Dhruv Jain **Research** Statement

My work aims to build **interactive systems to address pressing challenges in accessibility**. I focus on solving ambitious technical problems such as accurately classifying sounds in diverse contexts while also pushing towards deployable solutions with immediate real-world impact. My research covers the entire spectrum from inventing or reappropriating methods to sense physical events, to leveraging signal processing and machine learning techniques to interpret this data, to building interfaces for end-users to interact with this data. As my work is interdisciplinary, I collaborate across the sub-disciplines of computer science (*e.g.*, AI, computing systems) and outside of computer science (*e.g.*, design, psychology).

In my dissertation research, I focused on **advancing sound accessibility for deaf and hard of hearing (DHH) users**, resulting in several deployed systems for visualizing sounds in the home [15,16], providing customized sound feedback on smartwatches [17], and supporting conversation through augmented reality displays [4,10,11].

Outside of my dissertation, I have explored other problems in accessibility such as supporting sound awareness in virtual reality for DHH users [13,14], indoor navigation in diverse contexts for visually impaired users [5,12], and non-verbal mouth sound detection for users with motor-speech disorders [1].

Thus far, I have published 21 papers (14 as first-author) and 13 extended abstracts in premier HCI venues such as CHI, UIST, and ASSETS. My work has been honored with two best paper awards, four best paper honorable mentions, one best artifact award, and has helped procure $1.5m in funding. My systems have been open-sourced, have invited attention from industries Microsoft, Google, and Apple, and have been publicly launched (*e.g.*, one has more than 75,000 users [6]). My work also regularly receives media attention, including from *CNN*, *New Scientist*, and *Forbes*, and is included in teaching curricula—demonstrating broad impact.

ADVANCING SOUND ACCESSIBILITY

The world is filled with a rich diversity of sounds. However, these sounds can be inaccessible to DHH people, leading them to miss important events ranging from actionable cues (*e.g.*, a microwave beep) to safety related cues (*e.g.*, a fire alarm) to cues that make one feel present (*e.g.*, a bird chirp). My main thread of work addresses *how to improve sound and speech awareness for DHH users*. It is driven by my personal experiences as a hard-of-hearing individual [9,19]. Guided by this work, my ultimate vision is to **make sounds accessible to everyone, everywhere**, regardless of the disability, since *sound accessibility*, a field I termed [14], affects us all. We all find conversations difficult to hear in noisy bars, doorbells inaudible over a vacuum running, or may miss a phone ringing on a train.

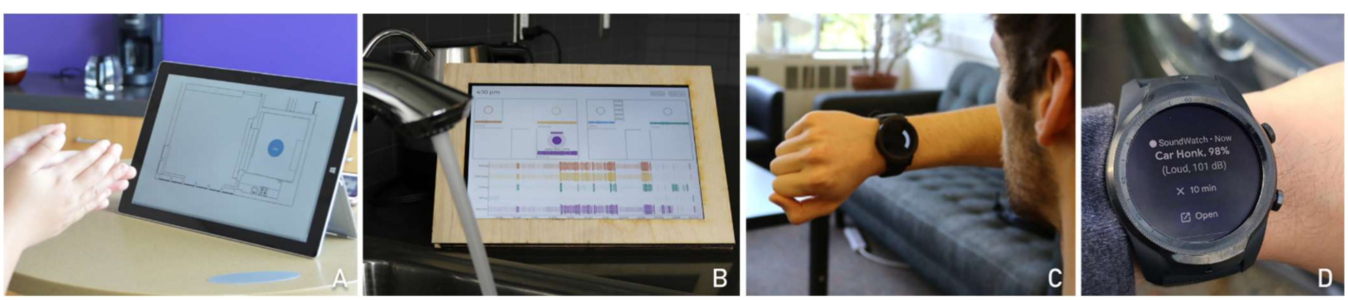

Figure 1: My dissertation research explored sound accessibility for DHH users following an iterative design process from large-scale surveys (not shown) to controlled evaluations of system-designs (A and C) to full system deployments in the field (B and D).

My work in this thread is guided by a large-scale online survey I conducted with 201 DHH participants at the beginning of my PhD [2]. This survey helped identify the most promising research questions to explore that are both technically challenging and also offer the most value to our end-users. In my PhD, I systematically explored several of these questions through an iterative human-centered research process ranging from formative studies to design and evaluation of prototypes in controlled environments to deployments of full systems in the field (Figure 1). Below, I describe four such explorations.

Visualizing sounds in the home. The first exploration addresses *how to classify and visualize sounds in the home*. Commonly used commercial products—such as flashing doorbells or vibratory alarm clocks—substitute specific auditory cues with visual or haptic feedback, but do not provide general awareness about sounds in the home.

