# Weaving Sound Information to Support Deaf and Hard of Hearing People's Real-time Sensemaking of Auditory Environments: Co-designing with a DHH User

JEREMY ZHENGQI HUANG

University of Michigan, Computer Science and Engineering, zjhuang@umich.edu

JAYLIN HERSKOVITZ

University of Michigan, Computer Science and Engineering, jayhersk@umich.edu

LIANG-YUAN WU

University of Michigan, Computer Science and Engineering, lyuanwu@umich.edu

CECILY MORRISON

Microsoft Research, cecilym@microsoft.com

DHRUV JAIN

University of Michigan, Computer Science and Engineering, profdj@umich.edu

Current AI sound awareness systems can provide deaf and hard of hearing people with information about sounds, including discrete sound sources and transcriptions. However, synthesizing AI outputs based on DHH people's ever-changing intents in complex auditory environments remains a challenge. In this paper, we describe the co-design process of SoundWeaver, a sound awareness system prototype that dynamically weaves AI outputs from different AI models based on users' intents and presents synthesized information through a heads-up display. Adopting a Research through Design perspective, we created SoundWeaver with one DHH co-designer, adapting it to his personal contexts and goals (*e.g.*, cooking at home and chatting in a game store). Through this process, we present design implications for the future of "intent-driven" AI systems for sound accessibility.

CCS CONCEPTS • Human-centered computing • Accessibility • Empirical studies in HCI

**Additional Keywords and Phrases:** Accessibility, AI, sound awareness, deaf and hard of hearing

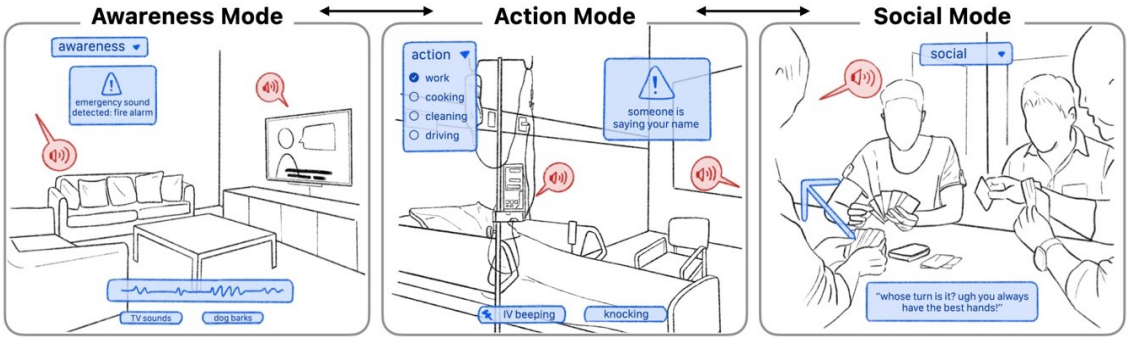


**Figure 1: The SoundWeaver's Prototype's Three Weaving Modes. SoundWeaver is an AI sound awareness system that dynamically weaves AI outputs based on DHH users' intents across personal contexts. The prototype contains three modes: Awareness, Action, and Social Mode. Awareness Mode provides general awareness of the environmental sounds through visualization of ambient sounds and optional sound identification. Action Mode facilitates active monitoring of**

**specific sounds related to a task. Social Mode uses captions and peripheral visualization of ambient sounds to facilitate social interactions. Users can freely switch between modes on the fly based on their real-time information needs.**

## 1 INTRODUCTION

Deaf and hard of hearing (DHH) individuals have limited access to sound information and often seek to enhance their understanding of their environments through sound awareness [8, 15]. To address this, HCI researchers have leveraged various AI models to design and develop intelligent sound awareness applications that reduce barriers to accessing sound information [8, 14, 31, 32]. For example, SoundWatch [32], powered by audio classification models, notifies DHH users of individual sound events. Speech recognition, developed initially to support captioning for DHH people, has since become an integral part of audio accessibility for mainstream devices and software [69, 70]. These systems draw on rapidly advancing models for machine understanding of speech and non-speech sounds, including those for audio classification [71, 72], acoustic scene understanding [35], and automatic speech recognition [17, 73, 74], which will only continue to grow and develop more complex capabilities.

However, current AI-based sound awareness applications usually have pre-configured outputs [32]. These outputs do not have semantic connections to DHH people's real-time contexts, goals, and information needs. For example, consider the following scenario of a DHH person ("Sam") working as a Barista in a coffee shop (Figure 2):

When a customer approaches to place an order, Sam activates live transcription to understand their speech. As she begins fulfilling the order (*e.g.*, preparing a latte), her focus shifts from speech comprehension to order completion. While making coffee, Sam needs to monitor when the machine starts and finishes brewing while simultaneously watching for new customers. For this purpose, she enables sound recognition. However, since the interface merely consolidates all sound information (Figure 2; Prior Approaches), Sam struggles to interpret the recognition results effectively. She must also keep live transcription running to detect when someone calls her name or when colleagues initiate conversation, forcing her to divide her attention between the transcription and coffee preparation.

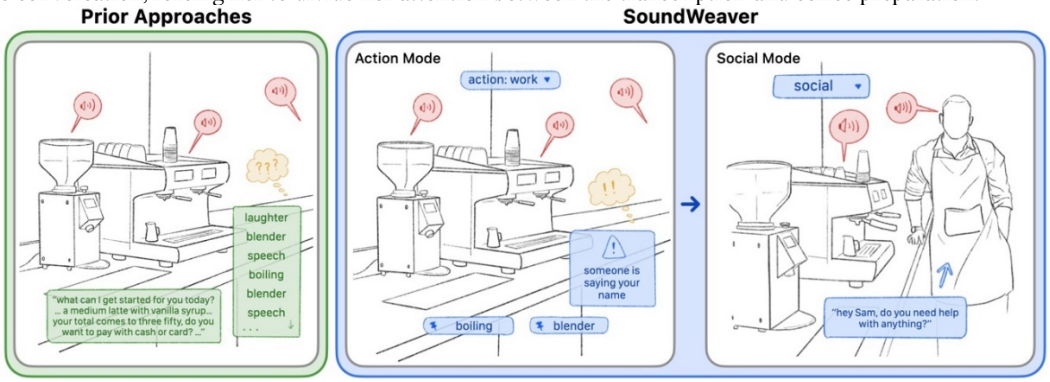


**Figure 2: An illustrative comparison of our system with prior approaches. Prior sound awareness systems assume pre-configured outputs (*e.g.*, showing sound events and captions regardless of context). In contrast, SoundWeaver adapts to users' different intents by synthesizing contextually appropriate information. For example, when the user ("Sam") is making coffee, Action Mode will help her monitor relevant sounds like "boiling" and "blender." When someone calls Sam, SoundWeaver will show this information on the display, leading Sam to switch to Social Mode, where the system focuses on displaying captions while maintaining some degree of awareness of the environmental sounds.**

The scenario above illustrates how current sound awareness systems, with their static design, often fail to meet DHH individuals' evolving information needs. As AI capabilities continue to develop, so does the potential for sound awareness systems to move beyond presenting discrete AI outputs toward delivering synthesized information that facilitates DHH people's real-time sensemaking and semantic understanding of auditory environments. To this end, this

work explores two critical aspects of human-AI interaction design in the context of sound awareness systems for DHH individuals. First, we explore an "intent-driven" AI system that adapts its behavior based on DHH users' intents, purposefully weaving together AI outputs to meet DHH people's dynamic information needs and support their sensemaking of complex auditory environments. Second, we seek to understand the relationships among AI systems, DHH users, and their environments, observing how these connections evolve through interactions.

We present our work from a Research through Design (RtD) perspective [66]. We started by reviewing previously identified design challenges and opportunities in the human-AI interaction and sound awareness technology space [1, 24, 75]. We also drew inspiration from established theories and models guiding the designs of accessible environments for DHH individuals, including DeafSpace [48], best captioning practices [37], and broader norms of Deaf Culture. Grounded in prior knowledge, we then worked closely with Declan[1], a Deaf participant, over six months to iteratively prototype an AI system that supports Declan's real-time sensemaking of auditory environments. Specifically, we learned about Declan's personal contexts (*e.g.*, daily routines, physical environments) alongside his specific information needs and preferences. Through a multi-stage grounded theory analysis, our findings elicited several design goals, including developing intent-responsive system behaviors and complementing Declan's trusted sensemaking approaches.

Informed by these design goals, we created *SoundWeaver*, an intent-driven AI system that weaves AI outputs about sound based on DHH users' needs and intents, synthesizing relevant sound information through a heads-up display. SoundWeaver was iteratively developed over multiple co-design sessions with Declan, supplemented by discussions with our team of mixed hearing abilities. The final SoundWeaver prototype contains three *weaving modes*: Awareness, Action, and Social (Figure 1). These modes facilitate distinct user intents:

1. **Awareness Mode.** Awareness Mode helps DHH users learn about the overall auditory environment and be aware of intermittent sound events.
2. **Action Mode.** Action Mode helps DHH users perform more action-intensive tasks by providing more active and focused monitoring of a selected set of sounds.
3. **Social Mode.** Social Mode supports social interactions while helping DHH users maintain certain levels of awareness of the auditory environments.

SoundWeaver lets users switch among the weaving modes instantly as their needs change with a simple button click. The system also supports granular customization (*e.g.,* selecting specific sounds for Action Mode), allowing users to adjust how the system behaves and delivers contextually relevant information. In Sam's scenario with SoundWeaver, Action Mode helps her focus only on essential work-related sounds (*e.g.,* blender, water boiling). When SoundWeaver notifies Sam that "*someone is calling her name*," she effortlessly transitions to Social Mode to speak with her colleague. This enables Sam to interact naturally with her environments based on her current needs and successfully complete her tasks (Figure 2; SoundWeaver).

We deployed and evaluated the SoundWeaver prototype in two of Declan's routine environments: his home and the game store he frequently visits. Through these field evaluations, we gained further insights into the potential design of AI sound awareness systems for DHH people's sensemaking of complex auditory information and how introducing novel sound awareness technology can foster new dynamics among DHH users, technology, and the situated environment (*e.g.*, interactions with a group of friends).

Overall, this work makes the following contributions:

[1] We use Declan as a pseudonym for our DHH participant.

1. We describe the iterative prototyping of SoundWeaver, an intent-driven AI sound awareness system that facilitates DHH people's real-time sensemaking. SoundWeaver adapts its information display based on DHH users' intents and purposefully weaves AI outputs based on DHH people's personal contexts.
2. We reflect on the considerations and tensions during the prototyping and field evaluation of SoundWeaver with a DHH co-designer and present design implications for future AI sound accessibility systems.

## 2 RELATED WORK

### 2.1 DHH Culture

The Deaf and hard of hearing community is a diverse group encompassing individuals marked by a wide range of experiences, backgrounds, and cultural identities. There are three primary models of understanding hearing loss: medical, social, and cultural-linguistic models [10, 49, 65]. The medical model views hearing loss primarily as a condition to be diagnosed and treated [76]. The social model shifts the focus from individual hearing loss to societal barriers limiting participation [47, 76]. The cultural-linguistic model recognizes Deafness as a unique cultural and linguistic identity rather than simply a disability. This model celebrates Deaf culture, defined by shared values, norms, and languages like American Sign Language (ASL) [10, 50]. ASL is a developed visual-spatial language with its own syntax, grammar, and nuances and can convey complex ideas, emotions, and narratives [41, 53, 77].

