

Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing

Dhruv Jain^{1,2}, Leah Findlater³, Jamie Gilkeson⁴, Benjamin Holland⁴,
Ramani Duraiswami², Dmitry Zotkin², Christian Vogler⁵, Jon E. Froehlich²

¹MIT Media Lab
Cambridge, MA

²Computer Science³Information Studies
University of Maryland
College Park, MD

⁴Montgomery Blair High School
Silver Spring, MD

⁵Technology Access Program
Gallaudet University
Washington, DC

djain@mit.edu, Leahkf@umd.edu, rduraisw@umd.edu, wdz@umd.edu, christian.vogler@gallaudet.edu, jonf@umd.edu

ABSTRACT

Persons with hearing loss use visual signals such as gestures and lip movement to interpret speech. While hearing aids and cochlear implants can improve sound *recognition*, they generally do not help the wearer *localize* sound necessary to leverage these visual cues. In this paper, we design and evaluate visualizations for spatially locating sound on a head-mounted display (HMD). To investigate this design space, we developed eight high-level visual sound feedback dimensions. For each dimension, we created 3-12 example visualizations and evaluated these as a design probe with 24 deaf and hard of hearing participants (Study 1). We then implemented a real-time proof-of-concept HMD prototype and solicited feedback from 4 new participants (Study 2). Study 1 findings reaffirm past work on challenges faced by persons with hearing loss in group conversations, provide support for the general idea of sound awareness visualizations on HMDs, and reveal preferences for specific design options. Although preliminary, Study 2 further contextualizes the design probe and uncovers directions for future work.

Author Keywords

Deaf; hard of hearing; head-mounted display; wearable; sound visualization; conversation support; accessibility.

INTRODUCTION

Persons who are deaf or hard of hearing rely on visual signals such as body language, facial expressions, and lip movement to interpret speech (*speechreading*) [4,5,9]. Knowing where to focus visual attention is a prerequisite for effective speechreading. While hearing aids and surgically implanted devices can improve speech recognition, they generally do not improve sound localization [21,23]. In this paper, we investigate visualizations on a head-mounted display (HMD) to increase sound awareness for the deaf and hard of hearing, particularly for group conversations with oral partners.

Prior work on visual aids for persons with hearing loss has focused largely on non-speech sounds (*e.g.*, an alarm or

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-3145-6/15/04...\$15.00
<http://dx.doi.org/10.1145/2702123.2702393>



Figure 1: Our proof-of-concept in use during a group conversation with four people. The wearer (not shown) is seated at the south end of the table. Arrows on the transparent HMD (Google Glass) direct attention toward active speakers. Arrow size indicates loudness. See supplementary video.

doorbell) presented on external displays such as desktops or mobile devices [11,17,18,27]. Though promising, these external displays require the user to turn their attention away from conversational partners. Moreover, these designs require algorithms that can accurately identify non-speech sounds, which is still an open area of research (*e.g.*, [34]). More similar to our work, Kaneko *et al.* [14] proposed a wrist-worn device with LEDs to indicate sound in one of eight directions to support group conversation. Their primary emphasis, however, was on sensing rather than visual feedback, and included only a minimal user study (1 participant with hearing loss). Here, we explore a head-mounted approach, which offers potential advantages such as increased glanceability and privacy compared to past solutions (Figure 1).

To investigate sound awareness visualizations for HMDs, we first developed a taxonomy of eight high-level design dimensions such as *perspective*, *loudness*, and *iconic representations*. For each dimension, we generated 3-12 specific visualizations, which were informed by known sound awareness needs [11,17,18] and our own experiences as persons with hearing loss.¹ We then performed two evaluations: a design probe study (Study 1) with 24 deaf and

¹ Two authors including the first have severe-to-profound hearing loss, which both helps to motivate and inform our work.

hard of hearing participants and a small, exploratory study (Study 2) of a proof-of-concept HMD prototype with four new participants. Study 1 focused on eliciting feedback and identifying promising visual designs, while Study 2 explored the experience of using real-time, HMD sound localization in group conversations.

Findings from Study 1 extend and reaffirm past work [3,9,28] on challenges in group conversations for people with hearing loss who employ speechreading. Specifically, we further motivate the need for glanceable, unobtrusive, and always-available sound awareness to help localize sound sources, focus attention on active speakers, and assist with speaker transitions. During our design probe, all participants supported the idea of head-mounted sound awareness information. We also identified statistically significant preferences for specific design options—for example, positioning sound indicators on the periphery of the screen and how to represent loudness. While preliminary, Study 2’s findings further contextualize the design probe and uncover directions for future work.

In summary, the contributions of this paper include: (i) the introduction of an HMD approach to provide sound awareness for the deaf and hard of hearing, with a delineation of eight relevant design dimensions; (ii) empirical results from a design probe study with 24 lip-reading and signing hard of hearing/deaf participants; (iii) a real-time proof-of-concept system that includes a non-wearable microphone array, and findings from a preliminary evaluation with 4 participants; (iv) design recommendations for HMD sound awareness systems, including visual design and physical form factor aspects. While our long-term goal is to provide sophisticated features such as real-time captioning and speaker identification, our focus here is largely on designing and evaluating ideas for spatially locating sound.

BACKGROUND AND RELATED WORK

We provide background on communication strategies, and cover related work on technology for sound awareness.

Communication Strategies & Sound Awareness

Persons with hearing loss use a variety of strategies to converse with partners who use spoken language, including gestures, two-way note taking, and facial expressions [9]. Researchers have developed classifications for these communication strategies [3,28], which we employ in our analysis. A *maladaptive* behavior detracts or inhibits communication, such as avoiding conversation or feigning comprehension [3]. *Adaptive* behaviors can be verbal or non-verbal, such as asking a speaker to repeat or simplify an utterance, explaining one’s hearing loss, or repositioning to improve one’s view of the speaker [4,9]. Our head-mounted approach should support non-verbal adaptation by helping to direct visual attention.

Even those who cannot hear any sound often understand at least parts of oral conversation through *speechreading*, where lip-reading combined with visual cues, such as facial expressions and body language, are used to interpret speech

[5]. Using a hearing device (*e.g.*, hearing aid or cochlear implant) does not eliminate the need to speechread [*ibid*].

