interrater reliability) between the two coders using the ReCal2 package [55] and resolved disagreements via consensus among our research team. The average Krippendorff's alpha value was 0.65, and the raw agreement was 83.1%. Finally, we organized the themes into subsections (step 5) and formed our narrative (step 6). We have attached the final codebook as supplementary material.

For the automatic quantitative log data, we calculated the average number of sound events per day throughout the three weeks and the total occurrence for each sound type.

5 FINDINGS

Our findings detail usage patterns of the SoundWatch app, observed errors in sound recognition, user interface suggestions, and privacy implications. We also elicit participants' ideas for improving mobile sound recognition in the future, including feedback on two mid-fidelity design prototypes to mitigate AI errors.

These findings represent data collected from (1) the two interviews and two weekly surveys completed by our participants, (2) an additional 16 email threads and 23 text messages initiated by our participants to provide more feedback, and (3) automated logs from the SoundWatch app. Quotes are verbatim from participants' responses but lightly edited for grammar.

5.1 Usage Patterns

Participants reported that they wore the watch consistently, except while sleeping and in situations where potential collisions with hard surfaces might happen (*N*=2) (*e.g.,* construction and warehouse). In addition, P10 did not wear the watch on the weekends. In terms of location, participants used the app at home (*N*=9), in social situations (*e.g.,* family dinner, visiting friends; *N*=6), in vehicles (*N*=6), at workplace and school (*N*=6), in commercial spots (*e.g.,* restaurants and stores; *N*=5), and in-transit (*N*=3). Table 3 details the specific scenarios of use in each location.

Context	Counts	Example Scenarios
Home	9	Living room, bedroom, kitchen
Social	6	Family dinner, visiting friends
Vehicle	6	Cars
Work/School	6	Classroom, office, warehouse
Outdoor	5	Hiking trails, walking on campus
Commercial	5	Restaurants, shopping malls
Crowd	4	Sports events, conferences
Transit	3	Bus, plane

Table 3. Participants' reported locations of use with examples scenarios.

SoundWatch provides an option to select/deselect among a list of sounds to enable/disable them from recognition. Almost all participants (*N*=8) enabled all the sounds listed. The other two participants disabled the bird chirping sound because of the frequent false positives. From the automatic logs, we found that the app recognized 112.7 sound events per day across all participants (*SD*=60.7) (169 on average (*SD*=79.9) for participants who disabled some sounds). Among the specific sound types, bird sound was recognized most frequently, closely followed by water running and dog bark.

Participants reported that they generally paid attention to every sound notification early in the study. For example, P4, P6, and P8 reported that, even if the prediction was inaccurate, a notification from the watch could indicate a sound occurring, and they would still try to locate the sound source. P6 explained:

> *"I would try to figure out what was making the sound… Sometimes it might be just false positives. But I never assumed it was nothing."*

However, towards the end of the study, some participants stopped paying attention, either because they became familiar and desensitized to the notifications (*N*=3) — such as P6 who reported that they "dismissed" or "ignored" unimportant sound feedback after several days of use or P7 who reported that when he felt a vibration on the street, he would *"automatically assume that it was vehicles passing"* — or because they got annoyed with the frequent errors (P9). We describe these errors in the next section.

5.2 Feedback on Sound Recognition

There were some impactful cases where SoundWatch helped participants monitor the state of the environment (*e.g.,* noticing their children opening the doors) and performing everyday tasks (*e.g.,* noticing incoming while walking on the streets). However, all participants (*N*=10) reported errors from sound feedback, with five explicitly stating that the app was not ready for daily use or long-term deployment due to several errors. Those errors broadly fell into four categories: false positive, false negative, misattribution, and registering background noise (*e.g.,* rain, TV sounds).

The most common error category was the false positives (*N*=6). Most of the false positives were about animal sounds. For example, among all sounds, birds got triggered the most frequently as the sound prediction (*N*=4); three participants reported that clothes like jackets and shirts can trigger bird predictions. This might be because of the resemblance between sound patterns of birds chirping and fabrics rubbing against other surfaces. Similarly, P2 reported receiving "duck/goose" predictions while driving on the highway, which might again stem from the similarity between car sounds (*e.g.,* engines) and duck quacks, suggesting a need for future algorithms to distinguish among similar sounding sounds.

For some participants, the frequent false positives were especially concerning because they caused desensitization to the app's notifications, potentially risking missing important sounds. P4 explained:

> *"When I got wrong predictions every time, I just learned to ignore it. But it's like the boy that cried wolf… what if next time, you ignore the [notification] and realize that, oh shoot, I should have paid attention."*

For P9, the false positives were so annoying that they reduced the usage on the last week of the study due to the "growing frustration" and "burnout."

The second most common category was the false negatives (*N*=4), where SoundWatch failed to recognize sounds enabled by the participants. Some of the missed sounds were critical. For example, P1 and P9 mentioned that emergency vehicles were sometimes not recognized on their way to school or work, possibly due to the environmental noise and the threshold settings for loudness (the lowest threshold for SoundWatch to recognize sounds is 40 dB). While we especially explained to the participants that SoundWatch should not be relied on in critical safety-related situations, accidental overreliance on the technology in the future could be dangerous and careful guidelines are needed before wide deployment. Other reports of false negatives concern everyday sounds that may be unique to their environments. For example, P4 reported that door-knocking was *"hit-and-miss."* This inconsistency in recognition might be caused by the variability in the materials of the doors. Similarly, the app's failure to recognize doorbells (P2, P4) and appliances like microwaves (P4, P8) might be caused by the variability in the makes or models of appliances.