Our work draws from established concepts, models, and theories in accessibility for DHH people to guide our co-design process with Declan. For example, *DeafSpace*, a conceptual framework for creating accessible physical environments, outlines space design guidelines that provide "full access to communication" and unique considerations for DHH people's cognitive, sensory, and emotional experiences [48, 78]. Even though DeafSpace primarily guides space design, these principles have design implications for AR- or HMD-based sound awareness technologies like SoundWeaver, where users perceive sound information as part of the physical space. For example, given that Deaf people perceive their surroundings through subtle visual cues [78], HMD-based sound indicators should communicate the changes in acoustic environments, such as SoundWeaver's use of changing colors and waveforms to indicate real-time ambient noise levels. The caption feature in SoundWeaver also follows established "best practices" to ensure readability [79], including placing captions at the bottom center of the view, using sans serif fonts with "medium thickness," and placing captions in a semi-transparent text box.

### 2.2 Towards Sound Accessibility with AI

HCI researchers have long studied sound awareness solutions to improve sound accessibility for DHH people. Early work explored visualizations based on sound characteristics like location, loudness, and pitch [23, 42, 63] and simple sound classification with shallow learning approaches like support vector machines [36] and decision trees [38]. Driven by recent advances in deep learning models for sound classification (*e.g.*, Convolutional Neural Networks [21, 52] and Recurrent Neural Networks [17]), HCI researchers have developed home, mobile, and wearable AI-based sound recognition systems [8, 31, 32, 60] that process audio signals from the environment and display information about recognized sound events. For example, SoundWatch [32] informed DHH users of environmental sounds through haptic feedback and visual notifications containing sound events and loudness. To mitigate AI's inherent uncertainties (*e.g.*, recognition errors), more recent mobile sound recognition systems enabled DHH people to teach the system through audio samples [33] and provide feedback [14]. Besides sound recognition, advances in automatic speech recognition also led to transcription applications in mobile devices [70, 73, 74] (*e.g.*, Google Live Transcribe [69]). To fulfill both sound

recognition and transcription needs, Guo *et al.* proposed an AR prototype that combined sound classification and ASR outputs in one interface [18].

Despite the progress, current AI-based sound awareness systems assume discrete, pre-configured outputs, which struggle to fulfill DHH users' dynamic and personalized sound information needs across different contexts (*e.g.,* driving and at work) [24]. While the conglomeration of multiple pieces of sound information [18] partially addresses this need, current visual representations remain static (*e.g.*, a textual description of sounds) regardless of the user's needs and context (*e.g.,* always showing transcriptions of the crowd speech in a coffee shop even though the user wants to focus on work). In contrast, prior work on non-sound related accessible technologies has explored new human-AI interaction designs that embodied the concept of "AI extenders," where AI is closely intertwined with human cognition to enhance information processing capabilities [20]. Examples of such applications include the scene-weaving concept, an interaction metaphor that presents information as strands of fabrics that could be "weaved" together into the precepted scene by individual blind and low-vision (BLV) users [2]. Similarly, Morrison *et al.* proposed "open-ended AI" as a facility for BLV children to make sense of social situations based on various spatial audio cues [45].

The current work extends this line of research to the sound awareness technology space. Specifically, we propose an intent-driven design for AI sound awareness systems, where various kinds of AI-based sound feedback are allocated purposefully to adapt to DHH people's real-time contexts and intents and complement DHH people's trusted ways of sensemaking of the environment. The scenario of Sam working in a coffee shop, illustrated in Section 1, exemplified the vision of such systems.

### 2.3 XR-based Sound Feedback Design

Our decision to implement SoundWeaver as a head-mounted display (HMD) application was informed by prior findings that, compared to traditional mobile form factors, HMD could provide a diverse set of always-on, easier-to-access sound information while reducing attention splits and the need to carry the device with hands [16, 29]. Prior work explored the 3D display of sound information, including captions, localization, sound sources and events, and visualization of acoustic signals. For example, one pioneering work conducted a design probe of visual feedback for sound information [28] with a Google Glass-based system. This work elicited user preferences across several dimensions of visual sound feedback, including arrow-based indicators for directionality, peripheral positioning of indicators, and the inclusion of loudness data. To address challenges in communication for DHH people, Jain *et al.* explored HMD-based captions on HoloLens and suggested designs that adapt to real-world contexts like light conditions and convey this contextual information (*e.g.*, speaker name and locations). SpeechBubbles [51] focused on the accessibility of group conversations by probing DHH people's preferred designs for in- and out-of-view conversations in Mandarin.

Regarding VR environments, researchers explored multimodal sound feedback ranging from visualization [11, 30, 39, 40] to sound modifications [9] and haptics [11, 30, 44]. For example, Jain *et al.* categorized sounds into dimensions such as sound source and sound intent and designed corresponding visual and haptic prototypes for VR sound feedback, such as waveforms for ambient sounds and textual displays for currently playing sounds (*e.g.*, torch crackling) and rhythmic haptics for critical information. SoundVizVR [39] built on this work and further examined the usability of different indicator designs for sound types and characteristics. EarVR+ [44] attaches physical LED lights and vibro-motors to traditional devices to inform DHH users of the localization results.

Building on prior 3D sound feedback designs, we carefully examined these designs by situating them in DHH people's personal contexts and preferences. This process allowed us to observe how sound indicators behaved over time and could produce unexpected results. For example, we found that peripheral arrow indicators, a design suggested by a prior

design probe [28], produced distracting visual flickers during rapid speaker transitions, such as in instances of turn-taking or overlapping dialogue, which impeded users' ability to focus on the conversation.

## 3 GENERAL METHODOLOGY

**Positionality Statement:** Our team comprises five researchers. The first author, Jeremy, is a graduate student at the University of Michigan who is hearing and has speech-related disabilities. Jeremy is learning ASL and, at the time of writing this paper, had two years of experience engaging with the DHH community. Jaylin is a graduate student at the University of Michigan who is hearing and had five years of experience researching accessible technologies. Liang-Yuan is hearing and had one year of experience engaging with the DHH community. Cecily had over ten years of research experience working with people with disabilities. Dhruv identifies as hard of hearing and had over ten years of experience engaging and researching with the DHH community. Dhruv has a Level 2 fluency in American Sign Language. Our collective experience as a mixed-hearing ability research team shapes our work, including a deep understanding of DHH culture, the evolution of our design artifact, the analysis of research data, and in-depth discussions with Declan, our participant.

**Our participant "Declan":** Declan is a 22-year-old male who identifies as Deaf. He has profound and non-congenital hearing loss; he lost hearing in his left ear in childhood and his right ear when he was 20. Declan prefers to communicate with both DHH and hearing people in sign language; his primary sign language is American Sign Language and Pidgin Signed English. While he frequently uses sign language, he can also read and communicate in English well. Declan was a good fit for our study because of three considerations. First, Declan was an early adopter of common accessible technologies that DHH people use, like Google's Sound Notification and Live Transcribe, which we presume would make him comfortable with adopting new technologies and giving critical feedback. Second, identifying as Deaf, Declan deeply understood the DHH community's cultural norms and preferences. Third, as a Deaf individual who conversed in both speech and ASL and with both hearing and Deaf people, Declan offered a unique perspective from the angles of both Deaf and hard of hearing populations. In our study findings, we describe other details about Declan, such as his occupation, hobbies, and personal contexts, in our study findings.

**Research Methodology:** We narrate our six-month co-design process with Declan from the Research through Design (RtD) [66] perspective. RtD combines design science with scholarly research, where the creation of artifacts is the research outcome and a means of generating new knowledge about how people use interactive technologies [66]. In our work, we created the SoundWeaver prototype as a design artifact that embodies intent-driven AI and a medium for continuous and critical reflections on our design decisions along the journey.

Our co-design process with Declan consisted of three phases. Phase 1 of our study was primarily formative, where we learned about Declan's personal contexts and routines and probed Declan's information needs across these contexts. In Phase 2, we continuously engaged with Declan through in-person co-design sessions and frequent communications of design ideas and considerations through text messages, emails, and video calls. During this phase, SoundWeaver evolved from a "brute-force" prototype [3] based on formative insights and our prior assumptions to a more polished, carefully designed experience. Phase 3 consisted of the field evaluations of the SoundWeaver prototype and general reflections on the entire research process.

Throughout the co-design process, we curated five types of data: meeting transcripts, sketches, emails/text messages, field notes, and video recordings. Our data collection and analysis process was deeply inspired by the Grounded Theory methods (GTM) [6, 13, 46] to curate and analyze this data. Since the start of our research, we have kept a working document as the collection of memos using a shared Google Docs file. Initially, this document contained our assumptions on the "ideal" sound awareness systems based on our prior knowledge. We kept the document updated whenever there

was new data or "quick thoughts." For example, when we received the meeting transcript for a design session, we encoded the transcript following open, axial, and selective coding and compared the newly generated codes with the current ones. The research team regularly jotted down thoughts on existing data through margin comments, particularly to relate to and compare Declan's experiences with those of our hard-of-hearing author. By the end of Phase 3, the number of open codes in our working document expanded significantly from 78 (from Phase 1) to 231. We did not group these codes into themes; instead, as is common in GTM, we carefully examined the relationship between important codes and iteratively developed the design goals. The number of these design goals fluctuated as the new data could reinforce, add to, or invalidate the current goals.

Our decision to engage closely with one DHH participant, Declan, was influenced by several factors. First, as reflected in Jain's pioneering autoethnographic work as a DHH traveler [26], getting in-depth longitudinal experiences from marginalized populations who might be less willing to travel could be challenging. Second, prior studies have demonstrated that when designing and introducing novel multi-algorithm AI systems, starting with a "deep engagement" with a single person and community might be the more responsible approach that allows close and critical examination of that person's perspective and environments from a more involved standpoint [45, 55].

We proceeded with the study only after obtaining IRB clearance and Declan's consent with IRB-approved consent forms.

## 4 PHASE 1: FORMATIVE STUDY

The goal of Phase 1 was to understand (1) Declan's personal contexts and routines, (2) Declan's current approaches to making sense of the environments, and (3) challenges across his daily routines and tasks due to hearing loss.

### 4.1 Method

Following the general methodology outlined in Section 3, we now detail the specific procedures for this study.

#### *4.1.1 Procedure*

We kicked off Phase 1 with a 150-minute meeting with Declan in our research lab on the University of Michigan campus. Before the session, Declan completed a background questionnaire about his demographic and hearing loss-related information. We introduced Declan to the logistics of co-design and the goal of creating a sound awareness system that facilitates his sensemaking of auditory environments. We stressed to Declan that he would be an active co-designer instead of a participant simply providing feedback to a system.

To understand Declan's personal contexts, the first author asked Declan about his daily routines on weekdays and weekends and visualized them on a chart paper with post-it notes (Figure 3). The first author then collaborated with Declan to examine the visualized routines by breaking them into *individual tasks*. For example, the "cooking dinner" routine was broken into tasks such as grabbing ingredients from the fridge, chopping and preparing ingredients, heating potatoes using microwaves, pan-frying, etc. The tasks were visualized using sketching tools (*e.g.*, colored pens; Figure 3). For each task, the researcher discussed Declan's challenges due to hearing loss and the existing strategies he applied to help address the challenges. For instance, in the "cooking dinner" routine, we highlighted the "heating potatoes" task and asked, "*How did your hearing loss make heating potatoes in the microwave difficult?*" and "*What kind of information would be helpful?*" The researcher repeated this in-depth analysis for each of Declan's routines.