In response, through an iterative design process (Figure 1a, 1b) [15], I created *HomeSound* [16] an IoT based system visualizes sound activity in the whole house. HomeSound contains a network of interconnected "picture-framed" displays running a deep-learning engine deployed in different rooms of the home (Figure 1b). Each display senses and classifies 19 common sounds in the home (*e.g.*, dog bark, doorbell) and interacts with a centralized server to produce a single across-home visualization of in-home sound activity.

I iteratively deployed HomeSound in eight homes of DHH people, starting with visualizations of simple but accurate sound characteristics (*e.g.*, loudness, pitch), before deploying more complex but error-prone sound information (*e.g.*, the sound type). After a three-week use, DHH participants reported being able to perform some important daily tasks more easily and quickly (*e.g.*, getting clothes from a dryer, turning off the kitchen fan). For some users, HomeSound fundamentally changed their understanding of sounds in their home. For example,

> *"I became aware of so many sounds in my home: dryer whirring, cutlery clinking, door opening, water running. My wooden home makes a lot of noise when I walk. I didn't realize we have a noisy home. Do we make more noise than hearing people?!"*

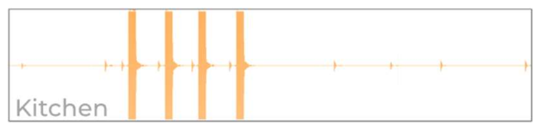

Figure 2: Distinct peaks in waveform were visible on HomeSound displays when a microwave beeped.

Importantly, the accurate visualizations of simple sound properties helped mitigate some errors caused by our imperfect sound recognition engine. For example, a user was able to recognize a microwave beep by using distinctive peaks in a waveform when AI failed (Figure 2). We also uncovered privacy concerns with our always-recording setup in a personal context of the "home". Overall, our findings provide important insights to design AI-based sound sensing systems accounting for concerns of social dynamics, space, privacy, and trust.

Providing sound awareness on portable devices. As the second key area, I explored *how to provide accurate sound awareness in portable environments*. After examining several portable devices, I chose to use smartwatch, the most preferred device by DHH people in our survey [2], since it is private, glanceable, and always-available on the wrist.

However, classifying sounds on a low-resource device and displaying sound information on a small watch screen is a challenging problem. To address, our team performed two smartwatch-based studies. First, we conducted a *Wizard-of-Oz* evaluation of several dense sound feedback designs [3], identifying preferences for visualizations and complementary vibratory patterns to accompany the visualizations—another benefit of smartwatch. Second, using transfer learning, I trained and compared four small deep-learning models to accurately classify sounds on smartwatches [17], identifying a promising model that performed close to state-of-the-art for non-portable devices (82% accuracy), while requiring substantially less memory and computational power.

Based on the above insights, I built *SoundWatch*, a real-time sound recognition app for commodity Android smartwatches [17]. SoundWatch uses our privacy-preserving sound recognition pipeline and can be switched among four system architectures (*watch-only*, *watch+phone*, *watch+cloud*, *watch+phone+cloud*). Through an evaluation with eight DHH users in multiple contexts (*e.g.*, homes, offices, outdoors), I uncovered several implications for future wearable accessibility systems such as the need for 'switchable' system architectures, multiple cascaded models, end-user control of the data, and a customizable user interface.

After improvements, we released the app with our best performing model and system architecture on Google Play Store. Thus far, it has been downloaded over 600 times. This line of work has received extensive press coverage, a best artifact award at ASSETS 2020, and was invited for CACM Research Highlights, which publishes the most significant recent research results across the field of computing. Now, sound recognition comes integrated in both the major smartwatch platforms—Apple Watch and Google WearOS.

Personalizing sound feedback. A key desired feature by DHH people in our survey [2] and system evaluations [16,17] was end-user personalization—such as training on new sound categories (*e.g.*, a new custom home appliance) or users' specific sounds (*e.g.*, my child's voice) [2,16,17]. Since sound accessibility has broad implications, supporting personalization could also help accommodate a diversity of other user groups beyond DHH such as those with auditory hypersensitivity or situationally impaired hearing users. However, AI-based systems typically require a large amount of training data which is difficult to obtain in-situ for custom sounds.

In response, I developed *ProtoSound* [18], an interactive system that allows users to personalize a sound recognition engine by recording only a few training samples (*e.g.*, five for each sound). The key idea is that by training a model repeatedly over datasets of limited training samples, it learns to train rapidly from a few samples in the field.

To evaluate ProtoSound, I quantified performance on real-life sound datasets I compiled [16,17] and conducted a field deployment study with 19 participants. Results show that ProtoSound significantly outperformed the best baseline (9.7% accuracy advantage), required minimal end-user effort to train (~10 minutes), and accurately learned sounds across many real-world locations (*e.g.*, homes, restaurants, grocery stores, and parks). To further reduce user-effort and improve accuracy, I am currently exploring another personalization technique, called *Thompson Sampling,* with plans to integrate the final algorithm into our SoundWatch app on Google Play Store.