Sound awareness is also important beyond conversations. In interviews with deaf and hard of hearing participants, Mankoff and colleagues asked about awareness needs for *non-speech* sounds [11,18]. Participants emphasized a desire to be more aware of sounds across contexts, including home (*e.g.*, alarms, doorbells/phone), transit (*e.g.*, honking), and work (*e.g.*, activities of coworkers, phones). While valuable, our focus in this paper is *localizing* rather than *identifying* sound. Future work should explore both.

Hearing Aids and Cochlear Implants

Sound localization affects sound awareness and speechreading even for users of hearing devices. Modern devices employ sophisticated digital signal processing that often incorporates noise suppression, dynamic gain control, and directional microphones (*e.g.*, [23,24]). The directional microphones amplify sound in front of the listener while limiting it from the sides and behind [19]; however, they generally do not improve the *localization* of sound without visual stimuli (speaker’s face in view) [23]. For bilateral hearing aid users (both ears), each device works independently, which has destructive effects on the binaural cues necessary to locate sound [1]. Even bilateral cochlear implant users encounter difficulties in localizing sound [8,21]. Multisensory hearing devices have also been proposed, such as to direct amplification based on eye gaze [10]; this still requires the user to know where to look.

Visual and Tactile Sound Awareness Approaches

Prior work has explored visual [11,14,17,18,32] and tactile [6,31,33] sound awareness approaches. The visual solutions focus largely on non-speech sounds and non-wearable solutions (*e.g.*, [11,17,18]). Closest to our work is Kaneko *et al.*’s [14] wrist-worn localization display mentioned in the Introduction. In contrast to our head-mounted approach, however, their device is low resolution and requires the wearer to look away from conversation partners.

Translation of speech information (*e.g.*, acoustic properties) to haptic patterns has also been studied [6,31,33]. For example, Yeung *et al.* [31] created a tactile display that transforms pitch information into a 16-channel vibro-pattern on the forearm. While tactile-based studies have generally shown positive outcomes—*e.g.*, in the perceptual enhancement of words and phonemes among lipreaders [6]—tactile devices remain an active area of research. We do not explore haptic feedback in this paper, but consider it complementary to our visual designs. Finally, real-time captioning is an active area of research. Because automatic speech recognition (ASR) is still far from a solved problem, recent work has looked at real-time editing of ASR [30] or forgoing ASR altogether by using crowdsourcing [16]. Sony recently released *Access Glasses* [26], which presents captioned text on glass lenses for moviegoers with hearing loss; however, the captions are prepared *a priori*. Jones *et al.* [13] examined methods to display a sign language interpreter

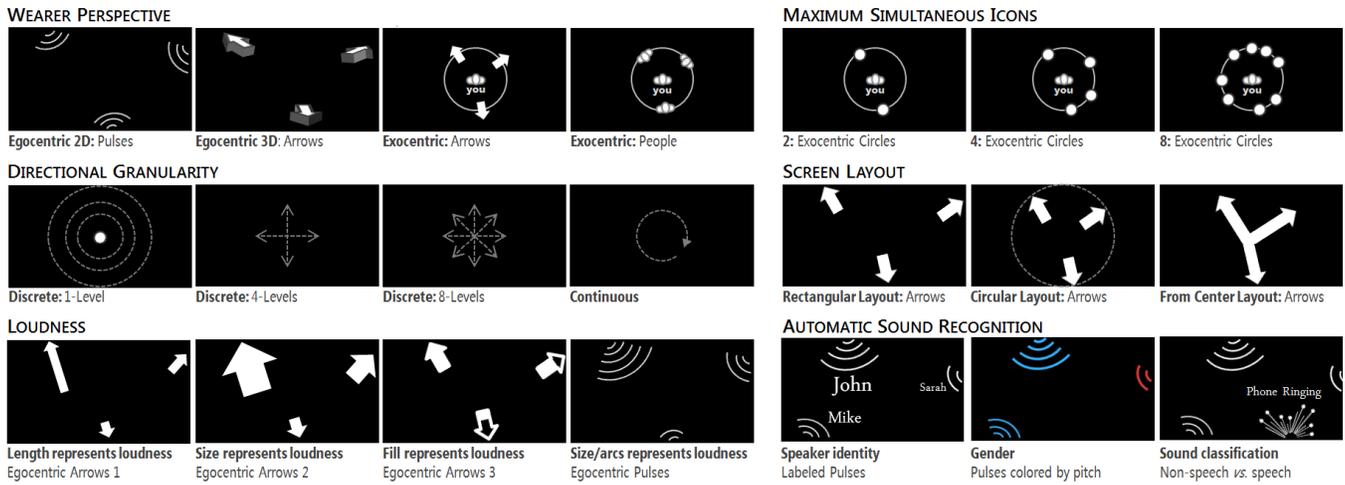


Figure 2: Examples of six of the eight design dimensions: wearer perspective, directional granularity, loudness, maximum simultaneous icons, screen layout, and automatic sound recognition attributes. Different sound indicator icons (*e.g.*, pulse, arrow) are also embedded throughout.

via an HMD. Reactions were mixed—some participants found value in having constant access to the interpreter regardless of head movement. A majority found it difficult to focus on both the interpreter in the HMD and the study tasks (*e.g.*, watching a movie). Though our focus is sound localization, captioning could be incorporated in the future but would also likely demand increased attention from the wearer.

Augmented Reality

Finally, most HMD work has occurred in augmented reality (AR), where virtual imagery overlays physical objects in real time (for reviews, see [15,35]). Though most AR work focuses on augmenting vision, audio has also been explored (*e.g.*, location-based audio cues [22]). Others have investigated how to accurately synthesize spatial audio so that it appears to emerge from AR objects (*e.g.*, [25]). Though AR systems have a potential cognitive cost because they require attentional resources (as with [13]), we could not find design guidelines for displaying peripheral visual cues in HMDs (some work exists for dashboard AR [20]).

DESIGN OF SOUND LOCALIZATION VISUALIZATIONS

We describe our design goals and eight design dimensions.

Design Goals

Informed by the research covered in the Related Work section and our own experiences as persons with hearing loss, we developed the following set of design goals for HMD-based sound visualizations:

- **Localize sound:** The visualizations should provide unobtrusive and accurate indicators of where sound occurs.
- **Glanceable:** The directional information should be easy-to-understand at a glance.
- **Responsive:** The visual cues should render in real-time.
- **Augment, not substitute:** The visualizations should augment and not replace the wearer’s own senses and communication strategies. For example, body language, head turns, and eye movements can often indicate who is speaking; the display should supplement these cues.