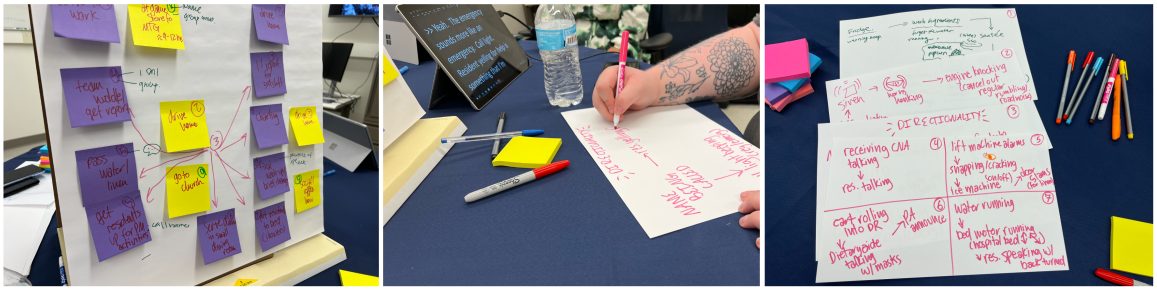

**Figure 3: Sketches from Phase 1's Formative Study. From left to right: (1) Declan's personal routines and contexts, (2) Declan sketching the "frames" and corresponding information needs within the routines, and (3) the numerous sketches of Declan's information needs across personal contexts.**

After the session, the first author kept in touch with Declan through text messages and emails. Following the analysis detailed in the following section, our team scheduled a 45-minute video call with Declan, during which he answered additional questions that surfaced from the analysis (*e.g.*, the spatial layouts of Declan's home and workplace).

#### *4.1.2 Analysis*

Phase 1 elicited five types of research data about Declan's personal contexts, routines, and experiences as a DHH person: transcripts from meetings and video calls, video recordings, sketches, and text messages and emails. We stored these data in the same Google Drive folder for convenient cross-referencing. We analyzed the data following the open, axial, and selective coding methods specified in Section 4. Specifically, the first author walked through the data and generated 57 open descriptive codes in the working document. During the walkthrough, the first author considered all routines and tasks to gain a holistic understanding of Declan's personal contexts and sensemaking approaches across contexts. The second and last author reviewed these open codes and added 21 more. We note that the last author related the open codes to his experience as a DHH person by commenting on the codes in the document. The research team met four times to translate open codes into insights that guided the design of the sound awareness system prototype. The team intermittently added their thoughts to the working document between meetings.

### 4.2 Findings

We first highlight the findings that contributed to eliciting design goals and considerations before discussing the latter. These findings came from our analysis of the raw results to carve out Declan's information needs across personal contexts.

#### *4.2.1 Declan's Personal Contexts*

Declan's workday routines include cooking, driving to work, and working at the nursing home. At the nursing home, Declan's routines consisted of team huddles and one-on-one reports, passing water and linens, taking care of patients (*e.g.*, getting residents up for activities, wash-ups, working on the IV pumps), charting, and serving dinner. For non-workdays, Declan would drive to the nearby city to play trading card games with his friends at a game store. Declan would also drive to the local church for the Sunday service and coffee hours.

#### *4.2.2 Varied Information Needs and Sensemaking Intents Across Contexts*

Across personal contexts, Declan had varied sound information needs based on intent:

**Awareness of the Environmental Sounds:** When entering a new environment, Declan usually "*did not know what to expect.*" Therefore, he hoped to get a holistic sense of the new ambient environment: "*Let's say I walk into a room… I want to immediately know how loud the room is. It's almost like a vibe check.*" Once situated in the environment, Declan wanted to know the individual sound events. Most of the time, Declan used visual and haptic feedback to make sense of the events. For example, at home, his partner would stomp on the ground to grab Declan's attention. At his place of work (nursing home), Declan used call lights and other visual feedback (*e.g.,* co-workers' reactions and monitors) to understand when residents needed his attention. However, sometimes those visual cues were not easily glanceable or available: "*When I am working at the station, I would have to frequently pop my head over and see if there are lights down in the hallway.*"

Declan told us that he was familiar with sound recognition systems (*e.g.*, Google's Sound Notifications) but was not impressed by their performance. He used ASL's description of objects as an analogy [25] for understanding sound events: "*Paint a broad stroke first, then describe details.*" Declan first wanted to access more general descriptions of sound events (*e.g.*, direction and real-time sound level) to incorporate other perceptual abilities (*e.g.*, vision) to make sense of the information before knowing specific details. "*Most of the time, I don't want to know what it is. I just want to know that it's here… If I find interest, then I can be like, 'What do you think it is?'*" He added. In addition to environmental awareness, Declan also wanted to increase awareness of the sounds he produced (*e.g.*, whether he was making too much noise or speaking too loudly).

**Active sound monitoring:** Besides awareness of intermittent sound events, sometimes Declan worked on specialized tasks requiring more specific, precise, and instant recognition of sounds. We refer to this need as "active sound monitoring." There were two general scenarios where Declan needed more active sound monitoring. The first scenario concerned fault detection – the presence of abnormal sounds and the absence of expected sounds. For example, Declan described his experience using self-checkout stations:

> *"Scanner beeping… like the 'doo' sounds at the checkout. I cannot tell if I actually scanned the item… like, am I shoplifting right now, or what's going on?"*

The last author, who is hard of hearing, related to this challenge and shared another case of failed fault detection with the research team, where he could not tell when something "*got stuck in the garbage disposal.*" Another scenario category concerned the processes (*i.e.*, sequences of events) with critical acoustic changes that required attention. For example, Declan could not tell when the pianist started or stopped playing during the church service. Similarly, at work, he had difficulty knowing the real-time status of the IV pump – standard beeps signaled normal operation, while melodic chimes indicated potential issues. As a result, he had to "*constantly pop his head*" to look at the monitor.

**Social Interactions:** Declan primarily used American Sign Language in Deaf environments or when communicating with people who used ASL. When interacting with hearing people, Declan used a combination of lipreading and captioning. During the discussion, Declan highlighted three information challenges:

1. Difficulty locating the speaker in predominantly hearing environments, especially in larger and more scattered spaces.
2. Increased reliance on captioning when lipreading became difficult (*e.g.*, communicating with people with "thick facial hair" or facing with his back against him).
3. Difficulty in recognizing people's calls for attention.

#### *4.2.3 Declan's Hearing Loss Created New Dynamics, Interactions, and Expressiveness*

In addition to challenges and information needs, we discussed Declan's existing strategies for understanding the auditory environment. This conversation surfaced valuable insights on how Declan created access intimacy [4, 67, 68] with social

interactions and used creative expressions as strategies to work and live around hearing loss. These findings sparked our reflections on how our novel sound awareness system could potentially impact Declan's relationships with technology, the environment, and the people around him.

Declan described his experience with the piano music played during the church service:

> *"Our church pianist, Anita, knows that I can't hear. So, if she is done playing and I am looking at her, she will look at me and nod her head or take her hands off the keys."*

Due to his limited access to the music, Declan did not know when the music would start (so he could pay attention). To address this, the church pianist and Declan turned to gestures (*e.g.*, exaggerating the palm being placed or taken off the piano gesture) as accessible visual cues to inform Declan of the start and end of the piano music session. However, there were times when a guest pianist was called to play. "*In that case, it would be nice to have someone telling me,*" Declan told us. Future systems should consider these situations and be careful not to forcibly introduce novel sensemaking processes that might disrupt the current ones.

Declan usually spent his Saturday afternoon playing *Magic: The Gathering* with his friends at the game store. When Nathan, Declan's friend who is Deaf and uses hearing aids, joined him, Declan trusted Nathan as the companion for achieving collective sound awareness:

> *"My friend, Nathan, is Deaf and uses ASL. He also has hearing aids. So, I am basically clocked out when hanging out with him. I'm like, OK, I am not going to be on the lookout for anything… if you hear something, you will let me know, because you know I can't hear anything. Hearing people don't always remember that."*

In this case, Declan and Nathan's interdependent relationship was more personal, intimate, and potentially effective for achieving sound awareness. We took this scenario into account in the design of our artifact by allowing Declan to stop the system's sensing behaviors and defer the task to his trusted companions.

To avoid being mistaken as ignoring other people, Declan used T-shirt graphics as a creative means to inform other people of his Deaf identity:

> *"A lot of people think I am being rude when I don't hear them… If I am going out grocery shopping or at the Pride event, I have T-shirts that have the 'I love you' ASL sign and says, 'I am not ignoring you, I'm Deaf.'"*

Declan's graphic T-shirts are a great example of how accessibility solutions can be expressive. Assistive technologies have traditionally prioritized functions over forms [34]. However, recent studies have found that people with disabilities consider aesthetics to be an important consideration when choosing ATs [34, 54]. Affirming this finding, we hope it sparks a new conversation about expressiveness and creativity in the design and use of accessibility tools.

#### *4.2.4 Design Goals and Reflections*

Our mixed-ability research team critically reflected on the Phase 1 findings and used them to shape the design goals of future sound awareness systems. We emphasize that these design goals are not merely direct products of Declan's experiences but a higher-level reflection of what a future sound awareness system should look like inspired by Declan's grounded experiences and established knowledge in DHH accessibility (*e.g.*, DeafSpace [48, 78]). We describe the design goals as well as the specific findings and thoughts that inspired them:

**DG1: The display of sound information should adapt to Declan's ever-changing intents across personal contexts.** Across Declan's personal contexts, we categorized the intents into three broad categories: awareness of environmental sounds, active, detailed monitoring of sounds related to specific tasks (*e.g.*, checking out at the grocery store, cooking, etc.), and social interaction—each intent category called for different kinds of awareness of sound information, leading to DG1.

**DG2: The delivery of sound information should complement Declan's trusted sensemaking approaches instead of replacing them.** Declan briefed us about his way of processing sound information: "*Paint a broad stroke first, then describe details.*" Specifically, he learned the general sound information (*e.g.*, loud noise) first, followed by specific details (*e.g.,* a loud blender sound) only when necessary. As designers, we should respect DHH people's trusted sensemaking approaches, avoid redundant information, and ensure that the system is compatible with them.

**DG3: Sound awareness systems should be designed with mindfulness toward their influence on social dynamics and connections.** While prior studies studied how social contexts could impact the acceptability of AI-based sound awareness systems [15, 24, 31], we bring to attention these systems' impacts on DHH people's existing social dynamics. For example, Declan had established unique access intimacy [67] through "collective access" of sound information with his friend Nathan and personal cues from the church pianist Anita, an important part of his Deaf identity. Aligning with DG3, the design of SoundWeaver should support these interactions, not replace them.

**DG4: The system should promptly visualize anomalies and other notable changes in the auditory environments.** Across different contexts, Declan has used his perception of visual cues to make sense of the environments (*e.g.*, using call lights to indicate if residents needed attention), which was consistent with prior theories in the Deaf Culture [78]. However, these real-world visual cues were not always accessible (*e.g.*, call lights being out of sight). Sound awareness technologies could reduce this barrier by providing always-on and glanceable visualizations of the changes or anomalies in the ambient acoustic environments.

