Dhruv Jain **Research** Statement

The world is filled with a rich diversity of sounds ranging from mundane beeps and whirs to critical cues such as fire alarms or spoken content. These sounds can be inaccessible not only to people with auditory-related disabilities such as those who are deaf or hard of hearing (DHH) but also to hearing people in many situations. We all may find TV dialogue difficult to hear in noisy bars, doorbells inaudible over a vacuum cleaner running, or may miss a phone ringing on a loud train. Indeed, an estimated 80% of TV captioning users are hearing [BBC, Verizon]. **I envision a future where sound information is accessible to everyone, everywhere**, enabling people to be more aware of their surroundings, easily communicate with others, and seamlessly access critical sound-based information in diverse contexts.

Achieving *sound accessibility* for all is a multi-faceted vision that requires innovation in many fields such as in engineering to build novel sensing and processing systems, in urban infrastructure to collect sound data, and in public policy to manage end-user privacy. As a human-computer interaction researcher, my research contributes to the engineering side of this problem by **developing interactive systems that leverage advances in machine learning, signal processing, and wearable technology to sense and provide sound feedback**. I also draw from my personal experience as a hard-of-hearing individual [5,13] to drive my research directions.

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 [3]). 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

Sound accessibility, a field I termed [8], lies at the intersection of *sound sensing* and *accessibility*. My research contributes *tools and techniques to sense and process sound data* and *interfaces to accessibly convey this data to the end-users*.

Sound Sensing Tools and Techniques. As the first key thread, my work explores sound sensing tools and techniques. For example, *SoundWatch* [11] is a real-time system that senses and alerts DHH users about important sound events (*e.g.*, dog bark, car honk). The core contribution was the development of a deep-learning based sound recognition system that, unlike prior work, could run fully locally on low-resource portable devices, thereby supporting privacy-preserving use in diverse contexts.

Figure 1: Example use cases of our SoundWatch app.

In another project, I explored techniques for end-user personalization of sound feedback. Prior sound awareness systems, including SoundWatch, are trained on large pre-defined sound datasets, and cannot accommodate custom user-specific sounds (*e.g.*, a custom home appliance, or my pet dog's bark)—a key desired feature by our DHH users [10,11]. This is a challenging problem since machine learning typically requires a large amount of training data, which is difficult to obtain in-situ. In response, I developed *ProtoSound* [12], a novel technique that allows end-users to personalize a sound recognition model by recording only a few training samples (*e.g.*, five short samples per 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.

Beyond solving ambitious technical problems, I also focus on providing immediate utility for our end users. Thus, we released the SoundWatch system as a smartwatch app on Google Play Store. Thus far, it has been downloaded over 600 times. While originally intended for DHH users, our app reviews reveal usage by hearing people as well—

e.g., *"to know if someone is ringing the bell when I am wearing headphones"*, demonstrating the broad utility of sound accessibility. This line of work has received extensive press coverage, a best artifact award at ASSETS, and was invited to CACM Research Highlights, which publishes the most significant recent research across computing. Now, sound recognition comes integrated in both the major mobile platforms—Apple iOS and Google Android.

Sound Feedback Design. As the second key thread, my research explores interfaces to provide unobtrusive and glanceable sound feedback to the end-users. The core contribution of this thread includes the design of visual and haptic substitutes of sound that account for perceptual differences among different modalities. For example, in the *HomeSound* system [10], which visualizes sound activity in the home, the persistence or duration of sensed sound events in a home was visualized spatially in a horizontal bar (Figure 2). This choice accommodated the human perceptual differences in sound and vision, which are perceived temporally and spatially respectively.

Figure 2: A HomeSound IoT display deployed in a home. The red box shows the persistence of sound visualized as a horizontal bar.

To design these sound feedback systems, I follow an iterative research process consisting of formative studies, design and evaluation of prototypes in controlled environments, and, crucially, deployment studies of full systems in the field. As an example, Figure 3 summarizes the four iterative prototype publications resulting from a design process to support speech awareness on heads-up displays. This user-centric process ensures that the systems seamlessly adapt to the preferences and the 'lived experiences' of our users. Indeed, the final system (Figure 3D) [2] contains carefully tuned customization options to serve the needs of each DHH user (*e.g.*, modifying caption placement, selecting the list of non-speech sounds).

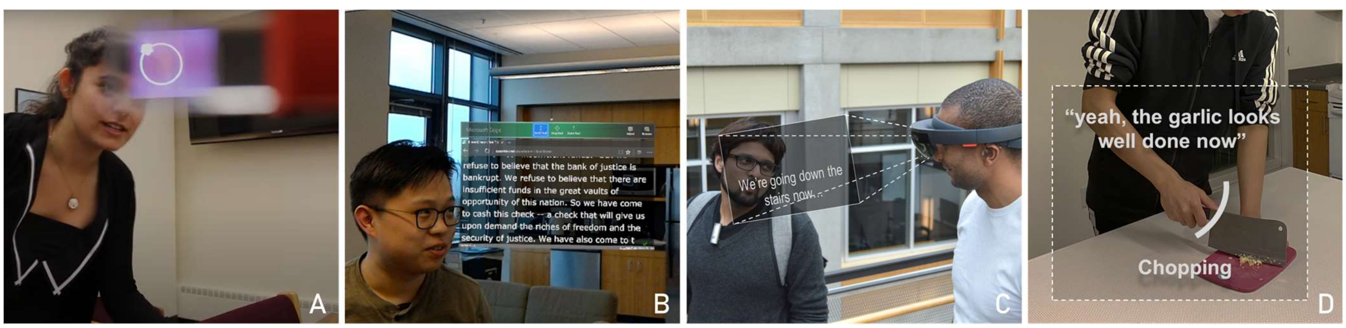

Figure 3: Iterative exploration of speech awareness on heads-up displays starting from proof-of-concept localization (A) [6] to an initial captioning prototype (B) [4] to a refined prototype (C) [7] and, finally, a full system combining captions, location, and sound cues (D) [2].