Another common error (*N*=3) was misattribution (classifying a particular sound as another). P3 and P5 reported that the system frequently registered beeping or buzzing sounds (*e.g.,* microwaves, medical devices, and basketball court buzzers) as fire or smoke alarms. While P5 was able to verify that there was no fire alarm present, P2, who identified as Deaf, was confused by the sound feedback:

> *"I was in the hospital… and am hooked up to a heart monitor. And I received a notification from the system. I was like, 'Wait a second, a smoke alarm?'"*

This experience demonstrates the risks of misattributions in critical settings and the need for the graceful handling of AI errors in sound recognition. Other cases of misattributions involved human speech and activities. For example, P7 reported that lecturers were registered as "dog barking." P10 encountered similar errors, saying the system would recognize the chat with colleagues as "duck/goose." However, interestingly, he thought these errors were more "entertaining" than alarming, as he would joke with his colleagues to "*check if any goose is paddling around.*" One possible way to address this issue is by implementing contextual awareness. For example, if the system recognized that users were in a hospital environment, it could temporarily tune up the confidence threshold required to report a fire/smoke alarm.

Despite many reported errors, participants did indicate that SoundWatch has been helpful for them in many ways. Half of the participants (*N*=5) reported that the app helped increase the awareness of their environments—for example, by helping them stay aware of their children's movements in the house through "door-in-use" and "footsteps" notifications (P8) and noticing the bird chirping on hiking trails (P3). Critically, the vehicle sound feedback helped P7 detect incoming traffic. He explained:

> *"Things like vehicles are actually very helpful, especially when I am outside… it's like shooting me a message and letting me know that there is a car coming."*

SoundWatch also helped participants perform everyday tasks. For example, the system reminded P8 of the running water in the kitchen sink. It also helped P6 and P7 notice door-knocking in hotel rooms and at home. P8 added that while the system is "not perfect," it was "generating an overall awareness that something was happening." These useful cases suggested that if the errors described previously were fixed, the system would have the potential to help DHH people monitor their environments (*e.g.,* watching the kids), performing everyday tasks (*e.g.,* noticing the door knock), and keeping them safe by notifying them of critical events (*e.g.,* incoming traffic). Indeed, seven participants explicitly mentioned that this technology holds many promises in the future. For example, P10 said:

> *"There is no question in my mind that the idea of having a wearable device being able to help people who can't hear things is very beneficial. My phone flashes every time I get a text message to help me notice it… this is like that."*

However, until these errors are completely fixed, system guidelines should carefully indicate that the system can fail and should not be relied on for safety-critical situations (*e.g.,* fire alarms, sirens).

5.3 User Interface

While the participants' sentiments on the SoundWatch's sound recognition engine were mixed, they generally (*N*=7) liked the design of the SoundWatch interface, saying it was *"intuitive"* and *"simple to use."* Below, we describe their comments on the app's visual display and haptic feedback.

5.3.1 Visual display

SoundWatch delivers sound feedback through push notifications on the watch. The sound feedback includes sound identity, loudness, and time of occurrence (see Figure 2). The notification UI also allows users to snooze the sounds for a period. Most participants (*N*=7) approved of the design of the visual feedback. For example, P8 thinks the amount of information is *"spot on."* P7 echoed: *"Everything basic is on the screen."* Both P6 and P8 liked the display of loudness information. For example, P6 said:

> *"The decibel is very useful when I am not wearing hearing aids… it gives me a sense of if it's something I should be concerned about."*

While P4 shared similar sentiments about the visual display, he also mentioned that the UI fonts could be challenging to read for someone who uses corrective glasses, suggesting the need to be able to customize the display:

> *"It would be nice if we can increase the font size… I would like the ability to make the [notification] a little bigger so I don't have to grab my glasses."*

Additionally, five participants desired more information, like the sound direction, to help locate the sound source. P6 explained:

> *"If I see water running, I will be wondering, where did the sound come from? So, you get to walk over all sources of water to locate it."*

P7 considered sound direction a more important characteristic than loudness, which differs from P8, who thought the loudness information was more helpful. This diversity in information preferences again points to the need for allowing customization to suit individual users' UI needs. Three participants mentioned that the sound identity displayed on the watch was too generic. For example, P4 stated:

> *"When the "vehicle" popped up on the watch, I don't know what it was pointing to… was it the water running, or is it the beeping sounds from the truck backing up? I had no idea."*

Future improvements can address this through more granular descriptions of sound objects, such as involving behaviors of the sound object (*e.g.,* changing "vehicle" to "vehicle passing").

5.3.2 Haptic Feedback.

While some participants appreciated the vibrational feedback about sound events (P6, P8), others expressed some frustrations. For example, five participants mentioned that the vibration is too subtle or short despite configuring the vibration intensity to the highest in system settings. P9 mentioned: *"[The vibration] could be a bit strong honestly… I have missed some notifications because of that."* P1 reported the same experience.