### 4.3 High-Level System Design and Specifications

Based on the design goals and critical reflections on Declan's personal contexts, we curated an initial system specification that would potentially inform the design of SoundWeaver. We hypothesized that the system design would be **intent-driven**, with three *modes* of weaving sound information to accommodate three broad intents for sound awareness highlighted in DG1. Users could freely switch among the three modes based on their real-time information needs. We list the three weaving modes and specify the user intents each mode supports:

- **Awareness Mode.** Awareness mode fulfills DHH people's needs for **awareness of overall auditory environments and individual sound events** specified in Section 4.2.
- **Action Mode.** Action Mode facilitates **active sound monitoring**, supporting more focused tasks that require continuous awareness of task-relevant sounds.
- **Social Mode.** Social Mode supports DHH people's **social interactions** while maintaining the necessary awareness of the auditory environments (*e.g.*, knowing the smoke alarm going off when chatting with friends).

The above three modes were directly mapped to the three categories of Declan's intents across personal contexts. On a high level, the three modes streamlined different kinds of awareness by weaving AI outputs to fulfill DHH users' dynamic information needs. Moreover, switching among the three modes would be an effortless process with minimal interactions needed. We envisioned the design of each mode to match Declan's preferred way of processing auditory information to fulfill the corresponding intent (DG2) and operationalized this vision through a co-design process with Declan in the following section.

## 5 PHASE 2: CO-DESIGN OF THE SOUNDWEAVER PROTOTYPE

Phase 2 consisted of a series of co-design activities that led to the creation of the SoundWeaver prototype.

### 5.1 Initial Prototype

Guided by the high-level system design, we began with an initial prototype that embodied the design goals listed in Phase 1. The goal of this initial prototype was not to provide a polished design solution but rather to serve as a starting point for our co-design process.

#### *5.1.1 Design*

The initial SoundWeaver prototype contained three buttons indicating the three modes of weaving sound information: Awareness, Action, and Social. Users could switch among the modes by pressing any of these buttons, and the SoundWeaver interface automatically reorganized information to match the new mode without the need for further customization (DG1).

**Awareness Mode:** The Awareness Mode contained three types of sound information: sound level, recognized sound events, and the textual description of the overall acoustic environment. To help DHH people notice changes in acoustic environments visually (a primary environmental sensemaking approach specified in DeafSpace [78]), we used dynamic, color-coded tooltip indicators to represent four discrete sound levels: quiet, ambient, loud, and very loud. We also interfaced with Audio Flamingo [35], an audio language model, to present textual descriptions of auditory environments. Labels of recognized sound events were hidden by default, but users could manually toggle their display on and off. This design aligned with DG2, which called for designs that complement Declan's preferred way of processing sound information (*i.e.*, "*Paint a broad stroke first, then describe details.*")

We present a vignette describing the potential use of this mode: when Declan entered a coffee shop, he noticed the loudness indicator turning from green ("quiet") to blue, showing an "ambient" label. To help him know what to expect in this environment, he asked the system to briefly describe the ambient acoustic scene (*e.g.*, "Multiple people chatting"). When Declan was fully situated in the context, he used the loudness indicator to help him judge if any event had occurred. When he noticed the loudness indicator turning from blue (ambient) to orange (loud), he toggled on the display of recognized sound events and saw a "blender" sound. He looked around and saw that the barista was blending smoothies.

**Action Mode:** Users could configure the task by assigning relevant sounds to monitor in a companion app. Once configured, all the selected sound labels would be pinned to the interface. If a sound occurred, the corresponding label would turn green; otherwise, it stayed gray. For example, Declan programmed the "cooking" task to monitor water boiling, microwave done, and sizzling sounds. These three sound labels were then pinned. When the microwave was done heating Declan's food and elicited a 5-second chime, the "microwave done" label turned green for 5 seconds.

**Social Mode:** Social Mode facilitated social interactions by providing captions while preserving awareness of environmental sounds. Informed by prior work on HMD sound visualizations [28], we used arrows as directional indicators of active speakers. Social Mode also included live transcriptions with medium-bold font and transparent backgrounds to ensure the visibility of texts (as suggested by best practices for captioning [79]) and the physical space. Finally, the initial design in Social Mode displayed labels of recognized sound events below the transcription.

#### *5.1.2 Implementation*

The prototype contained two components: the front-end interface and the back-end server. The front-end interface was implemented as a visionOS application based on an Apple Vision Pro running visionOS 1.2. The back-end server was based on an iPhone 13 Pro Max running iOS 17.3 and Firebase [80]. Due to the limited sampling range of iPhone and Vision Pro microphones, the system used an external clip-on microphone (DJI Mic 2). Real-time audio streams collected by the microphone were transmitted to the iPhone through Bluetooth. The iPhone handled the signal (*e.g.*, calculating sound levels) and AI (*e.g.*, sound classification) processing locally, except for acoustic scene understanding tasks.

Specifically, we used Apple's SoundAnalysis framework [72] to achieve real-time sound classification and Apple's Speech framework [81] to achieve live captioning. We configured the system to recognize 96 sound classes based on Declan's preferences and prior work probing DHH people's desired sound events [8, 15, 32]. We hosted the Audio Flamingo model on Google Cloud Platform [82] as an API for acoustic scene understanding tasks. The prototype interfaced with the Audio Flamingo model by uploading a 5-second audio clip with the prompt "*Please provide a description of the audio.*" The output of the model contained a textual description of the audio clip. The four discrete sound levels in Awareness Mode were: (1) Quiet: less than -50 dBFS, (2) Ambient: -50 dBFS to -30 dBFS, (3) Loud: -30 dBFS to -10 dBFS, and (4) Very Loud: -10 dBFS to 0 dBFS.

Users could configure Action Mode behaviors on the iPhone with a companion iOS application, which transmitted the configured tasks to Firebase as JSON data. All the real-time data, including sound recognition results, transcriptions, and sound level, were also relayed to Firebase. We implemented SoundWeaver's interface as a head-mounted display (HMD) application because HMD was rated as one of the preferred form factors for mobile sound awareness systems [15]. Declan reinforced this prior finding, telling us that "*having to pull up the phone*" was tedious and that having direct access to sound information "*through something like Google Glass*" would be "*extremely helpful.*" However, we acknowledge that the Vision Pro's form factor can be cumbersome for daily use and reiterate that our prototype served primarily as a design artifact for exploration.

### 5.2 Method

#### *5.2.1 Procedure*

Phase 2 consisted of a series of co-design activities to refine the SoundWeaver prototype iteratively. We invited Declan to the second 150-minute in-person meeting in our research lab. We first introduced Declan to the state-of-the-art AI for sound awareness, including sound classification, speech-to-text, and acoustic scene understanding. We discussed the capabilities, outputs, and limitations of these systems and emphasized that the raw outputs could be adapted based on his personal contexts and intents – the umbrella goal of the SoundWeaver prototype.

Then, Declan experienced the initial SoundWeaver prototype through a guided demonstration of the Vision Pro device. The demonstration was not for system evaluation; rather, it primarily focused on introducing the three weaving modes and the available AI tools (*i.e.*, sound classification, speech-to-text, and acoustic scene understanding) as design materials for the upcoming co-design activities. We also used this demonstration to familiarize Declan with visionOS and spatial applications to ensure that Declan could comfortably interact with our prototype in the future. After the demonstration, we asked Declan, "*What are your thoughts on the three modes?*" Declan approved this layout, saying: "*I like that you separated it into the three different [modes]. It was getting nebulous with all the [sound information] we talked about before.*"

With Declan's support on the three-mode design, we dove into the detailed user experience prototyping. Before the process, we reminded Declan that the goal of our design was not to replace Declan's perceptual and cognitive abilities. Instead, we were co-designing a tool for enhancing and complementing his existing ways of sensemaking. We revisited the sketches about Declan's personal contexts and routines from Phase 1 and discussed Declan's preferred sound information across the routines/frames, suitable modes for the desired information, and how the information should be presented on the head-mounted display. We used an iPad instead of paper as the canvas for sketching because it allowed us to use images of environments like living rooms and kitchens to help simulate a first-person view through an AR display.

Throughout the co-design process, we encountered several instances where the promises of the system's capabilities conflicted with technical constraints and user experience considerations. Our approach involved transparent

communication about these conflicts, followed by collaborative problem-solving with Declan to develop practical solutions. For technical challenges, we focused on workable alternatives. For example, upon realizing that ASR performed poorly in noisy settings, we decided that the system should explicitly inform Declan, allowing him to utilize lipreading as a "backup option." Similarly, Declan proposed a feature that pinpoints the direction of the person calling him. Given that this level of localization was currently difficult to achieve, we decided that using speech recognition to detect his name would be more reliable. For user experience challenges, the first author carefully stated and visualized the design limitations and worked with Declan to explore the alternatives. For example, we considered a design that would position sound events according to their physical directions (*e.g.*, sounds originating from the left would appear on the left side of the screen). However, we later determined this design would cause visual clutters and unnecessary distractions.

Following the initial meeting, we maintained ongoing communication with Declan via text messages over two months while iterating on the SoundWeaver prototype. We also provided regular asynchronous updates on the system's development progress and solicited Declan's feedback during this time. We also presented Declan with multiple design variations for specific features, asking him to evaluate each option and provide his rationale for any preferences.

#### *5.2.2 Analysis*

Phase 2 data consisted of meeting transcripts, text messages, and digital sketches about SoundWeaver's potential interfaces across Declan's personal contexts and routines. To guide the next iteration of the SoundWeaver prototype, we followed the same open, axial, and selective coding process on the mixed-format material as in Phase 1. The research team met weekly to discuss the codes and translate the insights into concrete design decisions for the prototype. This analysis yielded 91 new open codes.

### 5.3 Findings

#### *5.3.1 Reflections and Changes in Design Goals*

We reflect on the co-design session and present two additional design goals that emerged from it.

**DG5: Sound awareness systems should inclusively support individuals with intersectional disabilities and diverse identities.** During the co-design session, Declan disclosed that he was neurodivergent. As he explained, "*If I had too much information, I would go into a meltdown,*" highlighting how neurodivergent individuals can experience heightened sensitivity to busy or cluttered visual stimuli [12]. This revelation prompted our critical reflections on the broader implications for the system design. We recognized that similar considerations should extend to users with various combinations of disabilities and identities, such as DeafBlind individuals or DHH people with cognitive abilities. Recently, intersectional disabilities have received increasing attention in accessible technology research. For example, Harrington *et al.* advocated race and ethnicity as important constructs for integrating racial equity in accessibility work [19]. We echo this advocacy through the proposal of DG5 and hope that this work can spark more conversations on this important topic.

