A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users

Danielle Bragg Nicholas Huynh Richard E. Ladner
Computer Science & Engineering
DUB Group, University of Washington
Seattle, WA 98195 USA
{dkbragg,huynick,ladner}@cs.washington.edu

ABSTRACT

Sounds provide informative signals about the world around us. In situations where non-auditory cues are inaccessible, it can be useful for deaf and hard-of-hearing people to be notified about sounds. Through a survey, we explored which sounds are of interest to deaf and hard-of-hearing people, and which means of notification are appropriate. Motivated by these findings, we designed a mobile phone app that alerts deaf and hard-of-hearing people to sounds they care about. The app uses training examples of personally relevant sounds recorded by the user to learn a model of those sounds. It then screens the incoming audio stream from the phone's microphone for those sounds. When it detects a sound, it alerts the user by vibrating and providing a pop-up notification. To evaluate the interface design independent of sound detection errors, we ran a Wizard-of-Oz user study, and found that the app design successfully facilitated deaf and hard-of-hearing users recording training examples. We also explored the viability of a basic machine learning algorithm for sound detection.

CCS Concepts

•Human-centered computing → Sound-based input / output; Accessibility systems and tools;

Keywords

Sound detection, accessibility, deaf, hard-of-hearing

1. INTRODUCTION

Knowing which sounds are happening in one's surroundings can be useful. Auditory cues can signify important events happening outside of the line of sight. For example, a person shouting or gun firing might be heard but not seen. Furthermore, society relies exclusively on sound for communicating certain information. For example, cars honk to alert other drivers; alarms ring to announce important times and emergencies; loud speakers broadcast airport announcements; microwaves beep to tell us our food is cooked; and people ring doorbells and knock on doors to announce their arrival. These societal conventions make important information inaccessible to many deaf and hard-of-hearing people.

Non-technical sound awareness methods like visual inspection can be distracting and inconvenient, and technical solutions are often specific to individual sounds. For example, alarm clocks that ring loudly, flash bright lights, and vibrate are commercially available. Many deaf people also connect their doorbell to the home lights, so that the lights flash when the doorbell is rung. However, these solutions address individual sounds, and it can be expensive and inconvenient to purchase a different device for every sound. Even with many devices, some sounds cannot be covered because each person's life, and the sounds therein, is unique.

In this paper, we present the design of a personalizable mobile phone app to detect sounds that deaf and hard-of-hearing users find important. Guided by visual feedback, users train the app to identify the sounds they want to know about by providing recorded examples of those sounds. The user categorizes recordings into groups representing different sounds. Because the app learns models of sounds from training examples, it is flexible and gives the user control. Instead of buying a separate sound detector for each important sound, the user can download and train a single app. Furthermore, because it is a mobile app, the detector is portable. It accompanies the user throughout the day, detecting sounds in any location – at work, home, or in transit.

Our mobile app design provides sound detection for deaf and hard-of-hearing people through a ubiquitous device they likely already use. Most deaf and hard-of-hearing people we surveyed want a mobile sound detector (section 3.4). Because text messaging is so useful, mobile phone adoption is high in the deaf community, where vibration notifications are widely used. Deaf users typically choose phones that support vibration [28], one of the most useful mobile features for deaf users [29], and often carry their phone on the body to feel it vibrate [11]. Our app detects sounds with a familiar device (i.e., the mobile phone) and uses a popular notification medium (i.e., text and vibration).

We informed our app design through a survey on sounds that deaf and hard-of-hearing people want to know about, methods they currently use for sound awareness, and their design criteria for a sound detector app. We evaluated our app design through a Wizard-of-Oz in-lab user study where participants set up the app to listen for various sounds, and experienced the app detecting those sounds. We also ran

ASSETS '16, October 23 - 26, 2016, Reno, NV, USA

ISBN 978-1-4503-4124-0/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2982142.2982171

an offline proof-of-concept for a GMM (Gaussian Mixture Model) based sound detection algorithm.

The key contributions of this paper include:

- A survey on the design preferences deaf and hard-of-hearing users have for a mobile sound detector app.
- The design of a mobile sound detector app independently trainable by deaf and hard-of-hearing users.
- A user study exploring the usability of our sound detector app design for deaf and hard-of-hearing users.

2. RELATED WORK

Sound awareness techniques include non-technical solutions, commercial products, and research ventures. This plethora of sound awareness methods highlights the importance of sound awareness. However, the usability of a trainable sound detector for deaf and hard-of-hearing users has not been explored, which we provide. The acoustic event detection required is an active research area spanning signal processing and machine learning.

2.1 Sound Awareness Techniques

There is a wide array of approaches that deaf and hard-of-hearing people use for sound awareness. Some deaf and hard-of-hearing people are not interested in sounds, but there are many approaches adopted by those who are. Many sounds are accompanied by visual cues that deaf and hard-of-hearing people pay attention to or check for. Some deaf and hard-of-hearing people use hearing aids or cochlear implants to improve sound sensing. It is possible to amplify only particular sound sources in hearing aids or cochlear implants, using wireless streaming with loop systems, FM, or infrared. Tactile hearing aids and tactile vocoders that vibrate to relay sonic information can also be used to identify sounds. Hearing dogs are also trained to alert their owners to important sounds.

There are many products that detect specific sounds, or replace them with other types of signals. In many deaf homes, the doorbell or phone feeds into the light system, so that the lights flash when somebody rings the doorbell or calls. Alarm clocks are available that emit loud sounds, vibrate, or flash bright lights. Several companies, including Harris Communications [12] and Sonic Alert [34], specialize in such products for deaf and hard-of-hearing consumers. Sound detectors for specific sounds are also marketed to consumers who are not deaf or hard-of-hearing. For example, baby monitors and breaking glass detectors can be useful for hearing, hard-of-hearing, and deaf consumers alike.

More comprehensive sound detection systems are emerging on the market, further signifying that sound detection is an important problem. For instance, the Leeo Smart Alert [16] plugs into an outlet and listens for smoke and carbon monoxide alarms. When it detects either alarm, it calls the subscribed phone and plays a recording of the sound so that the user can verify that the alarm is going off, and respond appropriately. Audio Analytic [3] sells a suite of sensors that are installed in the home to detect specific sounds like a baby cry, smoke alarm, or window breakage. Consumers must work with the company to develop custom sensors, and the system is designed specifically for the home. OtoSense [24] provides software for monitoring industrial machine sounds, with a recently released app for personal use. The app detects two types of smoke alarms and lets the user specify additional sounds they want it to detect. However, the usability of these designs for deaf and hard-of-hearing users has not been studied.

