Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing

Dhruv Jain[1], Rachel Franz[1], Leah Findlater[1], Jackson Cannon[1], Raja Kushalnagar[2], Jon Froehlich[1]
[1]University of Washington, Seattle, WA [2]Gallaudet University, Washington, DC
djain@cs.uw.edu, franzrac@uw.edu, leahkf@uw.edu, jackscan@cs.uw.edu, raja.kushalnagar@gallaudet.edu, jonf@cs.uw.edu

ABSTRACT

Prior work has explored communication challenges faced by people who are deaf and hard of hearing (DHH) and the potential role of new captioning and support technologies to address these challenges; however, the focus has been on stationary contexts such as group meetings and lectures. In this paper, we present two studies examining the needs of DHH people in moving contexts (*e.g.,* walking) and the potential for mobile captions on head-mounted displays (HMDs) to support those needs. Our formative study with 12 DHH participants identifies social and environmental challenges unique to or exacerbated by moving contexts. Informed by these findings, we introduce and evaluate a proof-of-concept HMD prototype with 10 DHH participants. Results show that, while walking, HMD captions can support communication access and improve attentional balance between the speakers(s) and navigating the environment. We close by describing open questions in the mobile context space and design guidelines for future technology.

Author Keywords

Deaf and hard of hearing; conversations; moving; augmented reality; head-mounted display; real-time captioning.

INTRODUCTION

People who are deaf and hard of hearing (DHH) face several challenges when conversing with partners who use spoken language, such as background noise and lack of access to facial cues [5,10]. Even signed conversations require direct visual contact that may not always be possible. While prior work has investigated how these challenges impact the quality of social interaction and adaptive communication strategies [4,5,10], this work has focused primarily on static contexts (*e.g.,* group meetings, classroom lectures). In this paper, we investigate *mobile* contexts such as walking and transit, which present new potential challenges including varying background noise, changing lighting conditions, and increased visual attention split.

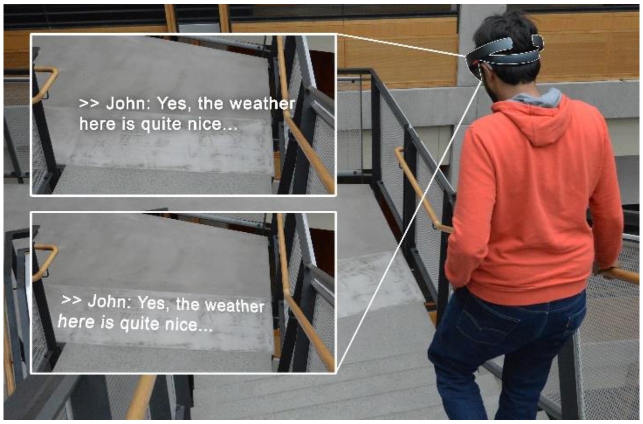

Figure 1: Illustrative image showing our lead author testing two different options for caption placement in our HoloLens prototype: (a) at a fixed distance from the eyes (here, 4m), or (b) automatically projected onto a surface (here, a floor).

Furthermore, accommodation technologies like real-time captioning have also traditionally been designed for and studied in static contexts and, thus, may not be appropriate for moving conversations. Some initial research has explored captioning on smartphones [23,35,38], an approach that requires shifting attentional focus away from the speaker and environment and onto a handheld device. Researchers have also recently proposed head-mounted display (HMD) captioning solutions [13,27,28,30,32], which reduce visual attention split by displaying the captions directly in the user's field of view. While offering potential for captioning in moving contexts, only a few researchers have discussed design, implementation and user evaluation of HMD captions [13,27,30]. Moreover, these evaluations were conducted in stationary contexts. We could not find any prior work exploring HMD captions in a mobile context.

In this paper, we report on findings from two studies investigating the communication needs of DHH people in mobile contexts and the possibility for HMD-based captions to support those needs. Through semi-structured interviews with 12 DHH participants, Study 1 explores communication challenges and coping strategies in moving contexts, preferences for potential real-time captioning solutions (smartwatch, phone, or HMD), and reactions to a captioning mockup on the Microsoft HoloLens. We found that most individuals missed a significant part of spoken conversations while moving and talking. Those who were able to engage in moving conversations divided their attention between looking ahead for personal safety and at the speakers for communication content. Extending findings in [19,21], we found that signed conversations were difficult when visual contact was limited (*e.g.,* hiking on a trail), and when hands were occupied (*e.g.,* driving). For real-time captioning,

ASSETS '18, October 22–24, 2018, Galway, Ireland

ACM ISBN 978-1-4503-5650-3/18/10…$15.00
https://doi.org/10.1145/3234695.3236362

participants preferred the HMD option and provided design suggestions (*e.g.,* ability to customize the captions).