- **360° sensing:** The visualizations should provide 360° spatial mappings of information. Study 1 investigates precision levels within this range (*e.g.*, 45° steps vs. 90°).
- **Adaptable:** Informed by [17,18], visual designs should be customizable to fit each user’s needs. Our examination of designs is broad, which helps to isolate specific, promising features that could be customizable in the future.

We also developed a set of secondary goals including: speaker recognition, real-time captions, topic analysis, descriptive conversational statistics (*e.g.*, who speaks the most), and emotion inferences. Though interesting and potentially valuable, most of these secondary aims require sophisticated machine-learning algorithms, which are still active areas of research. Thus, our design work here focuses on sound information that could be gleaned with current technology (*e.g.*, microphone arrays).

Design Dimensions

From the design goals, we generated eight high-level design dimensions (Figure 2). For each dimension, we created at least three different designs by manipulating one or more of the following features: size, shape, perspective, layout, and animation. To simplify our exploration and because Glass has a small color palette [7], we made limited use of color. Similarly, we restricted text use, which may be hard to read at a glance. Below, we describe each design dimension.

1. Wearer Perspective refers to the narrative mode used to render visual information. We evaluated two perspectives: *egocentric*, which presents data from a first-person perspective, and *exocentric*, which presents a disembodied view (*e.g.*, a top-down perspective). For the egocentric designs, we also evaluated 2D vs. 3D views. In the 2D view, indicators at the top of the screen correspond to the wearer’s front while those at the bottom represent behind.

2. Directional Granularity represents how precisely sound is shown on the display. We explored four levels of fidelity: *continuous*, *8-level discrete*, *4-level discrete*, and *1-level discrete*. While continuous shows the exact direction of a sound source, the 8- and 4-level designs visualize data at 45°

and 90° degree discretizations. The 1-level design is binary, showing that sound is occurring but not its location.

3. Loudness represents sound volume. To visualize loudness, we varied *length*, *size*, and/or *percentage fill* (similar to a bar graph). For example, with the *pulse* design, volume was represented by a proportional number of arcs, which were increasingly large for louder sounds.

4. Sound Indicator Icons are the visual shapes used to represent sound. For the egocentric designs, we explored *pulse* (called ‘rings’ in [17,18]), *arrow*, and *finger* icons. For exocentric, we used *pepole*, *arrow*, and *circle* icons.

5. Maximum Simultaneous Icons defines the maximum number of visual indicators to show simultaneously for concurrent sounds (e.g., overlapping speakers). We explored: *two*, *four*, and *eight*.

6. Screen Layout refers to where sound indicators are drawn on the display. We explored three layouts: a *rectangular* layout that positions indicators around the screen perimeter, a *circular* layout that positions indicators around a large, centered circle, and *from center*, which draws indicators in the center of the screen.

7. Conveying Sound Source refers to whether indicator icons should point towards the sound source (i.e., directing outward away from the wearer) or follow the path of the sound itself (i.e., directing inward toward the wearer).

8. Automatic Sound Recognition. Beyond localization, we explored more advanced features, including automatically identifying *who* is speaking, highlighting gender, performing sound classification (e.g., visualizing speech and non-speech differently), and real-time captioning.

STUDY 1: DESIGN PROBE AND EVALUATION

To assess our designs and gain a deeper understanding of problems faced in group conversation with oral partners, we conducted a two-part study with 24 deaf and hard of hearing participants.

Method

Participants

Twenty-four volunteers (12 female) were recruited through email and social media. Participants were on average 38.1 years old ($SD=15.8$, range 23–76). Twenty had profound hearing loss, while the remaining four had at least moderate hearing loss. Most reported congenital hearing loss (12) or loss in early childhood (7). Fourteen participants used a hearing device: 7 reported cochlear implants and 8 used digital hearing aids (1 reported using both). Finally, 19 participants (excluding P3, P7, P9, P19, P22) employed lip-reading during conversation at least some of the time. Participants were compensated \$50 for time and travel.

Design Probe

The design probe included two visual mediums: an iPad PowerPoint slide deck and a custom Android application for Google Glass. The iPad was used to show a breadth of visualizations for each design dimension, which allowed us

to elicit feedback on *overall* ideas rather than specific manifestations of a design. Glass provided a more realistic, medium-fidelity experience and allowed our participants to better understand how the visual designs would work in practice. A subset of designs was shown on Glass for each dimension except for that of maximum simultaneous icons. For example, to compare egocentric 2D vs. 3D designs, the iPad included six examples while only two were shown on Glass. The designs on the iPad were primarily static images similar to Figure 2, while the designs on Glass were pre-rendered animations of two common but difficult group conversation scenarios drawn from the first author’s own experiences: a small group meeting and a classroom setting (Figure 3). Each scenario has three speakers, one of whom is behind the wearer. Glass was controlled by a researcher via a Bluetooth-paired Android phone.

Procedure

The study procedure took on average 67 minutes ($SD=13$), and included a semi-structured formative interview and a design probe. Participants communicated with the research team verbally ($N=9$) or by typing in two-way chat ($N=15$).

Part One: We began with a questionnaire to collect demographics and background on the participant’s hearing loss. The researcher then conducted a semi-structured interview on problems encountered in group conversations, how the participant accommodated those problems, prior experience with computing or mobile devices to support group conversation, and ideas for future technology.

Part Two: At the beginning of the design probe, we introduced the idea of visualizing sound on an HMD, described the meeting table and classroom scenarios in Figure 3 with three animated design examples on the iPad (egocentric pulse, egocentric arrow, and exocentric arrow), and elicited initial feedback. Participants then put on Glass and viewed animated visualizations of the classroom scenario using the exocentric and egocentric arrow designs.

Following this introduction, for each design dimension, we provided a brief textual description, examples on the iPad, and, if applicable, animated visualizations on Glass. We elicited open-ended feedback on the designs and asked for specific preferences along with rationale. To maintain participant engagement and to ensure that they understood the designs, we periodically asked which scenario had been shown on Glass. After completing the first seven design dimensions but before showing examples for automatic sound recognition, we asked participants to describe their ideal designs; participants could do so verbally, textually, and/or with sketches. Finally, the automatic sound recognition examples were shown.