Improving speech awareness through augmented reality. I also examined *how to support speech conversations for DHH users* using AR-style head-mounted displays (HMD), which offer benefits of glanceability and reduced visual split over the traditional stationary captioning approaches. Again, an iterative design and evaluation process resulted in several prototype publications, starting with visualizations for spatially locating sounds [10], to initial [7] and refined [11] prototypes for speech transcription, and culminating in a full system [4] that displays multiple cues on an HMD (speech, location, non-speech sounds) (Figure 3). The final system is open-sourced (demo video).

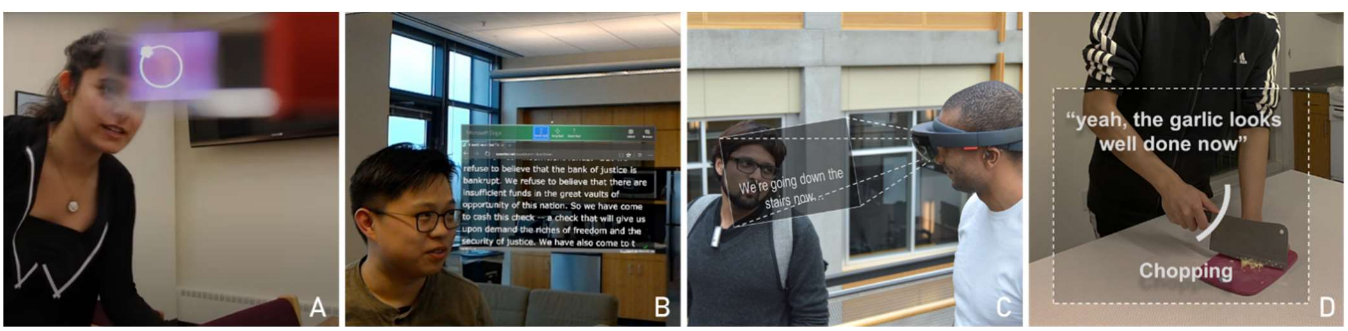

Figure 3: Iterative exploration of speech awareness on heads-up display starting from proof-of-concept localization (A) to an initial captioning prototype (B) to a more refined prototype (C) and, finally, a full system combining captions, location, and sound cues (D).

OTHER ACCESSIBILITY WORK

Outside of my dissertation, I have explored problems in augmented & virtual reality [13,14,20], haptics [8,14], and personalized technologies [9,19]. I briefly describe three selected threads of work with public prototype releases.

Improving sound awareness in virtual reality (MSR internship). Sound is a critical component of VR, needed to increase realism or convey critical cues, but is inaccessible to DHH users. To advance VR sound accessibility, I articulated a taxonomy of VR sounds [13], created a design space to map these sounds to visual and haptic feedback, and designed and evaluated several VR sound feedback prototypes [14]. This is the sole work in VR accessibility for DHH users, which won a best paper award [13] and culminated in open-source prototype releases.

Supporting indoor navigation (Undergraduate thesis). To support visually impaired users in India, I investigated low-cost 'grassroots' indoor navigation systems [6,12,12]. The final system, containing retrofitted $5 wall devices and a mobile app, was deployed in a national science museum for two years and was used by over 75,000 people.

Supporting non-verbal mouth sounds input (Apple internship). To enable people with severe motor-speech disorders to interact with mobile devices, I examined non-verbal mouth sounds (*e.g.*, pop, click) as an alternative input mechanism [1]. The current prototype, released in iOS 15, recognizes 14 sounds with usable accuracy, which can be mapped to user-defined sections such as "select an item" or "go back" to control smartphones.

FUTURE RESEARCH AGENDA

Here, I outline opportunities in three areas I am most excited to pursue.

Sound Sensing and Feedback. While my dissertation targeted DHH users, sound accessibility affects all of us. As described in the Introduction, I envision a future where sounds are accessible to everyone, everywhere. Achieving this vision will require innovations in many fields such as engineering, design, policy, and urban infrastructure. As an HCI researcher, I hope to contribute to design and engineering sides, building on my dissertation work to explore: (1) collecting sound datasets from diverse contexts, (2) building sound awareness systems that generalize across environments, and (3) designing visual and haptic sound feedback interfaces for diverse contexts. I hope to address questions such as: *how to build datasets that are culturally, spatially, and temporally diverse? how to build systems that are robust to ambient noise in different contexts? or, how to design interfaces that personalize to contextual user needs?* To guide my and others' work in the field, I am currently planning a CHI 2022 workshop on sound accessibility and formulating an opinion paper that outlines the history, current trends, and the future grand challenges in the field. Beyond accessibility, providing unobtrusive and glanceable visualizations of sound signals have many applications for many domains such as for home automation or medical imaging, which I also plan to explore.