Informed by these findings, prior work [13,30], and our own experiences as persons with hearing loss, we built a proof-of-concept prototype on the HoloLens that displays real-time captions in the user's field of view (Figure 1). In Study 2, we evaluated this prototype with 10 DHH people, who walked with our prototype through a campus building while talking to a research team member and passersby. Participants reported that our prototype assisted their understanding in at least some part of the conversation. However, four participants reported being distracted by the captions, which made it difficult to focus on navigation. We also received design suggestions, such as optimal placement of captions in 3D space and displaying non-speech information.

In summary, the contributions of our work include: (i) qualitative results from a formative study with 12 DHH participants highlighting challenges in mobile conversations, coping strategies, and technology use; (ii) insights from an evaluation of HMD-based caption with 10 DHH participants in a walking scenario; and (iii) design recommendations for future conversation support technologies for mobile contexts, particularly for HMD captioning.

BACKGROUND AND RELATED WORK

We cover background on challenges in spoken conversations for DHH people and related assistive technologies.

Challenges and Communication Strategies

DHH individuals face challenges in spoken conversations due to background noise, lack of facial cues, and lack of attention by hearing people [4,10]. These factors can result in social isolation and reduce conversation participation, which can negatively impact mental health and quality of life [33]. Common strategies to address these communication challenges include the use of gestures, two-way note taking, and turn-taking [10]. Demorest *el al.* [4] classify strategies as *maladaptive* (*i.e.,* that avoid or inhibit conversation) or *adaptive* (*i.e.,* that proactively adapt to the conversation), terms we employ in our analysis. Maladaptive behaviors include ignoring conversation, dominating conversations to avoid listening, and avoiding conversation with strangers [5]. Adaptive behaviors can be verbal or non-verbal [5,10], such as asking to repeat or simplify an utterance, explaining one's hearing loss, and repositioning to improve one's view of the speaker. These strategies may manifest differently in moving contexts, our focus, due to factors such as split visual attention and variable background noise and lighting.

Further, for DHH people to hold effective verbal and signed conversations, architectural spaces should be obstacle-free with balanced light, rounded corners and sound absorbers (see Deaf spaces [15]). These conditions seldom manifest in a mobile context, particularly in outdoor environments.

Speech-to-Text Systems

Our research is informed by prior work in speech-to-text systems that provide spoken information access to DHH people using either trained humans [39], automatic speech recognition (ASR) engines, or a combination of both (*e.g.,* [34]). Because ASR is still an active research area, recent work has looked at real-time editing of ASR-based text [11] or foregoing ASR altogether by using crowdsourcing [20]. Captions by trained human transcriptionists are highly accurate but transcriptionists are expensive and require prior scheduling [18]. Our study includes real-time captions from a trained transcriptionist.

Traditionally, real-time captions are displayed on a laptop or a shared large screen [18]. Two notable transcription systems include *Communication Access Real-Time System* (CART) [29] and *C-Print* [31] which differ in the transcription method, cost, accuracy, and availability. In both systems, captions are appended to a document, so users can scroll up and refer to the text history as needed. Users can also configure the caption color, font, size, and background. Our study employs CART because it is more accurate and widely available than C-Print.

More recently, researchers have begun to explore mobile and wearable captioning solutions. While most smartphone apps rely on a built-in ASR for transcription (*e.g.,* [38]), some systems have used a trained transcriber [23] or a crowd of untrained people [35] to transcribe on the go. We found only one prior work that evaluates phone-based captioning in mobile contexts (*i.e.,* in city, airport, and grocery stores) [23]. Results show that while the app was useful for moving contexts, participants did not appreciate the 3-5 minute transcription delay. Further, though portable, all smartphone apps require that users turn their gaze away from the speaker or environment to use the captions.

To reduce this visual split, some researchers suggest showing captions directly in the users' gaze area using an HMD [13,27,28,30,32]; however, [28,32] did not build or evaluate working systems. Out of the remaining three works [13,27,30], [13] and [30] are most relevant and report on caption design, implementation, and user evaluation on the HoloLens. In our previous work [13], we presented a 14-dimension design space for HMD captioning, one dimension of which was how caption design may change based on the *physical activity of speakers*. Our preliminary evaluation, however, only explored two design dimensions with a single participant in stationary contexts. We extend that work here.

Peng *et al.* [30] explored options for visual caption design on a HoloLens such as changing the number of lines, speaking order, and speaker location. They prototyped the most preferred design and solicited feedback from eight DHH participants. Results show that participants could more easily identify speakers and follow conversation with the HMD compared to the laptop captions. Though promising, their evaluation context was a stationary meeting held in a lab.

As an alternative to captions, Jones *et al.* [16] and Miller *et al.* [27] explored displaying a sign language interpreter via an HMD. In both cases, at least some participants found value in having the interpreter always in their view. However, while in [27] participants found it easier to follow a lecture with an HMD, in [16], a majority found it overwhelming to focus on both interpreter and other study tasks (*e.g.*, watching a movie).

In summary, while prior work suggests some potential advantages of HMD-based captioning (*e.g.*, reduced visual split, increased dialogue), this work has not been evaluated in mobile contexts—a gap which we address in our work.