**DG6: The system should carefully handle inference-based or interpretive information and not replace DHH people's reasoning process by forcefully interpreting outputs.** When demonstrating the acoustic scene understanding feature, we played the dog barking sounds from YouTube. The system presented the output: "*I hear an anxious dog barking.*" Declan was unsure about the model's behavior: "*It was attributing reasons to be the sounds… I was like, okay, the computer has no way of knowing that.*" We reflected on this sentiment and agreed that AI sound awareness systems should not replace DHH people's reasoning process by forcefully presenting interpretive outputs. This point was also reflected in the prior work, where ASL interpreters demonstrated that when it comes to interpreting "unknown

sounds," it was important to "show, not tell. [25]" Several studies in the broader accessibility research argued that assistive technologies should strive to be a "solution to a sensory problem" rather than the primary source of information [58, 62], a point echoed by Declan:

> *"We are using this as an accessibility tool – and it's not like, 'we want to know everything about everything.' Basically, we want access to the same information that hearing people have. And when a hearing person is in another room and hears a crashing sound, they are not going to know what the crash was about. So, I would not want the computer to try to figure out what the crash is about. I will go and investigate it."*

Other findings reinforced our existing design goals. In the initial prototype design, Action Mode used the alternation of colors to indicate the presence (green) and absence (gray) of sounds. Declan voiced concerns about this design in Phase 2 because it would alter the existing way Declan processed information:

> *"It's a tricky one because that design is a lot… I have a very specific way of processing information, and I don't think about the sounds that aren't there because I have never been in a situation where it's necessary. You are trying to help us, not change us. It's important to not impose a different way of processing information."*

This feedback reiterated the importance of designing sound awareness systems compatible with DHH people's trusted sensemaking approach (DG2). It also illustrated how designers' preconceptions can lead to "invasive" designs that are counterintuitive to marginalized communities and interfere with their existing ways of living – a pattern that resembles "colonization in design [83]." Furthermore, this experience reaffirmed the necessity of engaging in sustained, iterative co-design processes with target user communities to develop truly beneficial solutions.

#### *5.3.2 Prototype Evolution*

Throughout Phase 2, we worked closely with Declan to continuously iterate the SoundWeaver prototype. Here, we present its design evolution, also visualized in Figure 5.

##### *5.3.2.1 Awareness Mode Evolution*

**Sound level indicator:** In Awareness Mode, we replaced the discrete, colored-coded sound level indicator with a waveform that visualizes the sound levels for the most recent five seconds. This design change was based on Declan's preferences for symbolic over textual representations: "*Pictures are better than words. Deaf people born deaf typically have a lower rate of being able to read English.*" Moreover, we discovered that when sound levels fluctuated near the thresholds, the indicator colors frequently changed, which could cause considerable discomfort for neurodivergent DHH users.

**Displaying sound event labels.** In the initial prototype, once the display of sound recognition results was toggled on, it would remain active until manually deactivated. This design conflicted with Declan's preference for receiving "*occasional inputs*" about the identities of sound events. He felt that the more natural way for him was to have the system "*telling him something is going on*" and ask him if he wanted to know the names of the recognized sounds. Based on this feedback, we implemented a new mechanism: when the system detected a spike in sound level, a prompt "Show Sound ID?" would appear. Once the system received a "go-ahead" from the user, it displayed sound event labels for five seconds. Otherwise, the prompt would fade away after five seconds.

**Acoustic scene understanding.** Based on Declan's feedback about the system's overinterpretation of the acoustic scene (*e.g.*, attributing reasons), we adjusted the prompt to make the information more descriptive rather than analytical (*i.e.*, "*Please provide a neutral description of the audio without any extra details or interpretations in a short sentence.")*, aligning with DG6. We also fixed the output to shorter phrases to reduce the visual clutter and prevent information overload (DG5).

#### 5.3.2.2 *Action Mode Evolution*

The initial Action Mode design displayed all task-related sounds continuously, using green highlights to indicate sound occurrence. Based on Declan's feedback specified in Section 6.3, we implemented a more selective approach: users could designate specific sounds for continuous monitoring by "starring" them, which would pin these sounds to the heads-up display. Other sound events would only show up when they occurred. The research team debated whether the pinning system should be removed entirely but decided to keep it as an optional feature because Declan stated that visualizing the presence and absence of sounds could be helpful in certain situations (*e.g.*, ensuring the water was boiling).

#### 5.3.2.3 *Social Mode Evolution*

We omitted the sound recognition results and replaced them with the same sound level visualization waveform used in the Awareness Mode. This design change was inspired by two realizations. First, Declan emphasized that focusing on the transcription took precedence over identifying specific sounds in social settings like a game store. Nevertheless, he still wanted to maintain some degree of environmental awareness during conversations. Second, we recalled from the formative study that Declan's trusted friends would inform him of important sounds, a common practice in Deaf communities [78]. This dynamic effectively addressed Declan's sound identification needs, making system-generated recognition results unnecessary. This change aligned with both Declan's desired sound awareness approach (DG2) and preserved the access intimacy [67] between him and his trusted companions (DG3).

The caption component initially featured white text on a transparent background, based on our assumption that this design would minimize visual clutter and maintain Declan's visual awareness. However, Declan suggested enclosing the captions in a semi-transparent box to enhance readability while preserving the visibility of "*things behind it.*" He noted that other DHH individuals might have different preferences.

Recognizing the automatic speech recognition's performance limitations in noisy environments, we implemented a warning system that notified users about potential captioning accuracy issues when it detected elevated ambient noise levels. This design aligned with established guidelines for human-AI interaction [1, 75] about supporting the "efficient dismissal" of AI services.

#### 5.3.2.4 *Overrides for Critical Sounds*

During the co-design, we realized that critical sounds, like emergencies and name-calling, could be overlooked in the initial prototype due to the lack of explicit alerts. Reflecting on prior work in DHH sound awareness [8, 15, 24, 32], we concluded that critical and safety-related sound information (*e.g.*, emergency sounds and name-calling) should be displayed regardless of the current mode and updated the prototype accordingly (Figure 6; 4A and 4B).

## 6 PHASE 3: FIELD EVALUATION

In Phase 3, we deployed the second iteration of the SoundWeaver prototype to Declan's routine environments. We considered field deployment an integral part of the co-prototyping process because AI carried inherent uncertainty in its capability as a unique design material [61, 64]. Moreover, evaluating SoundWeaver in the field allowed us to probe two critical questions. First, how could "intent-driven" AI systems fulfill DHH people's dynamic information needs? Second, how would the SoundWeaver prototype blend into the intricate social and environmental dynamics across Declan's personal contexts? Exploring these questions in the wild allowed us to elicit new design opportunities for effective human-AI interaction design for sound awareness systems.

## 6.1 Method

### *6.1.1 Procedure*

We evaluated the second-iteration SoundWeaver prototype across two distinct personal contexts: Declan's home and a game store he frequently visited. We did not evaluate the prototype at Declan's workplace (*i.e.*, nursing home) due to the employer's concerns about the privacy of nursing home residents. We utilized a portable microphone (DJI Mic 2) for audio capture and transmission to address the audio sampling limitations inherent in Vision Pro and iPhone microphones.

At home, we first provided a guided demonstration of the SoundWeaver prototype and asked about Declan's first impression. Declan then used the system while performing three general tasks: engaging with his partner and pets, cooking, and participating in conversations. We simulated possible at-home sounds by knocking on the door, playing smoke alarm sounds through the phone, turning the faucet on and off, dropping books onto the floor, etc. We observed the interactions by mirroring the Vision Pro screen on an iPad. This session lasted for two hours.

At the game store, Declan used SoundWeaver while playing *Magic: The Gathering* with his four friends while sitting around a table. The first author observed the interactions from the side of the table and took field notes. The Vision Pro screen was mirrored on the first author's iPad as in the previous session. This session lasted for 2.5 hours.

Following each session, we conducted semi-structured interviews to learn about Declan's experience using the system and gather feedback on the sound information's quantity, presentation, and contextual appropriateness.

**Figure 4: Declan uses the SoundWeaver prototype when cooking at home (left) and playing card games with friends (right).**

### *6.1.2 Analysis*

Phase 3 data contained handwritten field notes and transcripts from two field evaluation sessions generated by Google's voice recorder app. We again followed Grounded Theory-inspired approaches (*i.e.*, open, axial, and selective coding) to analyze these materials. Phase 3 elicited 62 new codes. The research team met twice to analyze the open codes and compare them to the existing data.

## 6.2 Findings

### *6.2.1 Notes from Field Evaluation*

We present notable snippets of Declan's interactions with the SoundWeaver prototype at home and the game store that sparked reflections among the research team and guided the final iteration of the SoundWeaver prototype.

When Declan was petting his cats on the couch using the system, we called Declan's name to initiate a conversation. Upon seeing the system message "*Someone may have called your name,*" Declan looked up and manually activated Social Mode by tapping the "Social" button, transitioning from Awareness Mode to access the caption feature. In subsequent feedback, Declan noted that this mode-switching process felt "unnatural" and desired a more seamless transition. He suggested incorporating the name-calling message into an interactive button that would automatically activate Social Mode when selected.

When we simulated a visitor by knocking on the door, Declan's dog started barking. Declan was confused about the considerable fluctuations in the waveform. Instead of prompting the system to show sound recognition results, Declan immediately looked around to locate the sound source. When asked about this interaction, Declan explained that looking around was "*much faster*" than "*tapping a button to know what's going on.*"

In Phase 2, we changed the delivery of sound recognition results from manual toggles to a prompt-based delivery (*i.e.,* "Show Sound ID?") triggered by sudden sound level spikes. After extensive usage at home, Declan told us that the sound recognition results were "*much more refined*" than he expected, and he "*would not mind having it constantly on,*" ultimately favoring the manual toggle-based approach.

During the *Magic: The Gathering* game, Declan primarily used Social Mode due to the game's conversational nature. He noted two main concerns with the interface. First, the fluctuating waveform created a visual distraction that made it difficult to focus on the active speaker. Second, given the small group size (four players), Declan found the directional arrows indicating active speakers unnecessary and distracting, particularly during rapid speaker transitions. He suggested that displaying speaker names would be more helpful than directional indicators in small group settings.

Before the game session, the first author anticipated potential caption performance issues in two specific scenarios: multiple speakers talking at the same time and when the active speaker was positioned at a distance (*e.g.,* sitting diagonally across the table). These concerns were validated during the session. However, Declan and his companions devised an effective solution by implementing a "talking stick" approach, passing the portable microphone among speakers. This method notably enhanced both caption accuracy and reduced latency.

We also worried that inaccurate or incomplete captions might adversely impact Declan's understanding of the conversation and raised this concern with Declan. We asked Declan if the system should stop displaying captions to allow Declan to focus on lipreading when the caption performance became subpar. To our surprise, Declan told us that even when the caption was inaccurate, it identified some important words he failed to catch with lipreading; thus, he preferred leaving the caption on regardless of its performance. The session concluded after two hours when Declan removed the Vision Pro due to fatigue.