In addition to configuring vibration intensity and duration, P3 suggested giving users the ability to customize the vibration patterns to differentiate sound events with different levels of urgency:

> *"There could be different vibrations for different kinds of things… there could be light vibrations for sounds that were nice to know, like 'Oh, there is a bird on the tree,' and stronger vibrations for real urgent stuff like large vehicle approaching from behind."*

5.4 Privacy

During the initial interview, we explained the privacy-protecting measures, including local on-device processing and storage of sound information, and that only the aggregated data will be uploaded to the cloud (*i.e.,* Firebase). After three weeks of system use, most participants (*N*=7) reported no privacy concerns, but some understood that other people's potential concerns:

> *"I could see someone being sensitive to it… you know, they might think the microphones are spying on them." (P3)*

Interestingly, while P8 reported no privacy concerns, he mentioned, "*I've got Google Nests all around my house,"* saying that smart devices sensing information is *"inevitable."* This indicated an alarming lack of end-user agency in managing privacy.

Despite general acceptance of the system, three participants expressed privacy concerns at some point during the study. P2 stated: *"I think privacy is something that came to my mind, but it was not something that I let overrun."* This suggests that the perceived benefits may outweigh the potential privacy risks. P4 explained that most of his privacy concerns stemmed from *"not knowing how the system uses the data it gathers,"* indicating users' inherent skepticism toward novel technologies and the need for more system transparency beyond verbal explanations. This reflection was echoed by P7, who learned about the technical background of the SoundWatch system:

> *"I was a little apprehensive at first, but because of the fact that I know how the system works under the hood and that it cannot necessarily get the speech itself from the features you extract, I was fine with it."*

P7 added that he would *"take the watch off and keep it inside the drawer"* when the friends came over but stopped doing it over time. Indeed, most participants (*N*=7) reported no concerns from other people regarding the app's continuous sensing of sound information—although some close family and friends had initial questions about the sound sensing pipeline. These concerns were mitigated after participants explained how the system works, indicating that proper disclosure and transparency on system behavior is crucial in gaining social acceptability.

5.5 System Improvements

We present participants' design ideas for improving the app as well as their feedback on the designs that address AI errors and allow user-programmable sound recognition.

5.5.1 Design Ideas from Participants

During the interview, participants offered concrete design suggestions to improve mobile recognition systems. The above sections mentioned some of them, including sound localization and customized vibration patterns. Here, we present other suggestions.

Logging. Five participants suggested logging past sound events for retrospective viewing. While the smartwatch operating system's notification center could store recently recognized sounds, participants desired the ability to retrieve sound predictions *"further back"* (P5) and provide feedback to inaccurate predictions *"from memory"* (P9). Moreover, P6 said:

> *"Sometimes I would ignore [the notifications] when I was busy… I wish I could play back the sound, but louder, so you can hear it and figure out what it is."*

Logging sound events is not a novel suggestion. For example, Jain *et al.* [21], for example, designed an IoT system that visualized sound activities sequentially, which provided DHH users with insights into the behaviors around the house. However, they visualized sound history on a bigger tablet display, and it will be interesting to explore visualizations for a small smartwatch display or the paired smartphone app.

Customizable and adaptive systems. Participants desired more customizability and adaptability of the sound recognition system. For example, P4 suggested granular controls on the *base sound level*, a SoundWatch feature that allows users to adjust the loudness threshold for reporting sound:

> *"It would be nice if certain sounds like the trigger point can be individually [configurable] based on which sound, because if you move the [base sound level] setting up to 50 decibels, it affects everything, and it takes away my ability to hear all these other sounds."*

Other participants suggested adaptive features like prioritizing sounds based on contexts (*N*=2). For example, P5 suggested that the system can *"enable and disable sounds based on different settings."* P5 added that this idea was inspired by the adaptive feature on the hearing aids, where multiple *"preset programs"* are available for different environments.

5.5.2 Participant Evaluations of Human-AI Sound Recognition System Designs

To explore future designs of mobile sound recognition systems, we sketched a storyboard (Figure 3) and two mid-fidelity prototypes (see Figure 4) regarding future iterations of the SoundWatch app and asked participants for feedback.

The storyboard was about user-programmable sound recognition systems and was inspired by emerging research [20,42] and participants' reports of misattributions from the weekly survey. In the storyboard, a DHH user ("Amanda") trained a mobile sound recognition system to recognize her dog "Jessie" by recording sound samples of "Jessie" barking. Before we presented the storyboards, most participants had already suggested this feature during the interview. Their motivations for this suggestion came from their encounters of sounds not being recognized (*e.g.,* microwaves, doorbells, washing machines), remaining challenges not addressed by the system (*e.g.,* name-calling), and that the current sound categories were *"limited"* (P9). For example, P7 explained his suggestions about recognizing names:

> *"Someone calling my name is probably one of the most important things… and I noticed that people from different nationalities pronounce it differently. So it will be really helpful if I can record my friends calling my name, and the watch can let me know when they do."*

All participants (*N*=10) appreciated the ability to personalize the system to recognize sounds in their environments. For example, P6 explained*: "Everybody's home and work situations are different, so I can see it being very helpful.*"

Despite the general approval of this approach, participants brought up scenarios where recording sounds can be challenging, including encountering sounds that are not easily triggered (P8) or repeatable (P7), like fire alarms and dog barks. This concern is also reflected in prior work [20].