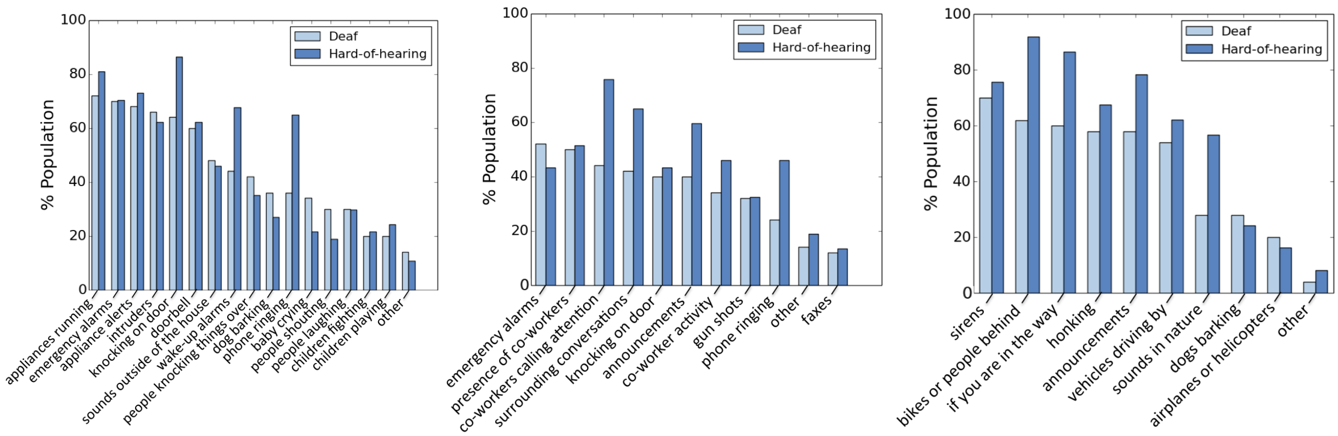
Several research projects have attempted to provide sound awareness for deaf people. Many of these systems focus on specific use cases, including detecting sounds in one specific deaf-blind person's home [8], and a chip for detecting sirens approaching from behind on the road [20]. Scribe4Me [18] is a mobile app that takes a more generalizeable approach to sound awareness. Users press a button to request detailed information about the last 30 seconds of audio, and the app uses a human-in-the-loop system for transcription. Peripheral displays depicting sounds for deaf people have also been explored (e.g. [13, 19]). Matthews et al. [19] provide a discussion of the sounds that deaf people care about. Because this research was conducted 10 years ago by interviewing a small set of people, we used it to inform our design of a large-scale web survey on deaf sounds of interest; we expect a higher quantity of responses, and those responses to be more up-to-date. Other applied sound detection research includes identifying stress in human voices (e.g., [17]), cough detection (e.g., [4, 15]), speech detection (e.g., [10, 33]), and voice recognition (e.g., [30, 21]).

2.2 Acoustic Event Detection

Acoustic Event Detection (AED) refers to the problem of identifying both when sounds occur within an audio stream and which sounds they are. Detection is more complex than sound classification because the temporal boundaries of the sounds must be determined. Creating a personalizeable sound detector app for deaf and hard-of-hearing users requires a particularly robust solution to AED. People care about a diverse set of sounds, ranging from babies crying to appliances buzzing, and the app must be able to model these different sounds. Training examples might be recorded with background noise or conflicting concurrent sounds; the phone's audio stream will include diverse environmental noise; and the microphone will be muffled when the phone is placed in a pocket. The AED algorithm must be robust to all these obstacles.

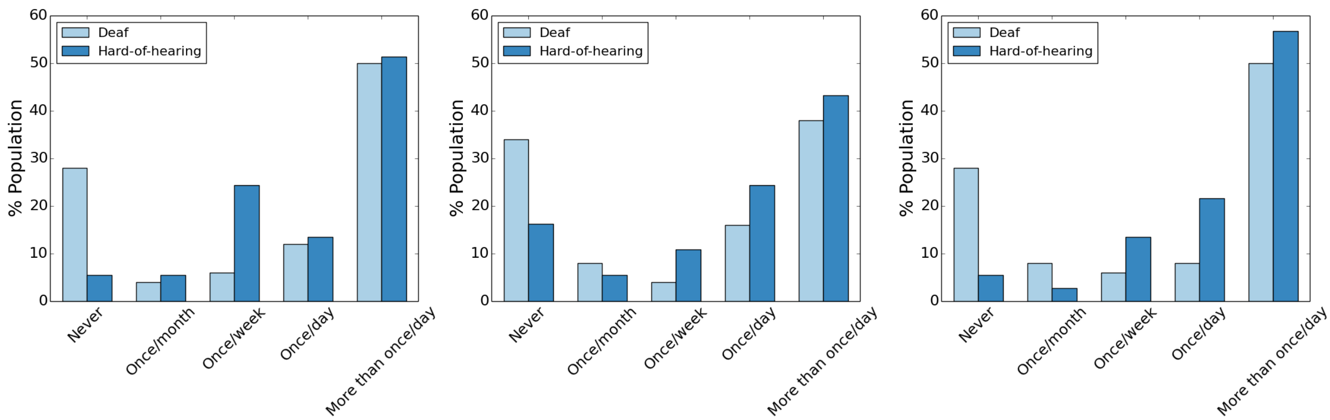
Model and feature choice greatly impacts AED accuracy. Spectral features like Mel-Frequency Cepstral Coefficients (MFCCs) represent the frequencies that make up a sound, and are commonly used to represent sounds. Because different features provide better signatures for different sounds, algorithms have been built to help determine the most effective features for particular classification tasks (e.g., [26]). The models for different sounds are built from these features. Researched models include Gaussian Mixture Models (e.g. [6, 31]), Support Vector Machines (e.g. [1, 9]), hierarchical models (e.g., [7, 1]), and random regression forests (e.g. [27]). Many of these solutions are tailored to specific sound types and environments, but a generalizable sound detection app must work well for any sound in any environment.

Once the sound model is built, different temporal methods are available for detecting sounds in the audio stream. Sliding window methods continuously determine whether a sound is present in a small window of recent sound. Hidden Markov Models represent the audio stream as a sequence of states (e.g. [6, 22, 23]). Each state represents the sound occurring at that time. Smoothing methods can help prevent the model from jumping from one sound to another.