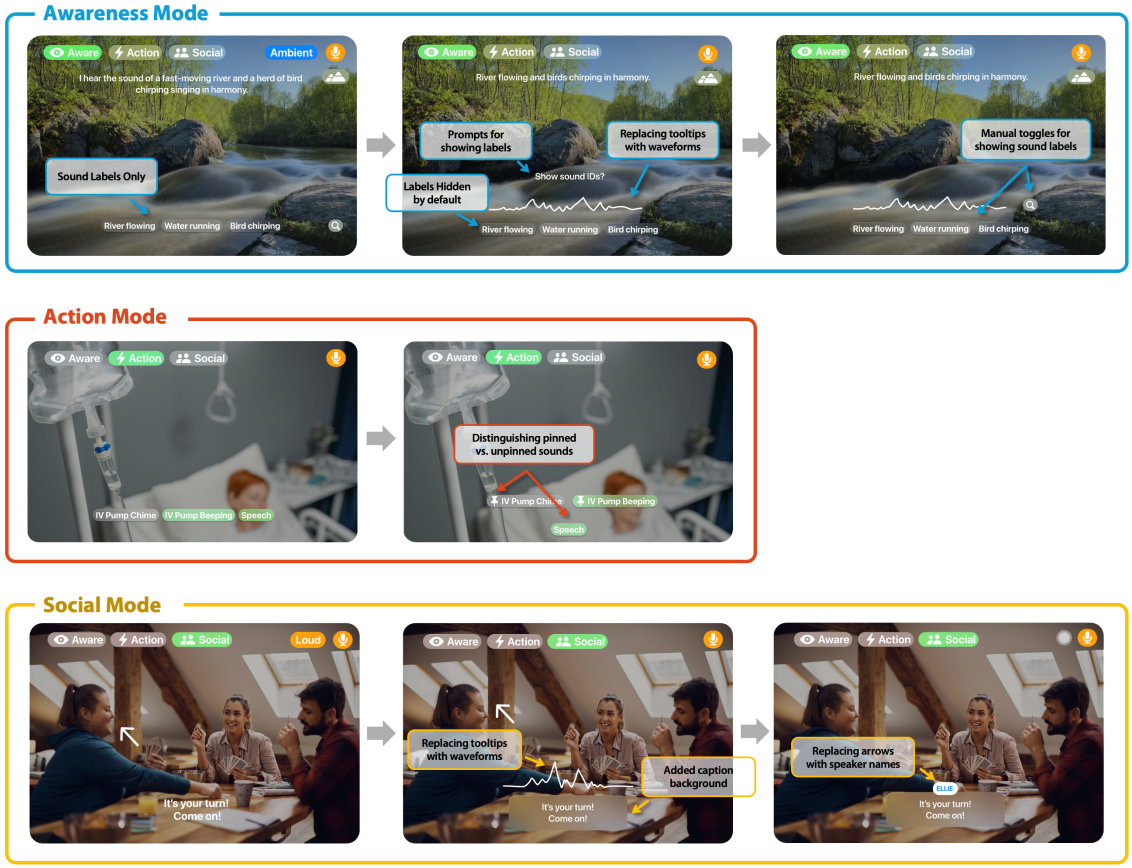


**Figure 5: The Design Evolution of the SoundWeaver Prototype Across Design Phases. Each color-coded section indicates the design iterations for one of SoundWeaver's three modes (Awareness in the blue section, Action Mode in green, and Social Mode in yellow). We also use floating tooltips with border colors that match the corresponding mode to highlight changes in the current iteration. For example, in the second iteration of the Social Mode design, we replaced the colored text box with waveforms as indicators for ambient sound level and added a semi-transparent background to the caption. Similarly, in the third iteration of Social Mode, we removed the arrow pointing to speakers and replaced it with speaker names to avoid visual distraction.**

*6.2.2 Reflections, Design Goal Updates, and Prototype Evolution*

Based on the findings, we made one update to DG1:

**DG1: The display of sound information should adapt to DHH people's ever-changing intents across personal contexts. Moreover, the adaptation process should require minimal effort from the users.**

The change to DG1 was inspired by Declan's comment that the current method of switching modes and retrieving sound recognition results felt "unnatural" and required constant manual inputs. This concern was also reflected in prior work, where DHH users expressed willingness to provide inputs to sound awareness systems but indicated that repeatedly asking for manual inputs would eventually lead to them "giving up" altogether [24].

We made three changes to the prototype. First, as Declan suggested, we made the name-calling alert interactive, enabling an automatic transition to Social Mode when tapped. We applied the same interaction pattern to the "speech" label in sound recognition results. Second, to address Declan's concerns about waveform fluctuations creating visual

distractions during conversations, we replaced the waveform-like sound level visualizer with a "sound bubble" positioned in the corner of the screen. The bubble expands as the sound level increases, providing a more subtle and less intrusive representation of sound. Third, we reverted Awareness Mode's sound recognition results to the Phase 1 design, reinstating the manual toggle feature. Our curated design goals across three co-design phases are presented in Table 1, while the final SoundWeaver prototype is shown in Figure 6.

Table 1. Our final curated design goals for intent-driven sound awareness systems to support Declan.

| CODE | DESIGN GOALS |
|---|---|
| **DG1** | The display of sound information should adapt to DHH people's ever-changing intents across personal contexts. Moreover, the adaptation process should require minimal effort from the users. |
| **DG2** | The delivery of sound information should complement DHH people's trusted sensemaking approaches instead of replacing them. |
| **DG3** | Sound awareness systems should be designed with mindfulness toward their influence on social dynamics and connections. |
| **DG4** | The system should promptly visualize anomalies and other notable changes in the auditory environments. |
| **DG5** | Sound awareness systems should inclusively support individuals with intersectional disabilities and diverse identities. |
| **DG6** | The system should carefully handle inference-based or interpretive information and not replace DHH people's reasoning process by forcefully interpreting outputs. |

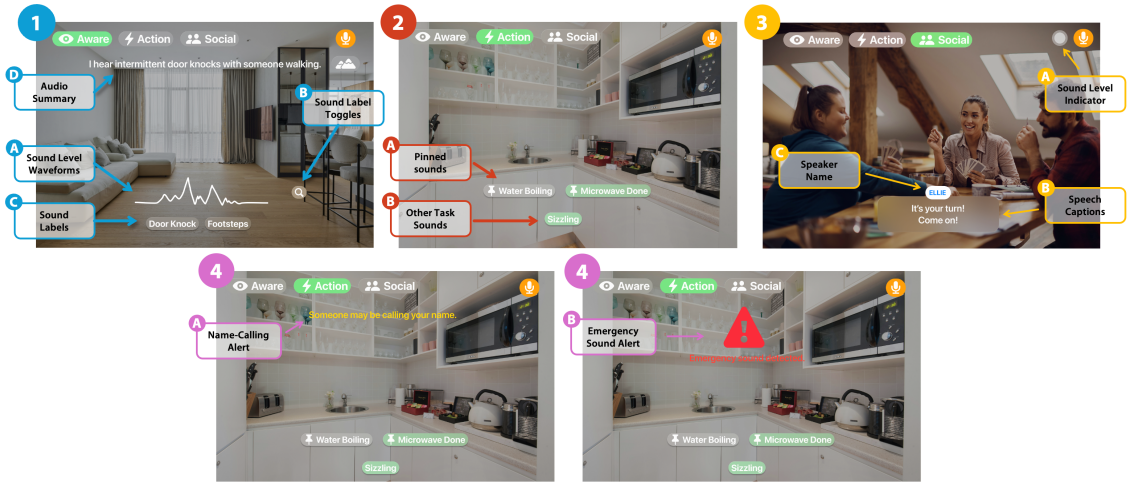


**Figure 6: The Final Design of the SoundWeaver Prototype. In Awareness Mode (Interface 1), the system displays a waveform indicating sound levels by default. Users can tap the sound label toggle (1B) to see the sound labels (1C). Users can also request a textual description of the auditory environment (1D). In Action Mode (Interface 2), the system displays task-related sounds (2A and 2B). Users can pin certain sounds (2A) to monitor them continuously, prompting the interface to reflect their absence and presence. Social Mode (Interface 3) displays the sound level as a pulse-like circular indicator (sound bubble) (3A), along with the speech caption (3B) and speaker name (3C). SoundWeaver will push name-calling alerts (4A) and emergency sound alerts regardless of the current mode.**

## 7 DISCUSSION

Here, we summarize and contextualize key findings in prior research, discuss further implications of our work, and state study limitations.

### 7.1 Intent-Driven Design of AI Sound Awareness Systems

Our design artifact, SoundWeaver, demonstrated how AI sound awareness systems can provide different information-presenting interfaces for individual-specific intents (as specified in DG1). The ability to modulate AI system behaviors based on Declan's intents encouraged his active and meaningful interactions with AI rather than imposing a predefined framework for the relevance of auditory information. For example, when we initiated a conversation, Declan's information priority shifted from environmental awareness to social interactions. Then, he switched from Awareness to Social Mode to match this new intent. This approach differs from previous non-personalized systems, where the AI outputs dictated the information received by DHH users (*e.g.*, [24]) and were agnostic to users' real-time information needs. We note that the three modes in the current prototype are tailored to Declan's individual experiences. Future work should engage with the wider DHH community to discover more intents requiring different ways of "weaving" sound information.

When comparing our approach to prior sound awareness systems that have explored end-user customization [8, 32, 33], we argue that the key difference lies in the abstraction of AI system behaviors. For example, ProtoSound [33] provides an interface that enables users to adapt the sound recognition system to recognize sounds in their personal contexts (*e.g.*, chimes from a DHH user's custom microwave). SoundWeaver's interface, on the other hand, abstracts away the complexity of customization and maps AI behaviors to Declan's self-knowledge of tasks to be accomplished (*e.g.*, to help me with [an intent], AI should do [behavior]). We do not argue for the superiority of either approach; instead, we encourage future work to envision an accessibility tool with the best of both worlds: how can AI systems be designed to not only align with user intents but also offer effective mechanisms for users to intervene, correct, or refine the system's behaviors?

As human-AI alignment gains traction in HCI research [5, 7, 43, 56], we hope our work inspires new conversations about developing AI-based accessibility tools that adapt to users' personal contexts and goals, particularly since AI systems have traditionally prioritized the majority [57]. However, designing intent-driven systems can be challenging because intents are "implicit feedback" [57] that can be difficult to observe; as a workaround, our system requires Declan to remain constantly aware of his real-time intents and assess whether the system aligns with it. The need for manual mode switching adds another layer of complexity, increasing the cognitive demand. We encourage future work to examine designs that better capture the implicit feedback and align with user intents without requiring constant interactions.

Another important consideration throughout the co-design process is the visual design of sound feedback. Specifically, we asked: What kind of information is necessary and, more importantly, relevant to the Declan's intents? How should this information be presented to Declan? These considerations become more important in head-mounted display (HMD)-based interfaces, as poorly designed visuals can easily lead to distraction, fatigue, and discomfort. Morrison *et al.* proposed *information density* [45] as a key factor in effectively supporting the social sensemaking of a blind child with HMD-based AI systems, as overly dense information can overwhelm blind and low-vision (BLV) users. In our case, designing low-density interfaces helped minimize visual distractions and prevent "*meltdowns*" caused by information overload for Declan. Additionally, we observed *design variability*, a term we defined to describe how SoundWeaver's sound indicator designs (*e.g.*, waveforms for ambient sound level and arrows indicating active speakers) respond to unpredictable changes in the acoustic environment. During the field evaluation at the game store, we noticed that the waveforms indicating ambient sound levels fluctuated drastically – an example of high design variability – which distracted Declan from focusing on the conversation. To address this, we replaced the waveform visualizer in Social Mode with a more subtle bubble-shaped peripheral visualizer (see Figure 6-3). Declan also reflected that in smaller and more intimate settings (*e.g.*, four friends sitting around a table) or situations where speakers frequently change, he

would prefer seeing the speaker's names in captions instead of directional indicators, as name changes would be less visually distracting than moving arrows.

### 7.2 Addressing the Invasiveness of Sound Awareness Technologies

A persistent theme throughout the co-design process was whether the design was "invasive." Here, we interpret invasiveness in two critical aspects:

1. Can our design interfere with the current social dynamics in Declan's circle?
2. Will our design disrupt Declan's existing sensemaking processes?