FUTURE RESEARCH AGENDA

My future vision is to achieve sound accessibility for all. The fields of computer science and information systems are at a unique position to contribute to this vision through innovations in sound sensing systems, sound feedback design, and sound datasets. I describe my future agenda within each thread below and discuss other applications.

Sound Processing Systems. Developing sound accessibility systems that work in diverse contexts and for multiple user groups is challenging due to varying acoustic conditions and user requirements. My studies have shown that by subtly involving end-users in the model training pipeline, we can build personalized systems that adapt to user-specific requirements and contextual needs [1,12]. 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, before these techniques can be widely adopted, many challenges remain such as *how to collect user data while preserving privacy? how to achieve the right balance between user agency and user effort? and, how to quantify field performance given the subjective user involvement?* As a next step, I would like to explore reinforcement learning techniques that can adapt the model to contextual needs *on-the-go* using minimal user effort. Human-AI sound sensing algorithms have wide applications, including for common technologies such as hearing aids or noise-cancelling earphones.

Sound Feedback Interfaces. Beyond processing sound information, unobtrusively conveying it to the end-users is key to achieving sound accessibility. My work has explored rich design spaces to convert sound to visual and haptic feedback [6,8,9]. However, many questions remain, which I hope to explore through controlled studies of multiple prototypes, such as *how to combine multiple sound information (e.g., sound type, localization, duration) on a small emerging display (e.g., heads-up display, smartwatch)? how to gain a user's trust by effectively conveying the confidence and error in sound processing? how to convey the associated semantics or the intent of a sound (e.g., 'actionable' cues such as a microwave beep or a fire alarm vs. 'experiential' information such as a bird chirp or wind blowing)?*

Sound Datasets. AI offers a great promise to improve accessibility since it can automatically learn from new data, allowing adaptation to a variety of users across environments. However, AI relies on datasets that are rooted in human problems, are unbiased, and are collected from diverse contexts. Huge disparities exist in current sound datasets which are collected by specific populations in limited contexts. I hope to increase dataset diversity by engaging marginalized groups such as people with disabilities or the global south in the data collection process. This is a challenging task. Consider for example, *how would someone who is deaf collect high-quality sounds if they cannot hear them?* To address this, our team is researching visualizations to convey the sample recording quality to DHH users [1]. I plan to expand on this effort through developing customized 'grassroots' physical sensing technologies and data recording interfaces. A key concern is privacy, especially for disabled populations (*e.g.*, deaf users may not know what sound data is collected about them), and I plan to address this through effective data visualizations.

Other Applications of Sound Feedback. Beyond accessibility, unobtrusively sensing and displaying sound signals have applications in many other domains such as home automation, ecological surveys, open surface mapping, and medical imaging, which I also hope to explore. One particular area I am excited about is artistic expressions of sounds around a user for provocation, reflection, and increasing environmental understanding. I envision making physical artifacts that encode sound data sensed either through active sensing (*e.g.*, through microphones embedded on a cup) or through passive collection (*e.g.*, using microphones deployed in a home). Through this work, I plan to reflect on questions such as *how might the artifacts integrate in and influence a user's life? how might personal preferences shape what sound data is collected and how it is visualized? and, how do these implications generalize to modern IoT technologies?* My recent work with personalized, DIY technologies [5,13] make me well suited to this task.

The above threads will benefit immensely from collaborations within (*e.g.*, AI, computing systems) and outside of computer science (*e.g.*, electrical engineering, psychology). I have collaborated with researchers from multiple universities and industries (*e.g.*, Google, Microsoft, Apple), who I will continue to associate with, while developing other partnerships. Together, I hope to advance sound accessibility in this world, which affects all of us.

REFERENCES

1. Steven M Goodman, Ping Liu, Dhruv Jain, Emma J McDonnell, Jon E Froehlich, and Leah Findlater. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *In IMWUT 2021*, 2: 1–23.
2. Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-mounted Display. *In ASSETS 2020*, 1–4.
3. Dhruv Jain. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. *In ASSETS 2014*, 225–232.
4. Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. *In DIS 2018*, 7–11.
5. Dhruv Jain, Audrey Desjardins, *et al*. Autoethnography of a Hard of Hearing Traveler. *In ASSETS 2019*, 236–248. **Best paper honorable mention.**
6. Dhruv Jain, Leah Findlater, Christian Volger, Dmitry Zotkin, Ramani Duraiswami, and Jon Froehlich. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. *In CHI 2015*, 241–250.
7. Dhruv Jain, Rachel Franz, et al. Towards Accessible Conversations in a Mobile Context for People Who are DHH. In ASSETS 2018, 81–92.
8. Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, *et al*. Towards Sound Accessibility in Virtual Reality. *In ICMI 2021*, 1–15.
9. Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, *et al*. A Taxonomy of Sounds in Virtual Reality. *In DIS 2021*, 160–170. **Best paper award.**
10. Dhruv Jain, Kelly Mack, Akli Amrous, Matt Wright, Steven Goodman, Leah Findlater, and Jon E. Frohlich. HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users. *In CHI 2020*, 1–12.
11. Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, and Jon Froehlich. 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.**
12. Dhruv Jain, Khoa Nguyen, Steven Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehilch. ProtoSound: A Personalized, Scalable Sound Recognition System for d/Deaf and Hard of Hearing Users. *In CHI 2022* (to appear).
13. Dhruv Jain, Venkatesh Potluri, and Ather Sharif. Navigating Graduate School with a Disability. *In ASSETS 2020*, 1–11. **Best paper honorable mention.**