(a) Sounds of interest *at home* (b) Sounds of interest *at work* (c) Sounds of interest *while mobile*

Figure 1: Sounds of interest to deaf and hard-of-hearing participants (a) at home, (b) at work, and (c) while mobile.



(a) How often sounds missed *at home* (b) How often sounds missed *at work* (c) How often sounds missed *while mobile*

Figure 2: Frequency of missed sounds (a) at home, (b) at work, and (c) while mobile.

Various onset detection methods have been developed and used to boost audio event detection accuracy (e.g., [5]). We explored the performance of a sliding window GMM sound detection algorithm using training data gathered in our user study. The state-of-the-art in sound detection is advancing, and we expect future work applying these algorithms to yield more accurate results.

3. SURVEY TO INFORM APP DESIGN

We conducted a survey to determine which sounds deaf and hard-of-hearing people care about, which methods they currently use for sound awareness, and what design criteria they would have for an app that detects sounds. The survey was approved by the University of Washington IRB and distributed online. It consisted of a combination of multiple-choice and free-response questions. Appendix A provides the exact questions. Participants were recruited by posting on Facebook and emailing relevant lists. We had 87 participants (51 female, 36 male). 50 were deaf, and 37 were hard-of-hearing. Ages ranged 18-99 (mean 42, std dev 17).

3.1 Sounds of Interest

The sounds that participants were interested in knowing about are presented in Figure 1. Participants selected sounds of interest from a list of options based on a previous sound awareness survey (i.e. [19]) and discussions with deaf colleagues. We also provided a write-in "other" option. They selected sounds for three scenarios: a) at home, b) at work, and c) while mobile.

The biggest differences between deaf and hard-of-hearing participants are likely explained by deaf participants having non-auditory ways of knowing when important sounds happen. For example, fewer deaf participants wanted to know about wake-up alarms and phones ringing than hard-of-hearing ones. Alarm clocks are available that ring loudly, vibrate, and have strong, flashing lights. Similarly, a phone can provide visual or tactile feedback when somebody calls, or can be connected to an external device that provides this feedback. It is likely that more deaf people know about and use these solutions. They are likely less concerned with knowing about a phone call or alarm clock going off because they already know when those events occur.

A small minority of participants wrote in additional sounds, which suggests our list was comprehensive. The home sounds participants added were: vehicles passing by, children having bad dreams, smoke and carbon monoxide detectors, appliances making unusual noises, water running, socializing, something dropping on the floor, gunshots, conversations, and distinguishing between multiple sources with similar frequency range. The sounds they wrote in for work were: dropping items, walking/running behind, moving carts, fire drill, printer, conversations, and baby sounds. The varied responses for work are likely due to varied work environments with different sounds. The sounds they added for mobile situations were: conversations, and sound location.

3.2 Missed Sound Frequency

To better understand where sound detection is needed, we asked participants how often they miss sounds of interest at home, at work, and while mobile. As shown in Figure 2, the majority of participants reported missing sounds in all three scenarios. About 50% of both deaf and hard-of-hearing participants thought they missed sounds more than once per day in each scenario. Most deaf participants either thought they never missed sounds, or that they missed sounds very frequently (more than once per day), as demonstrated by the U-shaped curve of the results in all three scenarios. Hard-of-hearing participants were more evenly distributed in how often they thought they missed sounds. It is possible that more deaf participants reported never missing sounds because they developed more reliable systems for knowing about the sounds they care about, or because they were less aware of missing sounds than hard-of-hearing participants.

3.3 Techniques for Sound Awareness

Participants reported using a wide range of techniques and devices for sound awareness, highlighting the importance of an all-purpose solution. The most widely used technique was to check to see if a sound happened (over 80% of both hard-of-hearing and deaf participants). The fewest participants relied on hearing dogs (under 30% of both hard-of-hearing and deaf participants). There was little variance in how much people relied on hearing dogs; each person either relied on a dog on a daily basis or not at all. Alarm clocks, and fire, smoke, or carbon monoxide alarms were the only alerting devices used by the majority of deaf participants. All other alert devices were used by a minority of both deaf and hard-of-hearing participants. The wide diversity of solutions with small user bases suggests that a general solution like a trainable sound detector app would be valuable. Instead of buying a separate device for many sounds they want to know about, users could download a single all-purpose app.

3.4 Need for a Mobile App

The vast majority of participants did not currently use any mobile apps for sound detection, but were interested in using a general sound detector app. Most participants did not use any mobile apps for sound detection (80% deaf, 89% hard-of-hearing). Those who did use sound detection apps reported using dictation software, software that connects to hearing aids or cochlear implants, and software that makes the phone flash or vibrate when receiving calls, alerts, or messages. None of the apps listed provide general sound detection (except OtoSense, which one participant used). Though most participants did not use mobile sound awareness apps, most wanted an app that would alert them to sounds of their choosing (88% deaf, 87% hard-of-hearing). The fact that the majority did not use any apps for sound awareness, yet wanted a sound detector app, demonstrates an unfulfilled user need.