Regarding social dynamics, we were concerned that Declan's usage of the SoundWeaver, which ran on the Apple Vision Pro headset, would make him self-conscious due to its intrusive form factor, as suggested by prior work [58, 59]. However, we were pleased to find that Declan felt comfortable using the system around his friends at the game store and that the friend group supported his system usage (*e.g.,* passing along the microphones to increase the automatic caption's accuracy). Declan's experience aligns with prior findings that DHH people generally feel more comfortable using sound awareness technologies around closer social circles (*e.g.,* family members) [15, 24]. Moreover, the friend group's supportive actions demonstrated how introducing novel assistive technologies can foster new social norms rooted in interdependence [4].

Beyond social acceptability, we carefully considered how SoundWeaver could disrupt Declan's intricate, sometimes intimate, social fabric based on his hearing loss and Deaf identity. For example, when designing interfaces for the context of "attending church service," we pondered whether introducing SoundWeaver would disrupt Declan's personal bond with Anita, the pianist who accommodates him by signaling the start and end of her music pieces through hand gestures. Another example of this dilemma arose in our discussion about the name-calling alert. We connected this feature to Declan wearing a graphic T-shirt that tells other people about his Deafness. While facilitating the recognition of name-calling can be beneficial in social events, this technology may diminish the need for Declan's existing workarounds, which are often far more expressive, creative, and social. These concerns, along with careful discussions with Declan, shaped our decision to only visualize ambient loudness peripherally in Social Mode, deferring most critical sound awareness tasks to Declan's trusted companions in the game store. Motivated by these concerns about SoundWeaver's impacts on social dynamics, we encourage future researchers to consider two factors when designing AI-based accessibility tools. First, when AI systems work as expected, does their adoption come at the cost of *access intimacy* [67, 68]? Second, in the spirit of Amershi *et al.* [1], who argue that AI systems should "gracefully degrade their services when encountering errors," can we design AI systems that, when encountering unexpected results, leverage a broader, interdependent support system to help users achieve their goals?

The second aspect of invasiveness emerged from the design artifact's negotiations with Declan's existing sensemaking models. Initially, we envisioned SoundWeaver as a "cognitive extender," [20] anticipating that this AI application could be tightly integrated and internalized into Declan's sensemaking process. However, our co-design sessions revealed a more nuanced perspective. While Declan appreciated the complementary knowledge SoundWeaver provided (*e.g.,* using captions for filling lipreading gaps), he firmly resisted the idea that the tool should be internalized or fundamentally alter his way of perceiving the world. This insight prompted us to shift our design approach from cognitive extensions to *augmenting* Declan's sensemaking abilities. The key difference lies in the potential risk when the system fails or becomes unavailable: the breakdown of extenders can result in a sum loss of the user's ability [20]. Our design strategy thus evolved from assumptions about "what is helpful" to a more considered approach that prioritizes the alignment with Declan's trusted sensemaking strategies, aligning with DG2. This shift was particularly

evident in the iterative developments of Action Mode, where we offered Declan more agency to selectively monitor sound events.

Besides the above two aspects, we note that current HMD devices may also be invasive due to their form factors. For example, our study used Apple Vision Pro, a headset similar in size to ski goggles and weighing about 650 grams. During the field evaluation, Declan chose to take off the Vision Pro device after two hours of use due to fatigue.

### 7.3 Reflecting on the Co-Design Process

To our knowledge, this work is the first study that synergizes user experience prototyping with the expertise of a DHH individual, leveraging their deep understanding of DHH culture, personal contexts, and sound information needs within a longitudinal co-design process. Here, we present three critical reflections on this process.

First, viewing Declan as an equal contributor to the design artifact allowed us to engage in thoughtful discussions that led to a system more closely aligned with Declan's cultural background and sensemaking abilities. During the formative study, Declan stated that the most important part of sound awareness for him was knowing the *presence* of sound. During the analysis, we connected this preference with DeafSpace, which emphasizes how Deaf people rely on subtle environmental cues to maintain awareness and navigate spaces [48]. This led us to ask, "How can we convey the presence of sound events through visual cues?" Since Declan identified loudness as a strong indicator of "*something happening*," we initially used color-coded textual indicators (*e.g.*, quiet, ambient, loud) to represent real-time sound levels. In Phase 2, Declan told us that "*pictures are better than words*" for him and suggested a waveform-based design to better convey temporal changes in loudness (*e.g.*, a sudden spike indicating a sound event).

Second, while the above examples demonstrate how an informed user experience prototyping process can help align AI systems' behaviors with the user's sensemaking approaches in accessibility tools, we caution future designers about the potential discrepancies between how users naturally process information and their interactions with AI. For example, in our second-iteration prototype, the system prompted Declan to confirm whether he wanted to access sound classification results whenever a spike in the sound level was detected. The classification results appeared only when Declan affirmed by tapping the prompt. While this interaction matched Declan's natural sensemaking process ("*Paint a broad stroke first, then describe details*"), during the field evaluation, he found it tedious. Upon realizing that the sound recognition model was more capable than expected, Declan preferred the manual toggle-based approach employed in the initial design, which persistently displayed sound event labels but allowed quicker access and dismissal. In this case, Declan's increased trust in AI capabilities shifted his priorities from compatibility with his natural sensemaking approach to a preference for efficiency. We encourage future work to consider this discrepancy, especially in technology design for marginalized communities, where responsibly balancing cultural norms, values, and usability is crucial.

Third, evaluating the SoundWeaver prototype in Declan's real-world contexts elicited design insights and challenges that may be difficult to acquire from lab settings. For example, while our initial design incorporated peripheral arrow indicators – theoretically supported as effective directional cues [28] – the rapid transitions among active speakers in group settings created distracting visual effects that impeded Declan's ability to focus on conversations.

### 7.4 Limitations and Future Work

While our design goals are deeply informed by the perspectives of one Deaf participant, our hard-of-hearing co-author, and relevant prior work in DHH culture and accessibility, we do not claim that they are exhaustive or will work as intended for all DHH people. We present these design goals as starting points for further exploration of intent-driven systems and eagerly look forward to future work that refines or expands them with feedback from the broader DHH community. While the SoundWeaver prototype was designed to be generalized beyond our Deaf participant's use cases,

we recognize that future work needs to evaluate and iterate the prototype through longitudinal studies with diverse members from the DHH population, further validating its utility and usability and refining its design. Since our work is primarily qualitative, we did not evaluate the performance of the Audio Flamingo model, which our system used for generating textual descriptions of the acoustic scene. We encourage future work to assess its performance and explore how it can involve human-AI interaction to ensure accuracy across diverse contexts.

While acknowledging the above limitations, we note that Research through Design with one participant is a powerful HCI methodology that has been used to elicit rich, situated insights for designing culturally conscious systems that accommodate diverse needs, particularly in the field of accessibility (*e.g.*, in [27, 45]), where population-wide preferences vary widely. As Jain *et al.* [27] argue, such insights are often difficult to obtain through traditional multi-participant studies.

## 8 CONCLUSION

Our work presents the co-design process of SoundWeaver, an intent-driven AI sound awareness system prototype that supports DHH people's sensemaking of auditory environments, developed in collaboration with a DHH participant. Reflecting on this design journey, we discuss its implications for the development of future sound awareness systems and other AI accessibility tools, particularly regarding the alignment of system behaviors with users' intents and personal contexts. We also highlight important considerations for designing socially and ethically mindful technologies to address accessibility challenges.

## References


[1] 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*, May 02, 2019. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300233

[2] Harshadha Balasubramanian, Cecily Morrison, Martin Grayson, Zhanat Makhataeva, Rita Faia Marques, Thomas Gable, Dalya Perez, and Edward Cutrell. 2023. Enable Blind Users' Experience in 3D Virtual Environments: The Scene Weaver Prototype. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. ACM, Hamburg Germany, 1–4. https://doi.org/10.1145/3544549.3583909

[3] Behance. Blog:: The "Brute Force" School of Design: Prototyping over Presentations. Retrieved September 6, 2024 from https://www.behance.net/blog/the-brute-force-school-of-design-prototyping-over-presentations

[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*, October 08, 2018. ACM, Galway Ireland, 161–173. https://doi.org/10.1145/3234695.3236348

[5] Elfia Bezou-Vrakatseli, Oana Cocarascu, and Sanjay Modgil. 2024. Towards Dialogues for Joint Human-AI Reasoning and Value Alignment. https://doi.org/10.48550/arXiv.2405.18073

[6] Melanie Birks and Jane Mills. 2010. *Grounded Theory: A Practical Guide*. SAGE.

[7] Angie Boggust, Benjamin Hoover, Arvind Satyanarayan, and Hendrik Strobelt. 2022. Shared Interest: Measuring Human-AI Alignment to Identify Recurring Patterns in Model Behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (*CHI '22*), April 28, 2022. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3491102.3501965

[8] 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*, October 23, 2016. ACM, Reno Nevada USA, 3–13. https://doi.org/10.1145/2982142.2982171

[9] Xinyun Cao and Dhruv Jain. 2024. SoundModVR: Sound Modifications in Virtual Reality to Support People who are Deaf and Hard of Hearing. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '24*), October 27, 2024. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3663548.3675653

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

[11] Pratheep Kumar Chelladurai, Ziming Li, Maximilian Weber, Tae Oh, and Roshan L Peiris. 2024. SoundHapticVR: Head-Based Spatial Haptic Feedback for Accessible Sounds in Virtual Reality for Deaf and Hard of Hearing Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '24*), October 27, 2024. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3663548.3675639

[12] Seungwon Chung and Jung-Woo Son. 2020. Visual Perception in Autism Spectrum Disorder: A Review of Neuroimaging Studies. *Soa Chongsonyon Chongsin Uihak* 31, 3 (July 2020), 105–120. https://doi.org/10.5765/jkacap.200018

[13] Tom Cole and Marco Gillies. 2022. More than a bit of coding: (un-)Grounded (non-)Theory in HCI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, April 27, 2022. ACM, New Orleans LA USA, 1–11. https://doi.org/10.1145/3491101.3516392

[14] Hang Do, Quan Dang, Jeremy Zhengqi Huang, and Dhruv Jain. 2023. AdaptiveSound: An Interactive Feedback-Loop System to Improve Sound Recognition for Deaf and Hard of Hearing Users. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*, October 22, 2023. ACM, New York NY USA, 1–12. https://doi.org/10.1145/3597638.3608390

[15] 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*, May 02, 2019. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300276

[16] 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* (*CHI '19*), May 02, 2019. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300276

[17] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech Recognition with Deep Recurrent Neural Networks. Retrieved July 19, 2024 from http://arxiv.org/abs/1303.5778

[18] 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 *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '20*), October 29, 2020. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3373625.3418031

[19] Christina N. Harrington, Aashaka Desai, Aaleyah Lewis, Sanika Moharana, Anne Spencer Ross, and Jennifer Mankoff. 2023. Working at the Intersection of Race, Disability and Accessibility. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '23*), October 22, 2023. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3597638.3608389

[20] José Hernández-Orallo and Karina Vold. 2019. AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, January 27, 2019. ACM, Honolulu HI USA, 507–513. https://doi.org/10.1145/3306618.3314238

[21] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. 2017. CNN Architectures for Large-Scale Audio Classification. Retrieved July 19, 2024 from http://arxiv.org/abs/1609.09430

[22] Jaylin Herskovitz, Andi Xu, Rahaf Alharbi, and Anhong Guo. 2024. ProgramAlly: Creating Custom Visual Access Programs via Multi-Modal End-User Programming. https://doi.org/10.1145/3654777.3676391

[23] 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.