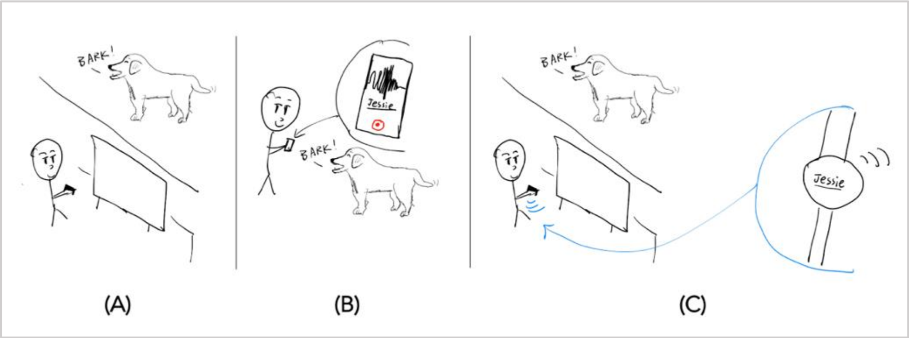


Figure 3. The storyboard for user-programmable sound recognition.

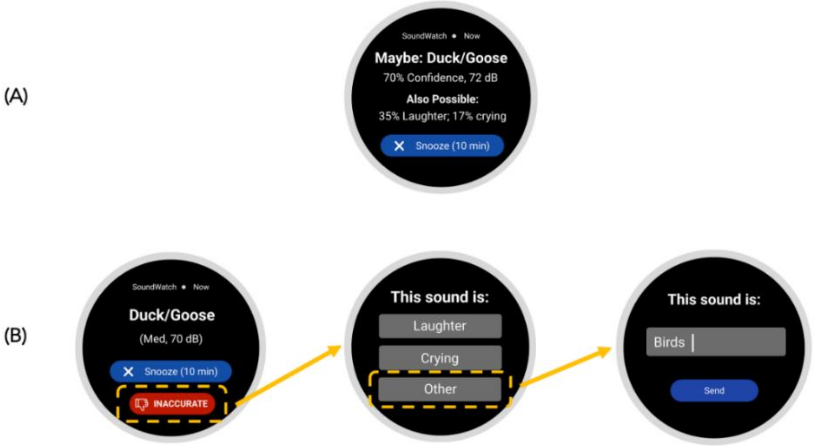


Figure 4. The two prototypes for Human-AI interaction designs. Prototype (A) acknowledges uncertainty by showing multiple possible predictions. Prototype (B) allows users to report and correct AI errors using the "inaccurate" button.

We then presented participants with two prototypes. The prototypes were about displaying the model's uncertainty information and supporting corrections of AI errors, both inspired by the Human-AI interaction guidelines proposed by Amershi *et al.* [2] and Google PAIR [47].

Showing uncertainty (A). The Human-AI interaction design guidelines suggest AI systems to "degrade AI system's services" when facing uncertainty [2]. The current SoundWatch system only reported one prediction that surpasses the confidence threshold (50%) for each sound event. In comparison, the prototype (Figure 3) acknowledged the potential errors the model makes by adding "Maybe" to the most confident predictions and other possible predictions. Most of the participants (*N*=7) appreciated this design, saying that it *"shows more options"* (P9), *"sparks curiosity"* (P9), and *"admits the system's own limitations"* (P10). However, two participants questioned its real-life effectiveness. For example, P5 thinks this design may report "three wrong predictions instead of one." P4 echoes the statement, stating that:

> *"At the end of the day, there still can be wrong predictions, and I am not able to do anything about it… so is it really useful?"*

Moreover, P9 thinks that showing the design *"takes up too much reading,"* indicating that balancing information display on mobile devices is essential to avoid information overload.

Error reporting and corrections (B). In the second prototype, we add a button that allows DHH users to report and correct inaccurate predictions. After users press the button, the interface will list two "runner-up" prediction results. Users can select either, but if neither result is correct, they can press "other" and enter the correct sound identity using the watch's input methods (*e.g.,* voice input or keyboard). Three participants mentioned this potential feature before we presented them with the storyboard, which is not surprising considering the frequent errors in our sound recognition model. Nine participants approved the design, and three preferred it over the previous one (showing uncertainty information). For example, P3 appreciated the added interactivity with sound feedback and collaboration with the AI, stating:

> *"It makes the deaf person feel like they are being heard and seen. By having them input what is important and teaming up with the device … that whole scenario is very empowering."*

Inspired by the design, P6 proposes a collaborative approach that involves other people to help assess and correct the sound recognition results:

> *"I wish that if I don't know what the sound is, I can just record it and send it to somebody else and have them figure it out… If we have a group with everybody doing this, it would be a lot more intelligent than having each person doing their own."*

Participants' willingness to report and correct errors indicates a strong potential for implementing reinforcement learning [23] into sound recognition systems by allowing users to report and correct misattributions. Moreover, as P6 suggested, this process can happen both collaboratively. For example, P9 described an incident where SoundWatch gave "duck/goose" feedback while driving and asked their partner to validate it. However, participants raised some concerns about this approach. A common concern was the effectiveness of the feedback (P5, P9, P10). Specifically, they were concerned about the excessive amount of feedback the system needs to reach a satisfactory level of sound recognition performance. This suggests the importance of implementing effective *and* efficient interaction paradigms for reinforcement learning. For example, P9 said:

> *"If I had to do more than two or three [corrections] an hour, I think that by day two, I will just give up using this feature."*

Another concern lies in scenarios where it may be challenging (*e.g.,* lifting heavy objects) or inappropriate (*e.g.,* meetings) to interact with the mobile device and give feedback (P6). This concern reaffirms the need to allow DHH users to review past sound events through features like event logging.