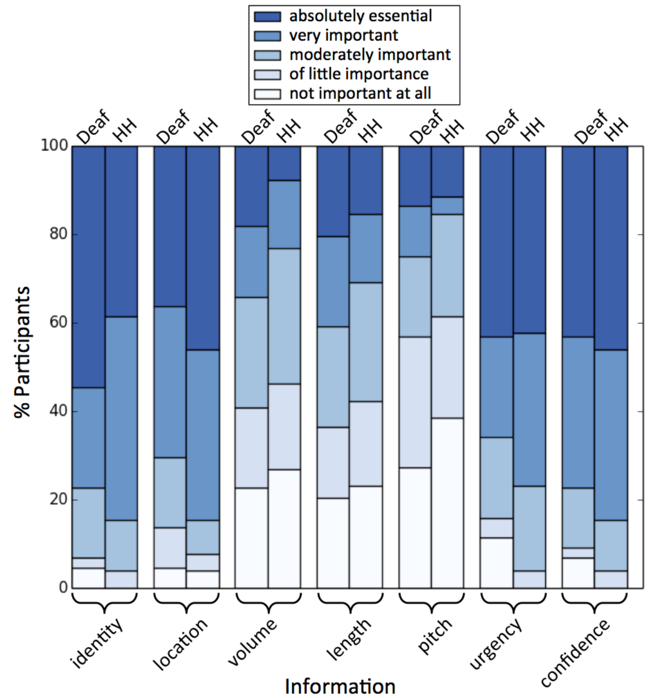
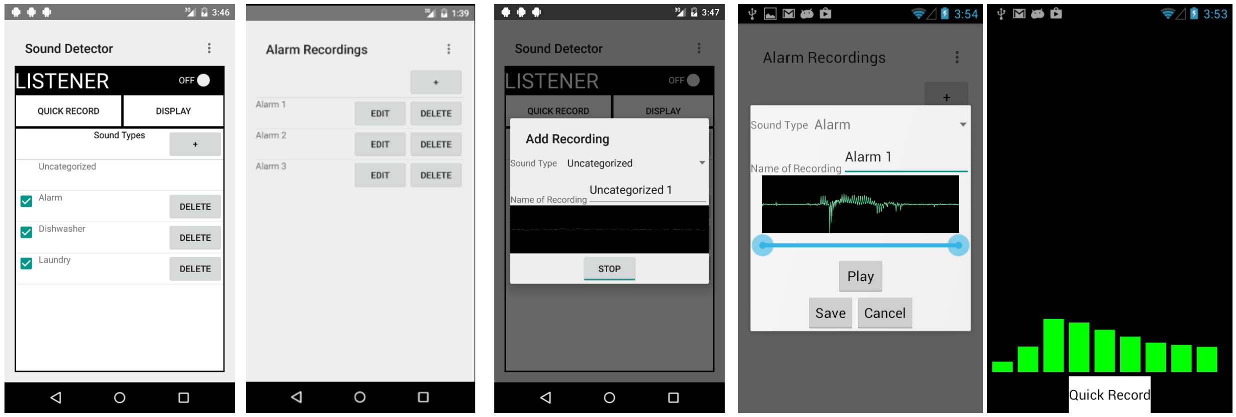


Figure 3: Desired information for app notifications. Pairs of bars represent deaf and hard-of-hearing (HH) participants.

We explored participants' design criteria for such an app. Figure 3 summarizes desired information about detected sounds. Participants most wanted to know about sound identity, location, urgency, and confidence in detection. Volume, length (duration), and pitch are less important. Participants also reported a higher tolerance for extra notifications than for missed notifications. Deaf participants were more tolerant of both missed sounds and extra notifications. Twenty-six (59%) deaf and ten (38%) hard-of-hearing participants would tolerate at least one extra notification per day. Twenty (45%) deaf and eight (30%) hard-of-hearing participants would tolerate at least one missed sound per day. It is likely that deaf participants would tolerate more errors because a faulty app would still provide an appreciable benefit, whereas hard-of-hearing participants need a more accurate app to provide a comparable benefit.

Our survey results suggest that autonomy and privacy are important to users recording examples of sounds. Hard-of-hearing participants were generally willing to record more training samples. Fourteen (54%) hard-of-hearing participants were willing to provide at least five training examples for a sound, whereas thirteen (30%) deaf participants were willing to do the same. Three (7%) deaf participants did not want to record any examples, but all hard-of-hearing participants were willing to record some. It is possible that rich visual feedback during the recording process would increase deaf users' willingness to record examples of sounds they do not hear. About half of our participants were willing to ask a hearing person to help record sounds, but many (especially deaf users) preferred autonomy. All hard-of-hearing participants were willing to ask for help, compared to 84% of deaf participants. In terms of sharing, 73% hard-of-hearing and 59% of deaf participants were "very willing" to share recordings. Reluctance to share is likely due to privacy concerns.



(a) The main screen. (b) Recording list. (c) Recording screen. (d) Editing screen. (e) Real-time display.

Figure 4: Screen shots of the app interface.

4. SOUND DETECTOR DESIGN

Informed by our survey and refined through iterative design, we designed a sound detector app to alert deaf and hard-of-hearing users to nearby sounds of interest. We implemented the design as an Android application using standard Android sound processing and data storage methods.

4.1 Interface Design

The app interface (Figure 4) allows users to train the app to identify sounds by recording examples of those sounds. The recording and editing screens provide visual feedback to support deaf and hard-of-hearing users independently recording sounds. Users create their own sound types (ex: "door knock" or "microwave beep"), and categorize their examples under the appropriate types. The app uses machine learning to model these sounds, and runs a sound detection algorithm on the incoming audio stream from the phone microphone to detect when these sound types occur. The user is then notified with a vibration and text notification letting them know which sound has occurred. The application also provides a display screen with a waveform of the current audio. The display can help users detect sounds that they want to record, gain an awareness of background noise, or find sound sources by seeing the waveform strengthen as they move closer to the source.

4.1.1 Main Screen

Because training the app to recognize different sound types is central to the app's functionality, the app's main menu provides a list of all sound types, as shown in Figure 4a. Users can check (or uncheck) the green box at the left of a sound type to make the app listen for (or ignore) that sound type. Sound types can be deleted by clicking DELETE or added by clicking the + at the top right. The "uncategorized" type appears at the top of the sound type list, and cannot be deleted. Clicking on a sound type brings the user to the recording list for that sound type. Through our iterative design process, we enlarged the listener switch to highlight the importance of turning it on and moved the sound display to a separate screen to avoid distraction.

4.1.2 Recording List

The recording list, displayed in Figure 4b, lists every recording that was given as an example of a particular sound type. To keep the interface clean, the list only displays the user-given recording names. Additional details are available in the editing screen when the user clicks EDIT. Clicking DELETE deletes the corresponding recording. The + button allows the user to add a new recording. The recording interface that then appears automatically classifies the new recording under the current sound type.

4.1.3 Recording Interface

The recording interface, pictured in Figure 4c, provides a waveform visualization for visual feedback during the recording process. The "start" and "stop" buttons allow the user to start and stop recording. The recording interface can be accessed in two ways: 1) from the quick record button on the main menu or display screen, or 2) by navigating to a particular sound type and clicking the + button to add a sample. If the recording interface is accessed through a quick record button, the recording is put into the "uncategorized" type. Otherwise, it is saved under the currently selected type. To highlight the importance of categorizing the sample appropriately, the app verifies with the user before saving an "uncategorized" recording.

4.1.4 Editing Interface

The editing screen, pictured in Figure 4d, allows users to move a recording to a different sound type, change its name, or trim the recording. A static waveform visualization provides a visual representation of the recording. The visual feedback can help users evaluate the content of recordings. For example, a pulsing alarm will be visualized as a series of peaks. If the recording does not look as it should, the user is free to trim it or delete and try again. Recordings are trimmed by dragging two sliders along the waveform visualization to frame the desired portion of the recording. The user can also play back the recording if they want to listen to it themselves or ask a hearing friend to check the quality of the recording.