[24] Jeremy Zhengqi Huang, Hriday Chhabria, and Dhruv Jain. 2023. “Not There Yet”: Feasibility and Challenges of Mobile Sound Recognition to Support Deaf and Hard-of-Hearing People. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*, October 22, 2023. ACM, New York NY USA, 1–14. https://doi.org/10.1145/3597638.3608431

[25] Jeremy Zhengqi Huang, Reyna Wood, Hriday Chhabria, and Dhruv Jain. 2024. A Human-AI Collaborative Approach for Designing Sound Awareness Systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–11. https://doi.org/10.1145/3613904.3642062

[26] Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E. Froehlich. 2019. Autoethnography of a Hard of Hearing Traveler. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, October 24, 2019. ACM, Pittsburgh PA USA, 236–248. https://doi.org/10.1145/3308561.3353800

[27] Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E. Froehlich. 2019. Autoethnography of a Hard of Hearing Traveler. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, October 24, 2019. ACM, Pittsburgh PA USA, 236–248. https://doi.org/10.1145/3308561.3353800

[28] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (*CHI ’15*), April 18, 2015. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/2702123.2702393

[29] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS ’18*), October 08, 2018. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[30] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. Towards Sound Accessibility in Virtual Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (*ICMI ’21*), October 18, 2021. Association for Computing Machinery, New York, NY, USA, 80–91. https://doi.org/10.1145/3462244.3479946

[31] 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*, April 21, 2020. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376758

[32] 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*, October 26, 2020. ACM, Virtual Event Greece, 1–13. https://doi.org/10.1145/3373625.3416991

[33] Dhruv Jain, Khoa Huynh Anh Nguyen, Steven 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*, April 29, 2022. 1–16. https://doi.org/10.1145/3491102.3502020

[34] Chloe Kent. 2021. Equipment aesthetics: the companies improving mobility aid design. *Medical Device Network*. Retrieved September 12, 2024 from https://www.medicaldevice-network.com/features/assistive-device-design/

[35] Zhifeng Kong, Arushi Goel, Rohan Badlani, Wei Ping, Rafael Valle, and Bryan Catanzaro. 2024. Audio Flamingo: A Novel Audio Language Model with Few-Shot Learning and Dialogue Abilities. Retrieved August 13, 2024 from http://arxiv.org/abs/2402.01831

[36] 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)*, December 2007. IEEE, Sivakasi, Tamil Nadu, India, 544–548. https://doi.org/10.1109/ICCIMA.2007.370

[37] Raja S. Kushalnagar and Christian Vogler. 2020. Teleconference Accessibility and Guidelines for Deaf and Hard of Hearing Users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, October 26, 2020. ACM, Virtual Event Greece, 1–6. https://doi.org/10.1145/3373625.3417299

[38] Yizhar Lavner and Dima Ruinskiy. 2009. A Decision-Tree-Based Algorithm for Speech/Music Classification and Segmentation. *EURASIP Journal on Audio, Speech, and Music Processing* 2009, (2009), 1–14. https://doi.org/10.1155/2009/239892

[39] Ziming Li, Shannon Connell, Wendy Dannels, and Roshan Peiris. 2022. SoundVizVR: Sound Indicators for Accessible Sounds in Virtual Reality for Deaf or Hard-of-Hearing Users. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (*ASSETS '22*), October 22, 2022. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3517428.3544817

[40] Ziming Li, Kristen Shinohara, and Roshan L Peiris. 2023. Exploring the Use of the SoundVizVR Plugin with Game Developers in the Development of Sound-Accessible Virtual Reality Games. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (*CHI EA '23*), April 19, 2023. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3544549.3585750

[41] Scott K. Liddell. 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511615054

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

[43] Malek Mechergui and Sarath Sreedharan. 2024. Goal Alignment: Re-analyzing Value Alignment Problems Using Human-Aware AI. *AAAI* 38, 9 (March 2024), 10110–10118. https://doi.org/10.1609/aaai.v38i9.28875

[44] Mohammadreza Mirzaei, Peter Kán, and Hannes Kaufmann. 2021. Head Up Visualization of Spatial Sound Sources in Virtual Reality for Deaf and Hard-of-Hearing People. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, March 2021. 582–587. https://doi.org/10.1109/VR50410.2021.00083

[45] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex Taylor, Geert Roumen, Camilla Longden, Sebastian Tschiatschek, Rita Faia Marques, and Abigail Sellen. 2021. Social Sensemaking with AI: Designing an Open-ended AI Experience with a Blind Child. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, May 06, 2021. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445290

[46] Michael J Muller and Sandra Kogan. Grounded Theory Method in HCI and CSCW. *Grounded Theory*.

[47] National Research Council (US) Committee on Disability Determination for Individuals with Hearing Impairments. 2004. *Hearing Loss: Determining Eligibility for Social Security Benefits*. National Academies Press (US), Washington (DC). Retrieved August 11, 2024 from http://www.ncbi.nlm.nih.gov/books/NBK207838/

[48] Joan Naturale. InfoGuides: DeafSpace: Principles and Elements of DeafSpace. Retrieved July 5, 2024 from https://infoguides.rit.edu/deafspace/principles

[49] Michael Oliver. 1996. *Understanding Disability*. Macmillan Education UK, London. https://doi.org/10.1007/978-1-349-24269-6

[50] Carol A. Padden and Tom L. Humphries. 1990. *Deaf in America: Voices from a Culture* (58896th edition ed.). Harvard University Press.

[51] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (*CHI '18*), April 21, 2018. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[52] Karol J. Piczak. 2015. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, September 2015. IEEE, Boston, MA, USA, 1–6. https://doi.org/10.1109/MLSP.2015.7324337

[53] Wendy Sandler and Diane Lillo-Martin. 2006. *Sign Language and Linguistic Universals*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139163910

[54] Aline Darc Piculo dos Santos, Ana Lya Moya Ferrari, Fausto Orsi Medola, and Frode Eika Sandnes. 2022. Aesthetics and the perceived stigma of assistive technology for visual impairment. *Disability and Rehabilitation: Assistive Technology* 17, 2 (February 2022), 152–158. https://doi.org/10.1080/17483107.2020.1768308

[55] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2020. Principles to Practices for Responsible AI: Closing the Gap. Retrieved September 6, 2024 from http://arxiv.org/abs/2006.04707

[56] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, Sushrita Rakshit, Chenglei Si, Yutong Xie, Jeffrey P Bigham, Frank Bentley, Joyce Chai, Zachary Lipton, Qiaozhu Mei, Rada Mihalcea, Michael Terry, Diyi Yang, Meredith Ringel Morris, Paul Resnick, and David Jurgens. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. (2024).

[57] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, Sushrita Rakshit, Chenglei Si, Yutong Xie, Jeffrey P. Bigham, Frank Bentley, Joyce Chai, Zachary Lipton, Qiaozhu Mei, Rada Mihalcea, Michael Terry, Diyi Yang, Meredith Ringel Morris, Paul Resnick, and David Jurgens. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. https://doi.org/10.48550/arXiv.2406.09264

[58] 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* (*CHI '11*), May 07, 2011. Association for Computing Machinery, New York, NY, USA, 705–714. https://doi.org/10.1145/1978942.1979044

[59] Kristen Shinohara and Jacob O. Wobbrock. 2016. Self-Conscious or Self-Confident? A Diary Study Conceptualizing the Social Accessibility of Assistive Technology. *ACM Trans. Access. Comput.* 8, 2 (January 2016), 5:1-5:31. https://doi.org/10.1145/2827857

[60] 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. https://doi.org/10.1145/3090082

[61] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. 2021. ProtoAI: Model-Informed Prototyping for AI-Powered Interfaces. In *26th International Conference on Intelligent User Interfaces*, April 14, 2021. ACM, College Station TX USA, 48–58. https://doi.org/10.1145/3397481.3450640

[62] Anja Thieme, Cynthia L. Bennett, Cecily Morrison, Edward Cutrell, and Alex S. Taylor. 2018. "I can do everything but see!" -- How People with Vision Impairments Negotiate their Abilities in Social Contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (*CHI '18*), April 21, 2018. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173777

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

[64] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 21, 2020. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376301

[65] A. M. Young. 1999. Hearing parents' adjustment to a deaf child-the impact of a cultural-linguistic model of deafness. *Journal of Social Work Practice* 13, 2 (November 1999), 157–176. https://doi.org/10.1080/026505399103386

[66] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 29, 2007. ACM, San Jose California USA, 493–502. https://doi.org/10.1145/1240624.1240704

[67] 2011. Access Intimacy: The Missing Link. *Leaving Evidence*. Retrieved July 8, 2024 from https://leavingevidence.wordpress.com/2011/05/05/access-intimacy-the-missing-link/

[68] 2017. Access Intimacy, Interdependence and Disability Justice. *Leaving Evidence*. Retrieved September 7, 2024 from https://leavingevidence.wordpress.com/2017/04/12/access-intimacy-interdependence-and-disability-justice/

[69] Live Transcribe | Speech to Text App. *Android*. Retrieved September 12, 2024 from https://www.android.com/accessibility/live-transcribe/

[70] Get live captions of spoken audio on iPhone. *Apple Support*. Retrieved July 19, 2024 from https://support.apple.com/guide/iphone/get-live-captions-of-spoken-audio-iphe0990f7bb/ios

[71] Google | yamnet | Kaggle. Retrieved August 1, 2024 from https://www.kaggle.com/models/google/yamnet

[72] MLSoundClassifier. *Apple Developer Documentation*. Retrieved August 1, 2024 from https://developer.apple.com/documentation/createml/mlsoundclassifier

[73] Introducing Whisper. Retrieved July 19, 2024 from https://openai.com/index/whisper/

[74] Speech-to-Text AI: speech recognition and transcription. *Google Cloud*. Retrieved July 19, 2024 from https://cloud.google.com/speech-to-text

[75] People + AI Guidebook. Retrieved July 19, 2024 from https://pair.withgoogle.com/guidebook

[76] Medical and Social Models of Disability | Office of Developmental Primary Care. Retrieved August 11, 2024 from https://odpc.ucsf.edu/clinical/patient-centered-care/medical-and-social-models-of-disability

[77] *Linguistics of American Sign Language, 5th Ed.* Retrieved August 11, 2024 from https://gupress.gallaudet.edu/Books/L/Linguistics-of-American-Sign-Language-5th-Ed

[78] DeafSpace - Campus Design and Planning. *Gallaudet University*. Retrieved August 12, 2024 from https://gallaudet.edu/campus-design-facilities/campus-design-and-planning/deafspace/

[79] Captioning Key. Retrieved August 13, 2024 from https://dcmp.org/learn/captioningkey#3

[80] Firebase | Google's Mobile and Web App Development Platform. *Firebase*. Retrieved September 8, 2024 from https://firebase.google.com/

[81] Speech. *Apple Developer Documentation*. Retrieved September 8, 2024 from https://developer.apple.com/documentation/speech/

[82] Cloud Computing, Hosting Services, and APIs. *Google Cloud*. Retrieved September 12, 2024 from https://cloud.google.com/gcp

[83] Colonized by Design - Industrial Designers Society of America. Retrieved September 8, 2024 from https://www.idsa.org/innovation_article/colonized-design/