Interestingly, P7 and P9 proposed a combined approach, where the sound feedback will show uncertainty information and allow DHH users to report and correct errors. Both participants recognized the risk of information overload with this approach, but P9 sketched an alternative smartwatch user interface where the runner-up predictions were visualized as buttons around the main prediction, and DHH users could provide feedback by pressing the one that represents the correct sound identity ("Laugh" and "Crying") or "Others" (O), thereby enabling faster interaction by effectively decreasing the user interaction steps for the second approach. (See Figure 5).

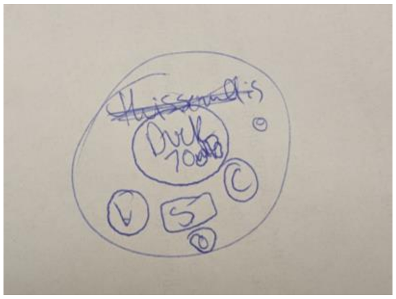


Figure 5. P9's sketch for a sound notification. The largest button ("Duck") at the center represents the main prediction and is surrounded by smaller buttons that represent other possible sounds ("Laugh" and "Crying"). The "S" button is for snoozing, and pressing the "O" (Other) button prompts users to input the correct sound identity if it is not listed.

6 DISCUSSION

Though prior evaluations of mobile sound recognition systems [22] provided insights into their short-term technical feasibility, we conducted a three-week field study of the SoundWatch system [22] to evaluate its long-term use and integration with DHH people's everyday lives. Some of our results contextualize prior findings—such as on social implications and user interface preferences [10,22]. We also report on new findings that can only emerge from an "ecologically-valid" longitudinal field study, such as the real-world utility of mobile sound recognition systems, usage patterns over time, the effect of privacy across different stakeholders, and implications of sound recognition in different contexts. We also delve into the design opportunities of mobile sound recognition systems and examine the technical and ethical issues that may arise during the deployment of this technology. Below, we discuss further implications of our findings and outline the limitations of our study.

6.1 Utility of Mobile Sound Recognition System

We used SoundWatch to represent mobile sound recognition systems inspired by DHH people's preference for using smartwatches for sound feedback [9], and evaluated its feasibility and utility in real-life through a three-week field study. Overall, the app's perceived utility was mixed. On one hand, the system demonstrated its capability to enhance DHH users' situational and environmental awareness, and—with increasing familiarity with the system—DHH users could potentially deduce sound events from contextual information (*e.g.,* interpreting footsteps prediction as children's movements around the house). Similar to the prior work about an in-home sound recognition application [21], SoundWatch helped DHH users perform everyday tasks, such as noticing door knocks or the water flowing down the kitchen sink. The current system extends this utility to mobile settings (*e.g.*, detecting incoming traffic). However, these benefits are limited to situations when the sound classification is accurate or, at the very least, predictable. This is evidenced by insights from both usage patterns (*e.g.,* the decline of system usage over time) and participants' testaments that they will strongly consider using SoundWatch daily only if the reliability issues of the sound recognition pipelines are addressed.

6.2 Challenges and Opportunities in Mobile Sound Recognition Systems

For the field study, we leveraged Google's open-source pre-trained model, a state-of-the-art approach, to classify sounds. While the controlled technical evaluation of the SoundWatch system demonstrates high accuracy, the system's inadequate real-world performance indicates the need for continued efforts on improving sound recognition in diverse contexts. Based on participants' feedback, we point out four technical challenges in developing reliable mobile sound recognition systems:

1. **Background noise.** Background noises like traffic, wind, or speech can influence the system's ability to recognize target sounds accurately.
2. **Variations in sound sources.** Diversity in sound sources, like doors made with different materials and beeps of microwaves from different brands, can lead to variations in sound characteristics like frequency, making it more difficult for the system to recognize accurately.
3. **Similar sounds.** The system may have difficulty differentiating among similar sounds without other information, such as visual inputs and contextual awareness. For example, if the system cannot recognize the medical setting, the beeping from medical devices may be identified as microwaves.
4. **Unexpected or rare sounds.** Real-life situations will involve sounds not included in the training and development of sound recognition pipelines, causing false negatives.

Informed by the above-listed challenges and participants' feedback and design suggestions, we identify four design opportunities for a reliable and robust mobile recognition system:

1. **End-user customization.** Participants' suggestions on customization-related features, such as distinct vibrational patterns for different sounds and adjusting UI fonts, reiterated the user needs reflected in the prior evaluation of the SoundWatch system [22]. Future work should expand the system's customizability by allowing granular controls on the system's sound sensing, user interface, and haptics. In terms of sound sensing, DHH users should be able to adjust the microphone sensitivity based on individual sounds and select different sound-sensing pipelines based on their needs. In terms of UI and haptics, the system should allow users to personalize the visual display of sound information (*e.g.,* replacing loudness with sound direction), adjust the intensity or duration of the vibrations, and assign vibrational patterns to certain sounds for quicker reactions (as explored by Goodman *et al.* [10]). Moreover, future designs should encourage creative customizations (*e.g.*, color rings for different sounds) since they can enable self-expression and motivate the active adoption of the technology [34].
2. **Context awareness.** Participants' frustration regarding the accuracy of sound recognition included the sound feedback that did not match the context (*e.g.*, the medical device beeps recognized as microwave). Even though SoundWatch allowed users to enable or disable sounds manually, this mismatch indicates a need for embedding context-aware capabilities in mobile sound recognition systems to automate or minimize the end-user effort to adapt the system to the environment. Indeed, as suggested by Wobbrock *et al.*'s influential ability-based design principles [41] future designs of mobile sound recognition systems should be responsive to users' environments.
3. **Graceful handling of AI errors and limitations.** Motivated by the frequent sound feedback errors participants reported and inspired by prior design guidelines of AI systems [2,47], we proposed three future-iteration designs of SoundWatch and asked participants for their opinions. The result suggested that future work should design and foster efficient interactions between DHH users and sound recognition pipelines by acknowledging system limitations and involving user input to report and correct AI errors while minimizing interruptions of current tasks. Moreover, inviting inputs from DHH people can also elicit a sense of empowerment and agency. However, supporting DHH users' interactions with AI is challenging because, as Goodman *et al.* [11] suggest, DHH people may not be able to effectively record or assess sounds and, as non-experts, may lack the knowledge of how AI will behave with their not-so-ideal sound samples

[24,39]. While prior work proposed several scaffolding techniques like visualizations (*e.g.,* waveforms and spectrograms) to address this challenge [11], future research should explore their feasibility on wearable interfaces.

4. **Encouraging collective information access.** Mobile sound recognition systems like *SoundWatch* foster a linear relationship where the system acts as a "messenger" between DHH people and the environment to enhance independence through sound information access. While prior work recognized the contributions of assistive technologies in supporting independent living [3,4], some disability studies scholars were skeptical about the concept of "independence" and "self-reliance" because they underscored the importance of community support and collective efforts to information access (*e.g.,* "interdependence" and "access intimacy") [4,44,45]. Furthermore, recent HCI work demonstrated that *interdependence* could be a valuable framework that guides the design of assistive technologies [4,40]. This value was reflected by both P6's suggestions of leveraging collaborative labeling of sound information among trusted ones and P9's case of validating sound feedback with the partner.

6.3 Privacy and Social Implications

While participants and people around them generally accepted the system's sound-sensing behaviors, some notable tensions emerged. First, SoundWatch introduced social tensions, reinforcing prior findings [6,33,35]. For example, P7 took off the smartwatch and put it in the drawer when his friend came over. Second, the acceptance of SoundWatch could have been partially driven by participants' risk-benefit assessments. For example, P2 thought about the app's potential privacy intrusions but later decided to be discouraged from using it. This contrast between privacy concerns and the perceived benefits of adaptive assistive technology is also reflected in prior work [16]. Finally, DHH users' inherent skepticism towards SoundWatch's handling of sound information suggests that future work should consider designs that facilitate the disclosure of system use and offer an easy-to-access interface to manage privacy settings. Future designs can also apply sensing pipelines that minimize the possibility of tracing back the original sound information collected from the users. For example, Iravantchi *et al.* [19] explored using inaudible frequencies above 20KHz to recognize sounds from everyday objects, showing promising results. Employing privacy-oriented pipelines like this may help lower DHH users' concerns about using mobile sound recognition systems in daily life.

6.4 Study Limitations

Our study has several limitations. First, our insights are drawn from participants' self-reported perspectives, and we do not have quantitative data on how well the sound classification worked. While the participants' detailed qualitative accounts demonstrated the perceived utility of the system, future work should also quantify field performance and corroborate our qualitative results. Second, our system only conveyed the sound identity (and other simple characteristics such as loudness). Future work should also explore designs for other complex sound properties, such as conveying source location, along with sound recognition to see if they help better discern sounds. Finally, we considered DHH participants (who may identify as deaf, Deaf, or hard of hearing) as a homogenous group while reporting our findings since past work [5] shows that these groups, despite their cultural differences, have synergetic access needs and preferences. Recruiting cross-culturally allowed us to explore mobile sound recognition with diverse users. Nonetheless, future work should examine how preferences may vary with DHH culture and hearing levels.

7 CONCLUSION

In this paper, we conducted a field evaluation of *SoundWatch*, a smartwatch-based sound recognition system, to understand the real-life feasibility of mobile sound recognition technology and identify future design opportunities. Our findings demonstrate the system's value in generating awareness for DHH users but also surface the challenges of designing and developing robust, reliable, and privacy-preserving sound recognition pipelines for mobile devices. We present design opportunities to address this challenge.

Acknowledgment

We acknowledge the contribution of people from Professor Dhruv Jain's previous lab at the University of Washington, where the initial work for this research was performed, particularly Professor Jon E. Froehlich and Leah Findlater.

References

[1] Taslima Akter, Tousif Ahmed, Apu Kapadia, and Swami Manohar Swaminathan. 2020. Privacy Considerations of the Visually Impaired with Camera Based Assistive Technologies: Misrepresentation, Impropriety, and Fairness. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Virtual Event Greece, 1–14. DOI:https://doi.org/10.1145/3373625.3417003

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300233

[3] Mirza Mansoor Baig, Shereen Afifi, Hamid GholamHosseini, and Farhaan Mirza. 2019. A Systematic Review of Wearable Sensors and IoT-Based Monitoring Applications for Older Adults – a Focus on Ageing Population and Independent Living. *J. Med. Syst.* 43, 8 (August 2019), 233. DOI:https://doi.org/10.1007/s10916-019-1365-7

[4] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Galway Ireland, 161–173. DOI:https://doi.org/10.1145/3234695.3236348

[5] Danielle Bragg, Nicholas Huynh, and Richard E. Ladner. 2016. A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Reno Nevada USA, 3–13. DOI:https://doi.org/10.1145/2982142.2982171

[6] Stacy M. Branham and Shaun K. Kane. 2015. The Invisible Work of Accessibility: How Blind Employees Manage Accessibility in Mixed-Ability Workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*, ACM Press, Lisbon, Portugal, 163–171. DOI:https://doi.org/10.1145/2700648.2809864

[7] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide*. SAGE Publications.