4.1.5 Real-time Display

The display screen, in Figure 4e, provides a visual representation of the current sound level. This feature can be used on its own to provide a sense of the current noise level, or can be used to help users identify when sounds they want to record happen. For example, if they want to record their dog barking, the visual display will jump every time their dog barks. When a sound of interest occurs, it can be recorded directly from this screen using the quick record button.

4.2 Implementation

We implemented the sound detector interface as an Android application. Audio data was managed by native Android classes, AudioRecord and AudioTrack. The two classes are designed to be used in conjunction with one another and function in similar ways. When the user records a sound, AudioRecord stores the microphone input as raw Pulse Code Modulation (PCM) data in a buffer (<1 s), which is transferred to external storage. The AudioTrack class inversely reads data from the external file into the buffer and plays it from there. Metadata on the recordings is stored in the phone using Android's built-in SQLite database. The waveform displays are generated from the raw PCM data. Our implementation includes all functionality, except for algorithmic sound detection. For our user study, we pushed event detection notifications ourselves using Parse (i.e., [25]), an open source backend API.

5. INTERFACE USABILITY STUDY

We ran a formative in-lab study to evaluate the usability of the sound detector interface design for deaf and hard-of-hearing users. During the study, participants set up the app to listen for two sounds: door knocks and an alarm clock ringing. For each sound, participants recorded and edited examples of the sound, saved the sounds in the appropriate category, and set the app to listen for the desired sound. Participants answered questions and provided open-ended feedback about their experience. We obtained IRB approval through the University of Washington and recruited by posting on Facebook and emailing relevant lists.

We had 12 participants (9 female, 3 male). Age ranged from 19-60 (average 33). Five identified as deaf (all from birth); four as hard-of-hearing (1 from birth, 3 from childhood); one as both deaf and hard-of-hearing depending on the context (from infancy); one as "hearing impaired" (from childhood); and one as mostly hearing with difficulty in noisy environments (from young or mid-adulthood). Seven participants (58%) reported having their mobile phone with them over 80% of the time at home, eight (75%) reported the same at work, and ten (83%) reported the same when mobile. This smartphone-equipped majority would be able to detect sounds throughout the day using our app. One participant did not own a smartphone and expressed frustration with smartphones in general. No participants used apps to monitor sounds outside of the study.

5.1 Study Procedures

The study took place in a lab setting, and an American Sign Language interpreter was made available to each participant. The study consisted of several steps: 1) watching a short demo video, 2) setting the app up to detect door knocks, 3) setting the app up to detect an alarm, and 4) receiving a sound detection notification. Participants used a Samsung Android phone with the app installed. After each task, participants answered specific questions about their experience, and provided freeform feedback. Participants were encouraged to ask questions and talk about their experience.

We explained that the app will detect each sound more accurately if the user provides more recorded examples and trims them, and participants were free to decide how many samples to record and which to trim. We provided the following instructions for setting up the app to listen for the door knock (and later for the alarm):

1. Create a sound category for door knocks (or the alarm).
2. Record examples of door knocks (or the alarm).
3. Tell the app to detect knocks (or the alarm).

In closing, we asked participants to configure the app to listen for door knocks but not the alarm, and triggered a door knock notification. Because we wanted to evaluate the usability of the app interface without detection accuracy confounding our results, we ran a Wizard-of-Oz experiment. Whenever a sound occurred that the app was configured to detect, we manually pushed a notification to the phone.

5.2 Study Results

We found the app design to be usable for deaf and hard-of-hearing users recording training examples of sounds. We evaluated usability through participants' ability to train and use the app, and their qualitative feedback. All participants successfully trained the app by recording, editing, and organizing samples appropriately, and noticed notifications. Participants' responses provide evidence that the training process was generally easy and notifications were appropriate. Areas for improvement include clearer instructions about sound categorization, larger buttons and checkboxes, and personalized notifications. We present participants' answers to questions about their experience in Figure 5, and provide a thematic analysis of their free-form feedback.

5.2.1 *Recording and Organizing Recordings*

All participants successfully recorded sound samples, and 91.7% agreed that "It was easy to record sounds." These participants understood that the user must categorize their recordings so that the app can "learn" those sounds, and felt that the app clearly supported this task. In the words of P9, it was "sleek and minimal which makes it easy to use." Others called it "intuitive," "simple," and "easy to use." Six participants specified the app's customizability as a strength. They liked that it could be trained to detect their personal sounds, and that it could handle a wide variety of sounds. They also enjoyed having control over their recordings.

The organization of sounds into categories confused some participants. Several asked us to clarify what a "sound category" was. Some expected the app to distinguish between individual recordings within a single category. Others thought a single example of a sound would be sufficient for the app to identify that sound. One person expected to provide more recordings for a variable sound (like knocking on different doors) than for a highly regular sound (like an electronic alarm). This expectation aligns with our vision of the app using the recordings in a single category to learn a model of that sound type. More training examples lead to more robust models, especially for highly variable sounds, and thus improved accuracy. Explanations satisfied our participants, and clearer instructions would likely reduce future confusion.

5.2.2 *Editing Recordings*

Participants generally enjoyed editing their training examples. The waveform visualization of the recordings was

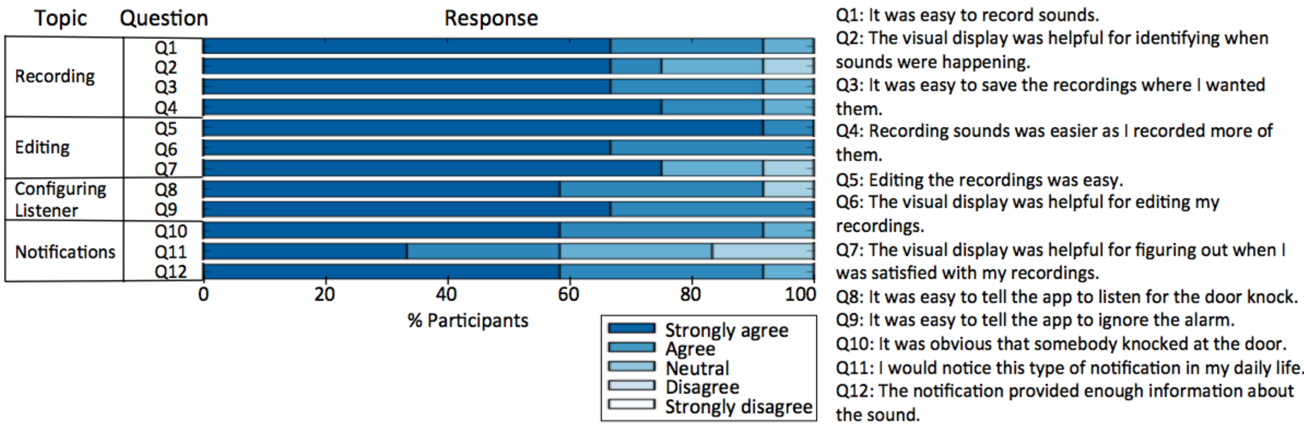


Figure 5: Participant responses to questions about their experience using the interface during the study.