Although not a controlled experiment, to reduce bias we counterbalanced the order in which animated visualizations were shown for each design dimension on Glass. To do so, we created six orderings and assigned an equal number of participants to each ordering. Full counterbalancing was used for sets with 2 or 3 designs, balanced Latin squares for sets of

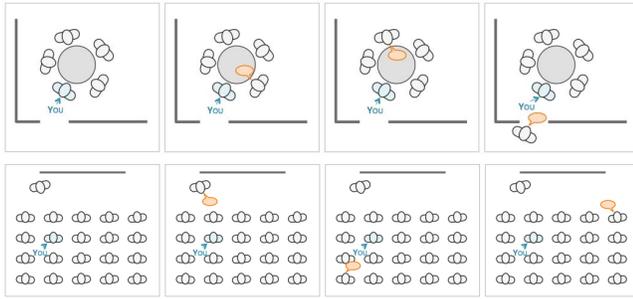


Figure 3: The two scenarios used for animated visualizations in Study 1: meeting table and classroom. Each scenario includes three non-overlapping speakers, one of whom is behind the Glass wearer ('You'). See video.

4 or 6 designs, and no counterbalancing was used for the automatic sound recognition set.

Data and Analysis

Sessions were video and audio recorded. Part One was transcribed for those sessions where participants spoke rather than typed (9 of 24 participants). We then conducted an iterative coding process on these responses [2,12]: (i) one researcher developed an initial codebook for each of 7 sections of the interview; (ii) two independent coders analyzed up to six randomly selected transcripts and met and refined the code set; (iii) the final code set was applied to the remaining transcripts by two independent coders. For this last step, Krippendorff's alpha across all codes was on average 0.68 ($SD=0.30$). Conflicting code assignments were resolved through consensus between the two coders.

For Part Two, we analyzed the distribution of preference votes across options for each design dimension using a chi-square (χ^2) test (except for automatic sound recognition and simultaneous icons, which were open-ended questions). Preference votes were collected verbally and, while we asked participants to make only one selection, they were sometimes unable to decide between multiple options. We note these cases below, for which we divided the participant's vote among their selections; for example, if a participant preferred egocentric and exocentric designs equally, a vote of 0.5 rather than 1.0 was assigned to each. Finally, open-ended justifications for preference were grouped based on emergent themes.

Findings

We first describe participants' experiences and challenges with group communication before discussing findings related to the visual designs. Quotes are pulled from the audio transcripts and two-way chat logs; some are lightly edited to fix typos and grammatical errors.

Part One: Formative Inquiry into Group Communication

Three themes emerged related to group conversation with oral partners: general problems, accommodations, and use of technological solutions.

General Problems. All participants agreed that communicating in a group with hearing persons can be challenging. Over half (15) mentioned problems directly

caused by multiple speakers. Commonly, this included issues following speaker transitions, for example:

"If one person finishes talking, I do not know who to look at next—that is my problem because hearing people can hear who the next person is, and what they are saying." (P20)

Other issues included the difficulty of following overlapping speakers and side conversations, and the need to remind people to look at them when speaking.

Accommodations. Participants were asked how they generally accommodate these problems. Fourteen reported using a variety of traditional adaptations, such as hearing aids, captioning services, or interpreters. For example, P3 discussed needing multiple interpreters for group meetings: *"The first interpreter will interpret. If someone interrupts, the second interpreter takes over and so on"* (P3). Nine participants mentioned using verbal accommodations, including asking a speaker to repeat an utterance or explaining about one's hearing loss. For example: *"I usually ask them to talk one at a time, and only when I am making eye contact,"* (P16). Seven participants reported using low- and/or high-fidelity technology such as pen and paper or two-way chat programs: *"I also use paper and pen to communicate if necessary or use computer technology such as Word and instant messaging."* (P22). Finally, seven participants mentioned maladaptive accommodations [3,28] that would detract from or prevent communication. For example: *"I usually avoid large groups,"* (P16), and *"I almost always interact with Deaf people. When I converse with hearing people it's usually 1:1 with interpreters."* (P4)

Computer Technology. Participants were asked about current and envisioned technology to support group conversations. For current technology, 16 participants mentioned using a smartphone, tablet, or laptop. For example, *"I use my iPhone notepad to communicate,"* (P10), and *"Colleagues have learned to keep their phone handy or use voice-to-text software,"* (P2). Five participants also cited assistive devices such as CART (Communication Access Realtime Translation), frequency modulation (FM) systems that transmit directly to hearing aids, and UbiDuo, a face-to-face communicator. Participants noted that these did not necessarily work well for group conversation, however. Finally, six participants stated that they had never used technology for group conversation: *"Never, there isn't any that I am aware of that would be helpful or that would enable real-time conversational fluidity,"* (P17).

For envisioned technology, 18 participants mentioned enhancements to common assistive technology devices (e.g., ASR on their mobile phone), 6 suggested technologies to indicate the direction of sound, 5 mentioned help for overlapping speakers, and 3 wanted better *in situ* collaborative typing. Automatic captioning was a dominant theme: *"...it would look as if subtitles are appearing over the person speaking like real-time/real-life transcribing,"* (P17). Still, participants realized that captioning is technologically difficult based on prior ASR experiences.

Part One Summary. Our findings confirm research in communication strategies of deaf and hard of hearing people with oral partners (e.g., [3,9,28]). Our participants used a variety of adaptive and maladaptive strategies for group conversation and expressed key challenges that may be addressed by our approach (e.g., missing speaker transitions, helping follow simultaneous speakers).

Part Two: Design Probe Findings

We provide initial reactions to the idea of head-mounted sound localization feedback, responses to each of the eight design dimensions, suggestions for new features, and an analysis of participant-sketched designs.

Initial Reactions. All 24 participants thought the idea of head-mounted visualizations for sound awareness was useful. A majority of participants (17) expressed that it would be helpful to know who is speaking and/or the direction of the sound source. For example, P7, who reported not being a good lip-reader, said that he would like to know where the speaker is so that the interpreter does not have to convey that information, which can waste time. Others emphasized the benefits to speechreading:

“I think it’s a great idea, especially for those that can lip read at least above a functional level... It would reduce the amount of time and effort to find the individual speaking if I have information where the sound is coming from, which would lead to less content loss.” (P17)

However, 4 participants also expressed that the direction of sound source combined with lip-reading would be insufficient to understand conversation and more help would be needed (e.g. interpreters, captions).