[8] Anna Cavender and Richard E. Ladner. 2008. Hearing Impairments. In *Web Accessibility*, Simon Harper and Yeliz Yesilada (eds.). Springer London, London, 25–35. DOI:https://doi.org/10.1007/978-1-84800-050-6_3

[9] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300276

[10] Steven Goodman, Susanne Kirchner, Rose Guttman, Dhruv Jain, Jon Froehlich, and Leah Findlater. 2020. Evaluating Smartwatch-based Sound Feedback for Deaf and Hard-of-hearing Users Across Contexts. In *Proceedings of the 2020 CHI*

Conference on Human Factors in Computing Systems, ACM, Honolulu HI USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376406

[11] Steven M. Goodman, Ping Liu, Dhruv Jain, Emma J. McDonnell, Jon E. Froehlich, and Leah Findlater. 2021. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2 (June 2021), 1–23. DOI:https://doi.org/10.1145/3463501

[12] Benjamin M. Gorman. 2014. VisAural:: a wearable sound-localisation device for people with impaired hearing. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility - ASSETS '14*, ACM Press, Rochester, New York, USA, 337–338. DOI:https://doi.org/10.1145/2661334.2661410

[13] Fabien Gouyon, François Pachet, and Olivier Delerue. 2000. ON THE USE OF ZERO-CROSSING RATE FOR AN APPLICATION OF CLASSIFICATION OF PERCUSSIVE SOUNDS. (2000).

[14] Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-mounted Display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Virtual Event Greece, 1–4. DOI:https://doi.org/10.1145/3373625.3418031

[15] Guojun Lu and T. Hankinson. 2000. An investigation of automatic audio classification and segmentation. In *WCC 2000 - ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*, IEEE, Beijing, China, 776–781. DOI:https://doi.org/10.1109/ICOSP.2000.891627

[16] Foad Hamidi, Kellie Poneres, Aaron Massey, and Amy Hurst. 2018. Who Should Have Access to my Pointing Data?: Privacy Tradeoffs of Adaptive Assistive Technologies. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Galway Ireland, 203–216. DOI:https://doi.org/10.1145/3234695.3239331

[17] Foad Hamidi, Kellie Poneres, Aaron Massey, and Amy Hurst. 2020. Using a participatory activities toolkit to elicit privacy expectations of adaptive assistive technologies. In *Proceedings of the 17th International Web for All Conference*, ACM, Taipei Taiwan, 1–12. DOI:https://doi.org/10.1145/3371300.3383336

[18] F Wai-ling Ho-Ching, Jennifer Mankoff, and James A Landay. Can you see what I hear? The Design and Evaluation of a Peripheral Sound Display for the Deaf.

[19] Yasha Iravantchi, Karan Ahuja, Mayank Goel, Chris Harrison, and Alanson Sample. 2021. PrivacyMic: Utilizing Inaudible Frequencies for Privacy Preserving Daily Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–13. DOI:https://doi.org/10.1145/3411764.3445169

[20] Dhruv Jain, Khoa Huynh Anh Nguyen, Steven M. Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehlich. 2022. ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–16. DOI:https://doi.org/10.1145/3491102.3502020

[21] Dhruv Jain, Kelly Mack, Akli Amrous, Matt Wright, Steven Goodman, Leah Findlater, and Jon E. Froehlich. 2020. HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376758

[22] Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. SoundWatch: Exploring Smartwatch-based Deep Learning Approaches to Support Sound Awareness for Deaf and Hard of Hearing Users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Virtual Event Greece, 1–13. DOI:https://doi.org/10.1145/3373625.3416991

[23] W. Bradley Knox and Peter Stone. 2015. Framing reinforcement learning from human reward: Reward positivity, temporal discounting, episodicity, and performance. *Artif. Intell.* 225, (August 2015), 24–50. DOI:https://doi.org/10.1016/j.artint.2015.03.009

[24] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ACM, Atlanta Georgia USA, 126–137. DOI:https://doi.org/10.1145/2678025.2701399

[25] R. Shantha Selva Kumari, D. Sugumar, and V. Sadasivam. 2007. Audio Signal Classification Based on Optimal Wavelet and Support Vector Machine. In *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, IEEE, Sivakasi, Tamil Nadu, India, 544–548. DOI:https://doi.org/10.1109/ICCIMA.2007.370

[26] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, ACM, Berlin Germany, 213–224. DOI:https://doi.org/10.1145/3242587.3242609

[27] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng, and Anhong Guo. 2022. ImageExplorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–15. DOI:https://doi.org/10.1145/3491102.3501966

[28] Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2009. SoundSense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, ACM, Kraków Poland, 165–178. DOI:https://doi.org/10.1145/1555816.1555834

[29] Tara Matthews, Janette Fong, F. Wai-Ling Ho-Ching, and Jennifer Mankoff. 2006. Evaluating non-speech sound visualizations for the deaf. *Behav. Inf. Technol.* 25, 4 (July 2006), 333–351. DOI:https://doi.org/10.1080/01449290600636488

[30] Tara Matthews, Janette Fong, and Jennifer Mankoff. 2005. Visualizing non-speech sounds for the deaf. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, ACM, Baltimore MD USA, 52–59. DOI:https://doi.org/10.1145/1090785.1090797

[31] Matthew S. Moore and Linda Levitan. 1992. *For Hearing People Only: Answers to some of the most commonly asked questions about the deaf community, its culture, and the" deaf reality"*. Deaf Life Press.