a particular strength, with 91.7% of participants *strongly agreeing* that "The visual display was helpful for editing [their] recordings." The display provides a visual representation on recording content that allows deaf and hard-of-hearing users to evaluate them without hearing them. P4 explained how the display helps understand and trim recordings of repetitive sounds like an alarm: "if a deaf person could not hear it but wants to record whatever the sound is, they could possibly see the repetition and edit it down to a certain amount." Only half of our participants played back a sample, yet *all* participants successfully used the sliders to frame the part of their recordings they wished to keep. As P9 summarized, "VERY easy to use – just drag and stop." Once familiarized with the editing process, several participants described viewing and trimming recordings as "fun." Two participants expressed difficulty controlling the sliders. Enlarging the slider area would improve ease in the future.

5.2.3 *Configuring the Listener*

All participants successfully configured the app to listen for door knocks and ignore the alarm, though several asked questions along the way. Configuration involved selecting the door knock category, unselecting the alarm category, and sliding the listener on. All these actions take place on the main menu. All participants agreed that "It was easy to tell the app to ignore the alarm," and all but 8.3% agreed that "It was easy to tell the app to listen for the door knock." Of the participants who found configuration easy, one described the tasks as "easy peasy" and another elaborated, "it's easy to check on and off the options." Of the participants who had difficulty, four mentioned that the check boxes for selecting and unselecting sound categories were too small, and consequently had trouble checking the boxes with their fingers. One participant forgot to turn on the listener once they selected the sound categories, and suggested increasing the listener button's size to draw attention to the button.

5.2.4 *Notification System*

While most participants (58%) found the notification design sufficient to alert them to sounds in daily life, we received more feedback on the notification design than for any other part of the design. All participants received a door knock notification. If they did not test the door knock detection themselves, we knocked on the study door for them.

The combination of haptic and visual feedback caught each participant's attention. Many were impressed when the app notified them, and responded with "Cool!" or "Neat!" Those who criticized the text display wanted a larger notification, more information like a timestamp, and longer persistence on the screen. One participant was concerned about missing notifications if the phone was not physically on them, explaining, "it'd be hard to detect... if it only vibrates and the phone is not in my hand." Several participants wanted to customize the notifications for different sounds. For example, the phone could use a different sequence of vibrations to notify the user about each sound type. Others wanted alerts sent over another modality or to another device. They suggested email, SMS, flashing phone lights, sonic alerts, and amplifying the detected sound.

5.2.5 *Use Cases*

Participants envisioned many uses of our app in their lives. Picking out sounds while watching TV or in a noisy environment can be particularly difficult, and several participants hoped that the app would do so for them. Others noted that the app would be useful when they chose not to use their cochlear implants or hearing aids, or when they were using their cochlear implants to listen to music or other content besides their surroundings. Several participants stay near the door and periodically check for guests they are expecting, and using our app to detect door knocks would give them more freedom. They were also interested in using the app to detect distant sounds. For example, they might want to know when the tea pot boils in the kitchen downstairs. Several participants mentioned that they would like to take the app home and try it out on their personal use cases.

Participants also suggested design improvements for real-world usability. Some sounds are common but difficult to record (e.g. a fire-alarm or ambulance passing), and having these sounds built-in would be convenient. One participant was concerned about the presence of background noise or competing sounds while recording. While the app provides some feedback about background noise by visualizing sound level (i.e. volume), adding explicit feedback about conflicting sounds might improve usability. Three participants were concerned about the time it takes to set up the app, and streamlining the training process or allowing users to share training samples would likely lower barriers to use.

6. SOUND DETECTION EXPLORATION

To explore the viability of a sound detection app trained by deaf and hard-of-hearing users, we implemented a basic sound detection algorithm and tested it on the training examples recorded in our user study. We model sounds as Gaussian Mixture Models (GMMs), a common technique in sound recognition and detection (e.x. [31]). The model's features are Mel-Frequency Cepstral Coefficients (MFCCs) which are commonly used for speech recognition [30]. We extract 14 MCFFs for each frame in the training samples.

We use a set of sliding windows to detect sounds in the incoming audio stream. Each sound (or class) has its own sliding window, spanning the average length of the class's training examples. The current window is classified as the sound whose GMM produces the highest normalized log-likelihood, the sliding windows are incremented, and the process repeats. The windows are incremented by 1/3 the size of the smallest window, a gap size found to perform well through experimentation. We smooth our classification by extending it to cover the expected duration of the sound, computed as the average duration of the training examples for that sound. We add a "white noise" class, trained on examples of office sounds and silence, to compete with the other sound types. An app notification is triggered when a window is classified as a sound other than "white noise" and the previous window does not share the same classification.

Table 1 shows our sound detection algorithm performance on recordings of door knocks and alarms from our user study. We ran 3-fold cross-validation on their examples. Test clips were formed by randomly inserting user recordings into longer streams of white noise collected in the study setting. We compared the events detected by our algorithm to the ground ground truth of audio events occurring where they were inserted in the longer streams. Each time the algorithm detected the inserted sound in the insertion range, we counted a true positive; each time the algorithm detected any other sound, we counted a false positive. Precision is the fraction of notifications sent that detected the right sound, and recall is the fraction of sounds that triggered a notification. F-score is a weighted average of precision and recall.

		Alarm	Knock
Uncleaned	Precision	1.00	0.41
	Recall	0.28	1.00
	F-Score	0.44	0.58
Cleaned	Precision	0.71	0.77
	Recall	0.98	0.41
	F-Score	0.82	0.54

Table 1: Accuracy of our sound detection algorithm using 3-fold cross-validation on recordings from our user study.