Wearer Perspective. Participants were nearly uniformly split in preference between the exocentric and egocentric wearer perspectives, with 13 and 11 votes, respectively. A chi-squared test on the distribution of votes was not significant ($\chi^2_{(1,N=24)} = 0.04, p = ns$). By far the most common reason (12/13) for preferring the exocentric perspective was that it showed the location of the wearer in reference to other speakers, which participants felt made it easier to locate direction compared to egocentric. For example, P15 said: *“I can better judge the direction if I have a [top-down] reference to myself. Pointing to front and back are difficult in egocentric.”* For participants who preferred egocentric, reasons included that it was more understandable, less cluttered, and easier to interpret.

When asked specifically about preference for 2D vs. 3D egocentric designs, participants were again split (12 each). For those who preferred 2D, reasons included simplicity, visual sharpness, and information density (i.e., the same information as 3D but in less space). For those who selected 3D, participants mentioned increased realism and a better sense of direction: *“In 3D, I just follow the arrow while in 2D, I need to remind myself I am in a 3D space,”* (P17).

Direction Granularity. Of the four direction granularities, participants preferred the more precise options

($\chi^2_{(3,N=24)} = 17.75, p < .001$). Fourteen participants selected continuous, the most popular option, because of its precision (e.g., P20 stated: *“The more precise the direction, the better it is.”*). Another perceived advantage, identified by 3 of the 14 participants, was that it could more easily support multiple speakers. The 8-level design was the next most popular (5 votes), with supporters emphasizing the balance between practicality and specificity: *“8 level is easier to locate than 4-level. Continuous is too specific,”* (P1). Only three participants preferred the 4-level and one participant selected the 1-level; these were generally seen as not providing enough information. For example, P20 did not like 1-level because he wanted the display *“to PINPOINT sound exactly where it is coming from.”*

Screen Layout. The rectangular and circular layouts were the most popular, with 10 and 9 votes, respectively, and 3 more participants split between the two; only 2 participants chose the from-center layout ($\chi^2_{(2,N=24)} = 6.81, p = .033$). For participants who liked the rectangular layout, the two most common reasons were that the position on the screen makes it easy to locate sound, and that the aesthetic is uncluttered. However, participants who preferred the circular layout generally found it easier to understand. One participant also mentioned that the circular layout freed up space on the display that could be used for other visual information. P3 captured some of the tradeoffs among the three designs:

“I’d prefer circular. Due to spatial orientation -- I know I’m in the middle. I don’t like rectangular due to the lack of visual aid to assist in spatial orientation. I don’t like from center because those arrows take up too much screen [space].” (P3)

Sound Indicator Icons. Participants were asked about sound icon preference for both an egocentric design and an exocentric design. With the egocentric design, pulses were the most preferred (14 votes) compared to arrows (6) and fingers (2); two participants were split between pulses and arrows ($\chi^2_{(2,N=24)} = 10.75, p = .005$). The most common reason for preferring pulses was that they represent sound in an intuitive way (9), for example: *“...easy to recognize that that’s where the sounds are coming from,”* (P2). For arrows, the most common advantage was that they clearly point to the sound source (3).

With the exocentric design, arrows, people, and circles were equally popular, with 9, 7, and 7 votes, respectively; one participant was split between arrows and people ($\chi^2_{(2,N=24)} = 0.44, p = ns$). The most common reason for preferring arrows, as expressed by 8 participants, was that they explicitly represent the direction of the sound. The people icon was seen as providing a good indication that someone is speaking, as opposed to, for example, a dog barking (of course, such a design would require automatic sound discrimination). Finally, the circles were appreciated for their simplicity. In summary, for the egocentric design, pulses were the most preferred, while for the exocentric design there was no clear preference among icon options.

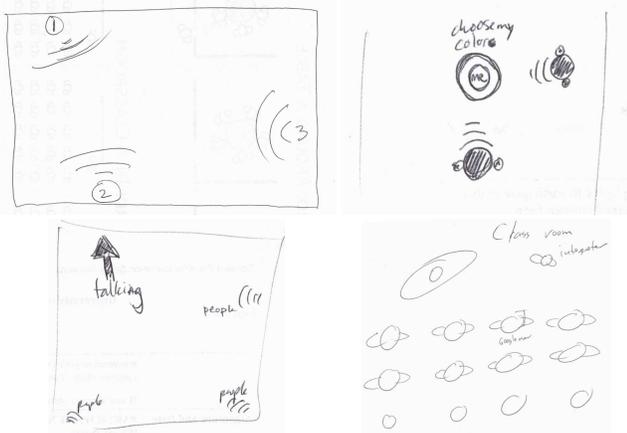


Figure 4: Ideal designs as sketched by our participants. (a) P19 extended egocentric pulses with numbers to show recent speaking order. (b) P14 drew an exocentric design that visualized all potential speakers; pulses indicate who is talking. (c) P14 also created an egocentric design where pulses represent recent speakers and a 3D arrow shows the current speaker. (d) P7 drew an exocentric design with room layout and people locations so that he could focus on his interpreter but still know who was speaking.

Conveying Sound Source. There was a strong preference for indicators that point outward to the sound source (17 votes) compared to indicators that point inward (3); four participants had no preference ($X^2_{(1,N=24)} = 7.04, p = .008$). The primary reason for choosing the outward indicators was that they target the sound source better, for example: “Outward tells me where to look. For inward, I’ll automatically think as if someone is talking to me,” (P6). In fact, 4 participants who chose outward mentioned that for inward, it looked like the speaker was talking to them, which may not be the case.

Simultaneous Indicators. Participants were asked how many overlapping speakers they would want to see visualized simultaneously on the display. This was an open-ended question and, as such, no chi-square test was conducted. The most frequent answer was 4 speakers (10 votes), 8.5 participants suggested 5 or more, and 5.5 wanted at most three speakers. So, while a maximum of four speakers may be reasonable, the variation in preference suggests that this design dimension should be customizable.

Loudness. We assessed how to visualize loudness by showing six options and soliciting overall feedback: egocentric arrows that varied in fill, size, or length, egocentric pulses that varied in size, and exocentric arrows and circles that varied in size. Among these options, the majority of participants (13.5) preferred egocentric pulses that varied in size to convey loudness. Reasons included that it was intuitive and natural. For example: “Egocentric pulses, it just seems natural based on my experiences with WiFi signal strength [icons]” (P17).