[32] Yuri Nakao and Yusuke Sugano. 2020. Use of Machine Learning by Non-Expert DHH People: Technological Understanding and Sound Perception. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, ACM, Tallinn Estonia, 1–12. DOI:https://doi.org/10.1145/3419249.3420157

[33] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, San Jose California USA, 4884–4895. DOI:https://doi.org/10.1145/2858036.2858130

[34] Halley P. Profita, Abigale Stangl, Laura Matuszewska, Sigrunn Sky, and Shaun K. Kane. 2016. Nothing to Hide: Aesthetic Customization of Hearing Aids and Cochlear Implants in an Online Community. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Reno Nevada USA, 219–227. DOI:https://doi.org/10.1145/2982142.2982159

[35] Kristen Shinohara and Josh Tenenberg. 2009. A blind person's interactions with technology. *Commun. ACM* 52, 8 (August 2009), 58–66. DOI:https://doi.org/10.1145/1536616.1536636

[36] Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Vancouver BC Canada, 705–714. DOI:https://doi.org/10.1145/1978942.1979044

[37] Liu Sicong, Zhou Zimu, Du Junzhao, Shangguan Longfei, Jun Han, and Xin Wang. 2017. UbiEar: Bringing Location-independent Sound Awareness to the Hard-of-hearing People with Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2 (June 2017), 1–21. DOI:https://doi.org/10.1145/3090082

[38] M Tomitsch and T Grechenig. DESIGN IMPLICATIONS FOR A UBIQUITOUS AMBIENT SOUND DISPLAY FOR THE DEAF.

[39] Joe Tullio, Anind K. Dey, Jason Chalecki, and James Fogarty. 2007. How It Works: A Field Study of Non-Technical Users Interacting with an Intelligent System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), Association for Computing Machinery, New York, NY, USA, 31–40. DOI:https://doi.org/10.1145/1240624.1240630

[40] Beatrice Vincenzi, Alex S. Taylor, and Simone Stumpf. 2021. Interdependence in Action: People with Visual Impairments and their Guides Co-constituting Common Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–33. DOI:https://doi.org/10.1145/3449143

[41] Jacob O. Wobbrock, Krzysztof Z. Gajos, Shaun K. Kane, and Gregg C. Vanderheiden. 2018. Ability-based design. *Commun. ACM* 61, 6 (May 2018), 62–71. DOI:https://doi.org/10.1145/3148051

[42] Jason Wu, Chris Harrison, Jeffrey P. Bigham, and Gierad Laput. 2020. Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376875

[43] Alina Zajadacz. 2015. Evolution of models of disability as a basis for further policy changes in accessible tourism. *J. Tour. Futur.* 1, 3 (September 2015), 189–202. DOI:https://doi.org/10.1108/JTF-04-2015-0015

[44] 2011. Access Intimacy: The Missing Link. *Leaving Evidence*. Retrieved July 28, 2023 from https://leavingevidence.wordpress.com/2011/05/05/access-intimacy-the-missing-link/

[45] 2017. Access Intimacy, Interdependence and Disability Justice. *Leaving Evidence*. Retrieved July 28, 2023 from https://leavingevidence.wordpress.com/2017/04/12/access-intimacy-interdependence-and-disability-justice/

[46] 2020. Important household sounds become more accessible. *Google*. Retrieved May 3, 2023 from https://blog.google/products/android/new-sound-notifications-on-android/

[47] People + AI Guidebook. Retrieved July 27, 2023 from https://design.google/ai-guidebook

[48] Accessibility - Hearing. *Apple*. Retrieved May 3, 2023 from https://www.apple.com/accessibility/hearing/

[49] TensorFlow Hub. Retrieved April 30, 2023 from https://tfhub.dev/google/lite-model/yamnet/tflite/1

[50] Live Transcribe | Speech to Text App. *Android*. Retrieved May 3, 2023 from https://www.android.com/accessibility/live-transcribe/

[51] Audio transcription for cloud recordings. *Zoom Support*. Retrieved May 3, 2023 from https://support.zoom.us/hc/en-us/articles/115004794983-Audio-transcription-for-cloud-recordings

[52] ReSound Smart 3D hearing aid app | ReSound. Retrieved May 3, 2023 from https://www.resound.com/en-us/hearing-aids/apps/smart-3d

[53] Real-time Call Caption App | Android & Iphone. *InnoCaption*. Retrieved May 3, 2023 from https://www.innocaption.com

[54] Nest Aware. *Google Store*. Retrieved May 3, 2023 from https://store.google.com/us/product/nest_aware?hl=en-US

[55] ReCal2: Reliability for 2 Coders – Deen Freelon, Ph.D. Retrieved July 31, 2023 from http://dfreelon.org/utils/recalfront/recal2/