Our results demonstrate that background noise in training examples can impact detection accuracy. One researcher took notes with a portable keyboard during the study, and loud typing sounds were present in many training examples. Because loud typing sonically resembles knocking, keyboard sounds during alarm recordings were often mistaken for door knocks. We removed typing noises with Audacity's noise removal tool [2] to produce "Cleaned" training examples. This boosted performance, and in particular increased alarm recall. Because deaf and hard-of-hearing users might not be aware of or recognize the impact of background noise, providing additional visual feedback on training example quality would be a powerful addition to the app design. While our algorithmic experiments are preliminary, the state-of-the-art in sound detection is advancing, and we expect that robust sound detection will be possible in the future.

7. CONCLUSION

In this work, we introduced the design of an app that detects sounds of interest to deaf and hard-of-hearing users. It is trainable by users who record examples of the sounds they want detected. Visual representations of both real-time audio and recorded sounds provide visual feedback on sonic content to allow deaf and hard-of-hearing users to independently train the app. Our design is informed by a widescale survey we ran on the design criteria that deaf and hard-of-hearing users have for such an app, including which sounds they want it to detect. We evaluated our interface design through a Wizard-of-Oz user study, and found the interface to be highly usable for deaf and hard-of-hearing users recording training examples of sounds. We also provided a preliminary exploration of a sound detection algorithm with training data from the user study.

A trainable sound detector app has many potential benefits: improved awareness in social situations where information is only communicated auditorily, freedom from visually checking if important events have occurred, the ability to turn off cochlear implants or hearing aids while still knowing about important sounds, and the consolidation of multiple detection methods into a single app. Allowing users to train the app gives users a great amount of control over the detector. They can customize it to personal sounds, and can expect high detection accuracy because it is trained on the exact sounds they want it to identify, recorded with the same device that will do the detection, and likely in the same environments where the detector will be expected to work.

There are several limitations to our current work. In particular, we did not implement a sound detection algorithm in the app because we did not achieve a high enough accuracy for diverse sounds in noisy environments with the methods we tried. Using Wizard-of-Oz sound detection for our user study allowed us to evaluate the app design without accuracy errors impacting the user experience. The algorithm we implemented is preliminary, and we expect to find appropriate algorithms in the future. Sensors are improving as industry pushes to gather data about users' environments, and virtual assistants already detect voice commands. The existence of virtual assistants like Apple's Siri (i.e. [32]) that run constantly but do not drain the battery suggests adequate efficiency is possible as well.

We plan to improve our interface design and release a complete working app. Because our study participants requested various notifications, we plan to support customization. For example, one participant could receive email alerts while another relays detected sounds to their cochlear implant. Sounds could even trigger complex responses, like turning on the porch lights when somebody knocks, by integrating into a smart environment or logic system like IFTTT (i.e. [14]). We will also explore active learning to help guide users about which categories of sounds need more examples. We plan to run a longitudinal study with the complete app to fully explore its usability. We hope that our work will result in a useful product, and encourage other sound detection researchers and developers to consider and evaluate the usability of their systems for deaf and hard-of-hearing users.

8. REFERENCES

[1] P. K. Atrey, N. C. Maddage, and M. S. Kankanhalli. Audio based event detection for multimedia surveillance. In *Proc. ICASSP*, volume 5, pages 813–816, 2006.

[2] Audacity. http://www.audacityteam.org, Accessed: 2015-09-28.

[3] Audio Analytic. http://www.audioanalytic.com, Accessed: 2016-05-06.

[4] S. Birring, T. Fleming, S. Matos, A. Raj, D. Evans, and I. Pavord. The leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough. *European Respiratory Journal*, 31(5):1013–1018, 2008.

[5] S. Böck, F. Krebs, and M. Schedl. Evaluating the online capabilities of onset detection methods. In *Proc. ISMIR*, pages 49–54, 2012.

[6] M. Casey. General sound classification and similarity in mpeg-7. *Organised Sound*, 6(02):153–164, 2001.

[7] C. Clavel, T. Ehrette, and G. Richard. Events detection for an audio-based surveillance system. In *Proc. ICME*, pages 1306–1309. IEEE, 2005.

[8] R. I. Damper and M. D. Evans. A multifunction domestic alert system for the deaf-blind. *IEEE Transactions on Rehabilitation Engineering*, 3(4):354–359, 1995.

[9] G. Guo and S. Z. Li. Content-based audio classification and retrieval by support vector machines. *IEEE Transactions on Neural Networks*, 14(1):209–215, 2003.

[10] J. Haigh and J. Mason. Robust voice activity detection using cepstral features. In *Proc. TENCON*, volume 3, pages 321–324. IEEE, 1993.

[11] J. Harkins, P. E. Tucker, N. Williams, and J. Sauro. Vibration signaling in mobile devices for emergency alerting: A study with deaf evaluators. *Journal of deaf studies and deaf education*, 15(4):438–445, 2010.

[12] Harris Communications. http://www.harriscomm.com, Accessed: 2016-05-06.

[13] F. Ho-Ching, J. Mankoff, and J. A. Landay. Can you see what i hear?: the design and evaluation of a peripheral sound display for the deaf. In *Proc. CHI*, pages 161–168. ACM, 2003.

[14] If This Then That (IFTTT). https://www.ifttt.com, Accessed: 2016-05-06.

[15] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld, and S. N. Patel. Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proc. UbiComp*, pages 375–384. ACM, 2011.

[16] Leeo Smart Alert. http://shop.leeo.com/pages/about-leeo-smart-alert, Accessed: 2016-05-06.

[17] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proc. UbiComp*, pages 351–360. ACM, 2012.

[18] T. Matthews, S. Carter, C. Pai, J. Fong, and J. Mankoff. Scribe4me: evaluating a mobile sound transcription tool for the deaf. In *UbiComp*, pages 159–176. 2006.

[19] T. Matthews, J. Fong, F. W.-L. Ho-Ching, and J. Mankoff. Evaluating non-speech sound visualizations for the deaf. *Behaviour and Information Technology*, 25(4):333–351, 2006.

[20] M. Mielke, A. Schäfer, and R. Brück. A mixed signal ASIC for detection of acoustic emergency signals in road traffic. *International Journal of Microelectronics and Computer Science*, 2:105–111, 2010.

[21] L. Muda, M. Begam, and I. Elamvazuthi. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*, 2010.

[22] P. Nordqvist and A. Leijon. An efficient robust sound classification algorithm for hearing aids. *The Journal of the Acoustical Society of America*, 115(6):3033–3041, 2004.