A chi-square test on the distribution of votes was significant ($X^2_{(6,N=24)} = 37.98, p < .001$). When asked about other ways to represent loudness beyond what we designed, the most common suggestion (4) was to use color, for example: “red for louder and pink/white very quiet,” (P10). Animation

speed was also suggested by 2 participants: “the pulses should be way faster if the speaker is loudest,” (P20).

Automatic Sound Recognition. Finally, we also asked about four advanced features that would require sophisticated sound processing algorithms: speaker identity, discriminating speech vs. non-speech sounds, real-time captioning, and gender. Almost all participants wanted the first three: speaker identity (22), non-speech sounds (22), and captions (23). For speaker identity, participants thought that it could make locating a person in a large group more efficient, for example: “Speaker name is a great idea; it would eliminate the need to look for moving lips” (P17).

As with the formative findings, real-time captioning was popular. Two new ideas arose about how it could work:

“I have to have interpreters anyway; maybe instead of having them follow me around, they could listen to what Glass hears, and caption it remotely.” (P3)

“If possible by technology, that would be nice. Just display the nouns, others can be implied. Also, real-time captioning is distracting and all consuming.” (P12)

The majority of participants were not interested in highlighting gender (12); three were unsure. It was deemed unnecessary (i.e., one can see by looking) or socially unacceptable. For the 9 participants who indicated interest in gender, they thought it would reduce visual search time: “It will help me to identify who is speaking” (P19).

Other Information. After the last design probe, we asked participants to suggest other information that we did not include in our own designs. Half of the participants wanted more complex cues such as emotion, conversation topics, or information on speakers (e.g., who spoke the most). For example: “I don’t know if Google Glass is able to identify the tone of the person... the emotional tone” (P6). Two participants wanted additional ways of identifying speakers (e.g., assigning colors to more easily discriminate them). One participant also expressed that too much information could lead to overload, so careful selection is required:

“I think it is too much to have names, arrows, circles, and whatever. I think one or the other will suffice.” (P4)

Ideal Design and Rationale. Roughly 70% into the design probe, we asked participants to describe and/or sketch their “ideal design for providing sound awareness feedback” on Glass (Figure 4). We purposely asked this question before showing loudness and automatic sound recognition designs to see if these features would emerge organically. While responses varied, 14 participants described or extended egocentric designs, 7 used exocentric as a base, and 3 participants used one of each. Seven designs included advanced features such as visualizing loudness (4), providing the exact location of people with a floorplan of the room (4), automatically identifying sounds (2), and automatically recognizing when speakers were talking to them (2). When sketching her egocentric 3D design, P20 described her thought process:

“The goal of this for me = pinpoint sound in the most exact, realistic way possible. Adding layouts confuses me. I want it to be simple and ‘point’ to the sound—as if I could ‘see’ the soundwaves coming towards me.” (P20)

Part Two Summary. Though all participants were supportive of HMD sound visualizations, preferences differed across design dimensions. Participants were split between egocentric and exocentric but preferred precise directional granularity, peripheral screen layouts, and pulse (egocentric) or arrow (exocentric) indicators with up to 4 simultaneous speakers. For advanced features, real-time captions, speaker identity, and discriminating speech from non-speech sounds were nearly uniformly desired.

STUDY 2: PROOF-OF-CONCEPT SYSTEM

To collect preliminary feedback on the experience of using working visualizations, we built a proof-of-concept system using Google Glass and a non-wearable microphone array for sound processing (*VisiSonics RealSpace Audio Sensor* [29]). We implemented two popular yet complementary designs from Study 1—egocentric pulses and exocentric arrows—and evaluated these with four new deaf and hard of hearing participants in lab-based group conversations.

System Description

The RealSpace Audio Sensor is a spherical 64-microphone array for localizing sound in three dimensions. It was connected via USB to a laptop running real-time sound source localization and beamforming software [36]. Every 250ms, the laptop sent a processed audio packet consisting of the intensity, azimuth, and elevation of all co-occurring sounds via WiFi to a custom Android application running on a *Samsung Galaxy Chat* phone. The top four highest intensity sounds were then transmitted to Glass via Bluetooth, which generated the feedback visualizations. While useful for a proof-of-concept assessment, this system had two limitations. First, the audio sensor was 20cm in diameter (*i.e.*, non-wearable), so had to be placed in the middle of the meeting table. As a result, the sensing was slightly offset from the participant and did not adjust to head movement. Second, low frequency voices were occasionally missed. See supplementary video for details.

Evaluation Method

Four new participants were recruited (2 female; 26–42 years old, $M=35.0$, $SD=7.5$). Three were profoundly deaf since birth, while R2 had severe hearing loss in both ears since early childhood. R2, R3, and R4 employed lip-reading at least sometimes. Study sessions were one hour long. After a background questionnaire and introduction to HMDs, the participant sat at a meeting table wearing Glass. For each of egocentric pulses and exocentric arrows, the following steps were completed. First, to familiarize the participant with the visualization, a researcher walked once around the room while talking. Second, the participant observed two scripted conversations among research team members: (i) a 2-person conversation with team members seated on opposite sides of the table; (ii) a 4-person conversation with team members

around the table, during which one person moved and spoke behind the participant. Third, all four team members and the participant had an open-ended conversation. All three conversations were 2–5 minutes long. Participants were asked open-ended questions about their experiences.

Order of presentation of the two visualizations was fully counterbalanced and participants were randomly assigned to orders. To reduce the impact of a specific conversation, we excerpted two 2-person scripts from the movie *Shawshank Redemption* and two 4-person scripts from *Ghostbusters*, and defined “travel” and “hometown” as open-ended conversation topics. These options were paired with visualization conditions using full counterbalancing.

Preliminary User Feedback

All participants thought the sound visualizations would be useful at least in some contexts. R1 and R3 felt the display helped with following the scripted conversations, for example, *“I think this [display] is very helpful for when somebody is speaking outside my hearing range, like behind me”* (R3). R2 and R4 did not find it useful for the study tasks, but nevertheless appreciated some aspects of it. For example, R2 commented that the *“approach would be helpful because my sound processor is not able to point where the sound was from.”*

Emergent Issues. Two important issues arose that will need to be addressed in future designs. First, the Glass display is slightly above and to the right of the wearer’s line of sight, which was problematic for R2 and R4. They found it difficult to attend to the display and lip-read, for example: *“I have to look up [at Google Glass] and look at people’s faces. It’s pretty hard to do it at once,”* (R4). A second issue was the potential for cognitive overload, mentioned, again, by R2 and R4. R2 found it distracting to simultaneously attend to his hearing aid and Glass, and was confused when multiple icons appeared at the same time. R4 felt that it could be distracting if small sounds frequently popped up (*e.g.*, paper rustling, coughing), although when asked she still wanted those sounds to be shown.