AR, VR Accessibility. Accessibility has always been an afterthought during the design of modern technologies, often leading to inaccessible or sub-par user experiences. Virtual and augmented reality technologies (XR) are at a crossroads in time when we can still codify accessibility best practices in this emerging medium from the start. Building on my work, such as on VR sound accessibility [13,14] or AR speech awareness [10,11], I will contribute to advancing three key areas in XR accessibility: accessibility of XR content and apps, accessibility of XR input and output devices, and accessibility-focused application areas. For example, I am currently exploring how to sense and show multiple dense sound information (*e.g.*, sound class, speech captioning, and sound source location) on a small VR-display to make VR content and apps accessible to DHH users. Through my work and interactions with the community, I hope to arrive at agreed-upon metadata, standards, and toolkits for making XR accessible to all.

Human-AI Systems. My personalization work leverages end-user interaction in the AI pipeline to adapt a system to users' specific sounds and environment. Such human-AI approaches combine the flexibility of humans with the computational power of AI to solve complex problems that are impossible with either alone. However, many questions remain such as *how to achieve the right balance between user agency and user effort? what device configurations and user interactions are most promising? or, how to quantify field performance given the subjective user involvement?* I will continue to explore these questions, particularly for resource-limited contexts such as accessibility or global south, generalizing the systems to varying geographies, abilities, and culture. To do so, I will leverage my existing connections with the specific communities (*e.g.*, Deaf groups or HCI4SouthAsia, a group I co-lead).

REFERENCES

1. Colin Lea, Zifang Huang, Dhruv Jain, Lauren Tooley, Zeinab Liaghat, Shrinath Thelapurath, Leah Findlater, and Jeffrey P. Bigham. 2022. Nonverbal Sound Detection for Disordered Speech. ***In Submission** to ICASSP 2022*.
2. 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 CHI 2019*, 1–13.
3. Steven M Goodman, Ping Liu, Dhruv Jain, Emma J McDonnell, Jon E Froehlich, and Leah Findlater. 2021. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *In IMWUT 2021*, 2: 1–23.
4. 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 ASSETS 2020*, 1–4.
5. Dhruv Jain. 2014. Pilot evaluation of a path-guided indoor navigation system for visually impaired in a public museum. *In ASSETS 2014*, 273–274.
6. Dhruv Jain. 2014. Path-guided indoor navigation for visually impaired using minimal building retrofitting. *In ASSETS 2014*, 225–232.
7. Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. *In DIS 2018*, 7–11.
8. Dhruv Jain, Brendon Chiu, Steven Goodman, Chris Schmandt, Leah Findlater, and Jon E Froehlich. 2020. Field study of a tactile sound awareness device for deaf users. *In ISWC 2020*, 55–57.
9. Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E Froehlich. 2019. Autoethnography of a Hard of Hearing Traveler. *In ASSETS 2019*, 236–248. **Best paper honorable mention.**
10. Dhruv Jain, Leah Findlater, Christian Volger, Dmitry Zotkin, Ramani Duraiswami, and Jon Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. *In CHI 2015*, 241–250.
11. 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 ASSETS 2018*, 81–92.
12. Dhruv Jain, Akhil Jain, Rohan Paul, Akhila Komarika, and M Balakrishnan. 2013. A path-guided audio based indoor navigation system for persons with visual impairment. *In ASSETS 2013*, 1–2.
13. Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. A Taxonomy of Sounds in Virtual Reality. *In DIS 2021*, 160–170. **Best paper award.**
14. 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 ICMI 2021*, 1–15.
15. Dhruv Jain, Angela Carey Lin, Marcus Amalachandran, Aileen Zeng, Leah Findlater, and Jon Froehlich. 2019. Exploring Sound Awareness in the Home for People who are Deaf or Hard of Hearing. *In CHI 2019*, 94:1-94:13.
16. Dhruv Jain, Kelly Mack, Akli Amrous, Matt Wright, Steven Goodman, Leah Findlater, and Jon E. Frohlich. 2020. HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users. *In CHI 2020*, 1–12.
17. 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 ASSETS 2020*, 1–13. **Best artifact award.**
18. Dhruv Jain, Khoa Nguyen, Steven Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehilch. 2021. ProtoSound: A Personalized, Scalable Sound Recognition System for d/Deaf and Hard of Hearing Users. *In CHI 2022* (to appear).
19. Dhruv Jain, Venkatesh Potluri, and Ather Sharif. 2020. Navigating Graduate School with a Disability. *In ASSETS 2020*, 1–11. **Best paper honorable mention.**
20. Dhruv Jain, Misha Sra, Jingru Guo, Rodrigo Marques, Raymond Wu, Justin Chiu, and Chris Schmandt. 2016. Immersive scuba diving simulator using virtual reality. *In UIST 2016*, 729–739.