[23] S. Oberle and A. Kaelin. Recognition of acoustical alarm signals for the profoundly deaf using hidden markov models. In *Proc. ISCAS*, volume 3, pages 2285–2288. IEEE, 1995.

[24] OtoSense. http://www.otosense.com, Accessed: 2016-04-10.

[25] Parse. http://www.parse.com, Accessed: 2016-05-06.

[26] G. Peeters and X. Rodet. Automatically selecting signal descriptors for sound classification. In *ICMC*, pages 1–1, 2002.

[27] H. Phan, M. Maas, R. Mazur, and A. Mertins. Random regression forests for acoustic event detection and classification. *IEEE/ACM TASLP*, 23(1):20–31, 2015.

[28] M. R. Power and D. Power. Everyone here speaks txt: Deaf people using sms in australia and the rest of the world. *Journal of deaf studies and deaf education*, 9(3):333–343, 2004.

[29] M. R. Power, D. Power, and L. Horstmanshof. Deaf people communicating via SMS, TTY, relay service, fax, and computers in australia. *Journal of deaf studies and deaf education*, 12(1):80–92, 2007.

[30] L. Rabiner and B.-H. Juang. Fundamentals of speech recognition. 1993.

[31] D. A. Reynolds. Speaker identification and verification using gaussian mixture speaker models. *Speech communication*, 17(1):91–108, 1995.

[32] Siri. http://www.apple.com/ios/siri, Accessed: 2016-05-06.

[33] J. Sohn, N. S. Kim, and W. Sung. A statistical model-based voice activity detection. *Signal Processing Letters, IEEE*, 6(1):1–3, 1999.

[34] Sonic Alert. http://www.sonicalert.com, Accessed: 2016-05-06.

APPENDIX

A. SURVEY QUESTIONS

1. How often do you miss sounds that you want to know about? For example, a door knock, baby crying, or car honking.

	Never	Once per month	Once per week	Once per day	More than once per day
At home:	○	○	○	○	○
At work:	○	○	○	○	○
When mobile:	○	○	○	○	○

2. At home, what sounds do you care about? Check all that apply.

- ☐ Emergency alarms
- ☐ Wake-up alarms
- ☐ Doorbell
- ☐ Knocking on door
- ☐ Phone ringing
- ☐ People shouting
- ☐ People laughing
- ☐ Children fighting
- ☐ Children playing
- ☐ Baby crying
- ☐ People knocking things over (ex: pots banging, vase breaking, plates breaking)
- ☐ Intruders
- ☐ Dog barking
- ☐ Appliance alerts (ex: dryer beeping, microwave beeping, tea pot boiling)
- ☐ Appliances running by accident (ex: garbage disposal on, faucet on)
- ☐ Sounds outside of the house (ex: people shouting outside the window)
- ☐ Emergency alarms

Other: ____________

3. At work, what sounds do you care about? Check all that apply.

- ☐ Presence of co-workers
- ☐ What your co-workers are doing
- ☐ Co-workers trying to get your attention
- ☐ Surrounding conversations
- ☐ Knocking on door
- ☐ Emergency alarms
- ☐ Phone ringing
- ☐ Faxes
- ☐ Announcements
- ☐ Gun shots

Other: ____________

4. When mobile, what sounds do you care about? Check all that apply.

- ☐ Vehicles driving by
- ☐ Honking
- ☐ Sirens
- ☐ Airplanes or helicopters
- ☐ Bikes or people coming up behind you
- ☐ Whether you are blocking another person (ex: ”excuse me”, or ”watch out”)
- ☐ Dogs barking
- ☐ Sounds in nature (ex: birds chirping, water in a stream, thunder)
- ☐ Announcements (ex: airport or train station announcements)

Other: ____________

5. How often do you use the following for sound awareness?

	Never	Once per month	Once per week	Once per day	More than once per day
Vibration sensing (ex: through the floor)	○	○	○	○	○
Checking to see if the sound happened (ex: checking to see if somebody knocked on the door)	○	○	○	○	○
Sound alerting devices (ex: flashing lights for the doorbell)	○	○	○	○	○
Hearing dog	○	○	○	○	○
Assistive hearing devices	○	○	○	○	○

6. What visual or tactile alerting devices do you use, if any?
   - ☐ Baby monitor
   - ☐ Sound monitor (for any type of sound)
   - ☐ Doorbell or door knock signaler
   - ☐ Fire, smoke, or carbon monoxide alarms
   - ☐ Security alarms
   - ☐ Motion detector
   - ☐ Telephone signaler
   - ☐ Alarm clock
   - ☐ Weather alert
   - ☐ Hearing dog

   Other: ☐

7. Do you use any apps on your mobile phone to provide sound awareness?
   ○ Yes
   ○ No
8. What mobile apps do you use for sound awareness? ☐
9. Suppose there is a new mobile phone app that can detect sounds. You tell it which wounds to listen for, and it sends you alerts when it hears those sounds. For example, you can tell it to listen for knocking on the door. Then every time it hears a knock on the door, it vibrates and a message appears on your phone screen. Would you be interested in using this app?
   ○ Yes
   ○ No
10. (If no:) Why would you not be interested in using such an app to detect sounds? ☐
11. (If yes:) What sound would you most want the app to detect? ☐
12. When the app detects your sound, how important is it that the app tells you the following information?

	Not important at all	Of little importance	Moderately important	Very important	Absolutely essential
What it is	○	○	○	○	○
Where it comes from	○	○	○	○	○
How loud it is	○	○	○	○	○
How long it lasts	○	○	○	○	○
How high or low the pitch is	○	○	○	○	○
How urgent it is	○	○	○	○	○
How sure the app is	○	○	○	○	○

13. Suppose the app tells you that your sound happened when it did not. How often can this happen, so that you would still use the app?

2-3 times per day	Once per day	Once per week	Once per month	Never
○	○	○	○	○

14. Suppose the app missed your sound and did not send you an alert. How often can this happen, so that you would still use the app?

2-3 times per day	Once per day	Once per week	Once per month	Never
○	○	○	○	○

15. In order to recognize a particular sound, the app needs samples of that sound. For example, before it can recognize your microwave beeping, the app needs recordings of the microwave beeping. How many samples would you be willing to record of your most important sound?

0	1-2	3-4	5-10	more than 10
○	○	○	○	○

16. How willing would you be to ask a hearing person for help to record sounds?

Not willing	Reluctant, but willing	Very willing
○	○	○

17. How willing would you be to share your recordings with other people using the app, so that they do not need to record the same sounds?

Not willing	Reluctant, but willing	Very willing
○	○	○