Contexts. Two participants thought that the system would be useful in a public setting when someone is trying to get their attention. For example, R4 felt it would reduce the chance of inadvertently being rude by ignoring someone she could not hear. R3 found it useful for social group settings, when *“everyone’s participating in the conversation but I don’t know that they are talking.”* However, R4 was not convinced that he would use the device with a group of hearing friends:

“I might not need it because they would want me to understand better by real conversation rather than expecting to read from Google Glass.” (R4)

Preferred designs. Three participants preferred arrows to pulses because they more clearly showed who was speaking or were bolder and easier to see; the fourth participant (R1) had no preference. This feedback highlights the importance of glanceability: *“Because they [the arrows] were bold, you know, I can see them easier without having to [squint]”* (R3).

The sound indicators changed in size based on loudness, which R4 found useful, but R3 said was information her hearing aid already provides.

DISCUSSION

While participants in both studies had varied reactions to specific design dimensions, *all* were supportive of the general idea of providing sound visualizations on an HMD. Here, we reflect on implications and study limitations.

Design Reflections

Based on the exploration of our eight design dimensions, we provide the following design recommendations:

- **Wearer Perspective.** Both egocentric and exocentric designs were well received, so either could be used.
- **Directional Granularity.** Precision is valued. Use high directional granularity, at least 8-level if not continuous.
- **Screen Layout.** Sound indicators should be positioned on the periphery of the screen rather than in the center.
- **Sound Indicator Icons.** For an egocentric design, pulses are recommended. For an exocentric design, no clear pattern emerged, thus arrows, people, or circles could all be used. Other better options may also exist.
- **Conveying sound source.** Arrow-based sound indicators should point outward toward the sound source. For other indicators (*e.g.*, pulses) preference may differ.
- **Simultaneous indicators.** Visualizing up to 4-5 overlapping speakers should satisfy most users, although some users may want to customize this attribute to suit their needs or the environment.
- **Loudness.** Loudness is a desired attribute. Of the six designs shown, egocentric pulses were the most popular.
- **Automatic Sound Recognition.** Of all design dimensions, advanced features were the most uniformly desired, particularly: speaker identity, differentiating between speech and non-speech sounds, and automatic captioning. Gender was not as popular; however, participants did like having additional information on how voices sounded.

Similar to prior studies [17,18], our findings also support the need for customizability. Though strong preferences existed for certain features (*e.g.*, high directional precision) others were mixed (*e.g.*, egocentric *vs.* exocentric). Moreover, preferences may change depending on context (*e.g.*, conversation with friends *vs.* strangers).

The design probe employed relatively simple visualizations for sound localization. How best to combine these with the more advanced features that were requested by participants is an open question. Some participants recognized the potential for information overload from features such as speaker recognition and captioning—as an example, one suggestion was to show only nouns rather than full captions. Although many of these more advanced features are in themselves open areas of research, once they are feasible to incorporate, the ability to turn them on only as needed will likely be key.

As seen in Study 2, Google Glass is not ideal for HMD sound visualizations. It physically interferes with some cochlear implants and behind-the-ear hearing aids, and the display is not always visible in peripheral vision. An ideal HMD

would: accommodate existing hearing devices, include automatic head-tracking that updates visualizations based on head orientation, be lightweight, comfortable, and accurate, and contain a large transparent display superimposed over the eye. Information would need to be presented on the display in a location that does not prevent or distract a person from speechreading. Obviously, a wearable audio sensing solution is also needed, which we did not examine. For accurate sound localization, we predict that at least eight microphones will be needed, which could be positioned on the HMD itself [37]. Our research group is working on a wearable headband version with 16 microphones (extending [37]), which should resolve sound sources with a root-mean-square error of 0.1 radians in the presence of noise.

The visualizations tested were based on a set of design goals informed by past work and our own experiences as persons with hearing loss. The understandability of the localization information, overall positive response from participants, and responsive proof-of-concept system suggest that we were successful in meeting at least some of these goals. However, future work will need to implement and evaluate a completely wearable, interactive system. Properly assessing glanceability and the ability of such a system to augment existing communication strategies, for example, will require observations of conversation dynamics and wearer behavior with realistic use and over a longer time period than our studies allowed.

Limitations

First, most of our study participants had profound hearing loss, but future work should consider a larger, more diverse sample. Second, in the design probe study, we did not show an equal number of all design dimension combinations. Participants had greater exposure to egocentric and outward designs, which may have influenced preferences. Where possible, however, design exposures were counterbalanced (*e.g.*, the order of presenting egocentric *vs.* exocentric designs). Moreover, we solicited preference feedback immediately after first showing a design to mitigate exposure effects. Third, both studies are based on initial reactions in a laboratory rather than long-term use with a working system. Fourth, the audio sensor used in Study 2 had limitations (size, placement, sensing frequency range), which negatively impacted the experience of at least one participant. Future work should apply our findings to a functioning, interactive wearable system that is ultimately deployed and studied longitudinally.

CONCLUSION

As the first work in the area of sound visualizations on HMDs for the deaf and hard of hearing, this paper explored a broad range of novel designs and a preliminary working prototype to examine and uncover promising elements for future work. Study 1, with 24 participants, reaffirmed challenges faced by persons with hearing loss in group conversations and identified design preferences for HMD sound visualizations. Responses highlighted uniform preference for some design dimensions and a need to make

others customizable. We then conducted a preliminary evaluation of a proof-of-concept non-wearable prototype with four participants. Participant feedback supported the potential of our approach, and uncovered additional directions for future work. Our findings have implications for HCI and hearing researchers, hardware developers of HMD technology, and persons with hearing loss.

ACKNOWLEDGMENTS

We thank Ji Hyuk Bae, Yumeng Wang, Meethu Malu, and Alina Goldman for their help.

REFERENCES

1. Van den Bogaert, T., Klases, *et al.* Horizontal localization with bilateral hearing aids: Without is better than with. *J. of the Acoustical Society of America* 119, 1 (2006), 515.
2. Braun, V. and Clarke, V. Using thematic analysis in psychology. *Qual. Research Psych.* 3, 2 (2006), 77–101.
3. Demorest, M. & Erdman, S. Scale Composition and Item Analysis of the Communication Profile for the Hearing Impaired. *J. Sp. Lang & Hear Res* 29, 4 (1986), 515–535.
4. Demorest, M.E. and Erdman, S.A. Development of the Communication Profile for the Hearing Impaired. *Journal of Speech and Hearing Disorders* 52, 2 (1987), 129–143.
5. Gallaudet University. Resources to Develop Speechreading Skills. 2008. <http://goo.gl/HALGdz>.
6. Galvin, K.L., Ginis, J., Cowan, R.S., *et al.* A Comparison of a New Prototype Tickle Talker™ with the Tactaid 7. *Aus. and NZ Journal of Audiology* 23, 1 (2001), 18–36.
7. Google Inc. Google Developers Glass Style. <https://developers.google.com/glass/design/style>.
8. Grieco-Calub, T.M. and Litovsky, R.Y. Spatial acuity in 2-to-3-year-old children with normal acoustic hearing, unilateral cochlear implants, and bilateral cochlear implants. *Ear and hearing* 33, 5 (2012), 561–72.
9. Hallam, R.S. and Corney, R. Conversation tactics in persons with normal hearing and hearing-impairment. *International journal of audiology* 53, 3 (2014), 174–81.
10. Hart, J., Ouceanu, D., Sohn, C., *et al.* The Attentive Hearing Aid: Eye Selection of Auditory Sources for Hearing Impaired Users. *INTERACT 2009*, 19–35.
11. Ho-Ching, F., Mankoff, J., & Landay, J.A. Can You See What I Hear?: The Design and Evaluation of a Peripheral Sound Display for the Deaf. *CHI 2003*, 161–168.
12. Hruschka, D., Schwartz, D., St. John, D., *et al.* Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Fld Methds* 16, 3 (2004), 307–331.
13. Jones, M., Lawler, M.J., Hintz, E., *et al.* Head Mounted Displays and Deaf Children: Facilitating Sign Language in Challenging Learning Environments. *IDC'14*, 317–320.
14. Kaneko, Y., Chung, I., and Suzuki, K. Light-Emitting Device for Supporting Auditory Awareness of Hearing-Impaired People during Group Conversations. *iSystems, Man, and Cybernetics* (2013), 3567–3572.
15. Van Krevelen, D.W.F. and Poelman, R. A Survey of Augmented Reality Technologies, Applications and Limitations. *Intl J. of Virtual Reality* 9, 2 (2010), 1–20.
16. Lasecki, W.S. and Bigham, J.P. Real-time Captioning with the Crowd. *interactions* 21, 3 (2014), 50–55.
17. Matthews, T., Fong, J., Ho-Ching, F., and Mankoff, J. Evaluating non-speech sound visualizations for the deaf. *Behaviour & Information Tech.* 25, 4 (2006), 333–351.
18. Matthews, T., Fong, J., and Mankoff, J. Visualizing non-speech sounds for the deaf. *ASSETS'05*, 52–59.
19. McCreery, R.W., *et al.* An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *American J. of Audiology* 21, 2 (2012), 295–312.
20. Medenica, Z., Kun, A., Paek, T., & Palinko, O. Augmented Reality vs. Street Views: A Driving Simulator Study Comparing Two Emerging Navigation Aids. *MobileHCI'11*, 265–274.
21. Mueller, M.F., Meisenbacher, K., Lai, W., & Dillier, N. Sound localization with bilateral cochlear implants in noise: how much do head movements contribute to localization? *Cochlear Implants Intl* 15, 1 (2014), 36–42.
22. Mynatt, E.D., Back, M., Want, R., Baer, M., and Ellis, J.B. Designing Audio Aura. *CHI'98*, 566–573.
23. Picou, E., Aspell, E., & Ricketts, T. Potential benefits and limitations of three types of directional processing in hearing aids. *Ear and hearing* 35, 3 (2014), 339–52.
24. Picou, E., Ricketts, T., & Hornsby, B. How Hearing Aids, Background Noise, and Visual Cues Influence Objective Listening Effort. *Ear and Hearing* 34, 5 (2013).
25. Sodnik, J., Tomazic, S., Grasset, R., Duenser, A., and Billingham, M. Spatial Sound Localization in an Augmented Reality Environment. *OZCHI'06*, 111–118.
26. Sony. Sony Access Glasses. <http://goo.gl/0DKFoQ>.
27. Tomitsch, M. & Grechenig, T. Design Implications for a Ubiquitous Ambient Sound Display for Deaf. *CVHI '07*
28. Tye-Murray, N., Purdy, S., and Woodworth, G. Reported Use of Communication Strategies by SHHH Members. *J. of Speech Lang & Hearing Res* 35, 3 (1992), 708.
29. Visisonic Inc. VisiSonics RealSpace™ Audio Camera. <http://visisonics.com/realspace-audio-camera>.
30. Wald, M. Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time. *ICCHP'06*, 683–690.
31. Yeung, E., Boothroyd, A., and Redmond, C. A Wearable Multichannel Tactile Display of Voice Fundamental Frequency. *Ear and Hearing* 9, 6 (1988), 342–350.
32. Yoo, I. & Yook, D. Automatic sound recognition for the hearing impaired. *IEEE Trns CE* 54, 4 (08), 2029–2036.
33. Yuan, H., Reed, C.M., and Durlach, N.I. Tactual display of consonant voicing as a supplement to lipreading. *J. of the Acoustical Society of America* 118, 2 (2005), 1003.
34. Zhang, Z. and Schuller, B. Semi-supervised learning helps in sound event classification. *ICASSP'12*, 333–336.
35. Zhou, F., Duh, H.B., & Billingham, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. *ISMAR'08*, 193–202.
36. Zotkin, D.N., *et al.* Chain architecture: An efficient hardware solution for a large microphone array system. *Proceedings of Meetings on Acoustics*, 2013, 3302–3322.
37. Zotkin, D.N., Duraiswami, R., and Gumerov, N.A. Plane-Wave Decomposition of Acoustical Scenes Via Spherical and Cylindrical Microphone Arrays. *IEEE Trans Audio, Speech, & Language Processing on* 18, 1 (2010), 2–16.