Sound Awareness Technologies

Besides captioning, past work has explored general sound awareness for DHH people using visual [2,12,17,24–26] and tactile approaches [8,36,37]. Most visual solutions are based on non-portable devices such as desktops [12,24,25], but researchers have also explored mobile and wearable sound awareness solutions. For example, Bragg *et al.* [2] and Mielke *et al.* [26] used smartphones to recognize and display environmental sounds (*e.g.*, phone ringing, sirens). Gorman [9] and Kaneko *et al.* [17] displayed the direction of sound sources using a wrist-worn device and HMD respectively. Jain *et al.* [14] used a design probe method to explore HMD visualizations for different sound characteristics. This lattermost work highlighted the importance of supporting sound localization and suggested that speaker tone, conversation topics, and contextual information (*e.g.*, who spoke the most) may be valuable to explore. Participants in our studies also identified the need for additional sound cues such as speaker tone and environmental sounds.

Besides visual feedback, tactile approaches have been employed to provide speech information such as voice tone [37] or frequency [36]. Researchers have also tried methods to completely substitute hearing with tactile sensation (*e.g.*, [8]), but this has shown little promise. We do not explore tactile feedback in this paper but consider it complementary.

STUDY 1: ASSESSING NEEDS IN MOBILE CONTEXTS

To assess the needs and potential technologies for DHH people in mobile contexts, we conducted a formative study.

Participants

Twelve volunteers (five males, six females, and one not disclosed) were recruited through email, social media and snowball sampling (Table 1). Participants were on average 34.5 years old (*SD*=15.3, range 18–66). Eight had profound hearing loss, while the remaining four had at least mild hearing loss. Most reported onset as congenital (*N*=9). Ten participants used a hearing device: two used cochlear implants, seven used digital hearing aids, and one reported using both. Nine participants (excluding P5, P6, and P12) employed speech-reading at least some of the time.

For communication, eight participants preferred sign language, and four preferred to speak verbally. Three participants reported that when conversing verbally with hearing people, they understood more than 81% of the speech; three understood 61-80%; four understood 41-60%; and the remaining two could barely understand speech (<20%). Participants received $25 as compensation.

Table 1. Demographics of participants in Study 1 (P1, etc.) and Study 2 (R1, etc.), covering age, gender, hearing loss, lip-reading, preferred communication method, and percentage of speech understood in verbal conversations. "ND" means Not Disclosed.

ID	Age	Gender	Hearing Loss	Lipreads?	Prefers	%speech
P1, R1	23	F	Profound	Yes	Sign	41-60%
P2, R5	18	ND	Profound	Yes	Verbal	> 81%
P3, R3	24	F	Profound	Yes	Oral	41-60%
P4	55	M	Severe	Yes	Sign	41-60%
P5	32	F	Profound	No	Sign	< 20%
P6	21	F	Mild	No	Oral	>81%
P7	28	M	Profound	Yes	Sign	41-60%
P8	35	F	Profound	Yes	Sign	61-80%
P9	66	M	Profound	Yes	Sign	> 81%
P10, R4	32	M	Severe	Yes	Oral	61-80%
P11, R2	23	F	Moderate	Yes	Sign	61-80%
P12, R6	57	M	Profound	No	Sign	< 20%
R7	28	F	Profound	No	Sign	< 20%
R8	31	F	Profound	Yes	Sign	41-60%
R9	28	F	Profound	Yes	Sign	41-60%
R10	54	F	Profound	Yes	Sign	21-40%

Method

The study procedure included a two-part, semi-structured formative interview and took about one hour. We investigated challenges and strategies for conversation in moving contexts, ideas for future technologies, and, for those showing preference for HMDs, a brief captioning technology mockup on HoloLens. Participants communicated with the research team by typing in a shared Google Doc. When desired, sign language was also used for minor clarifications or small talk.

Part 1. The session began with a questionnaire to collect demographics and background on the participant's hearing loss. The researcher then conducted a semi-structured interview on the frequency, location, and social context of mobile conversations, problems encountered during these conversations, how the participants handled those problems, and the impact of physical space (*e.g.*, architectural layout). Three mobile scenarios were explicitly explored: walking, public or personal transport, and other recreational activities such as sports, hiking, or kayaking.

Part 2. We presented three real-time captioning ideas for mobile contexts (phone, smartwatch and HMD; see Figure 2) and asked for the participants' preferences with rationale. If a participant preferred the HMD for at least one scenario, we introduced the experience of HMD-based captions using a physical mockup on HoloLens. Because we wanted feedback on the general concept of HMD captioning, we asked participants to imagine a future lightweight version of the HMD for this exercise. Our HoloLens mockup presented a single line of scrolling text from an example script at a fixed distance from the wearer. Participants briefly wore the HoloLens and explored our mockup while walking around the room and turning their heads at different angles. We asked for their thoughts about the overall concept and to describe or draw any other design ideas to support mobile conversations.

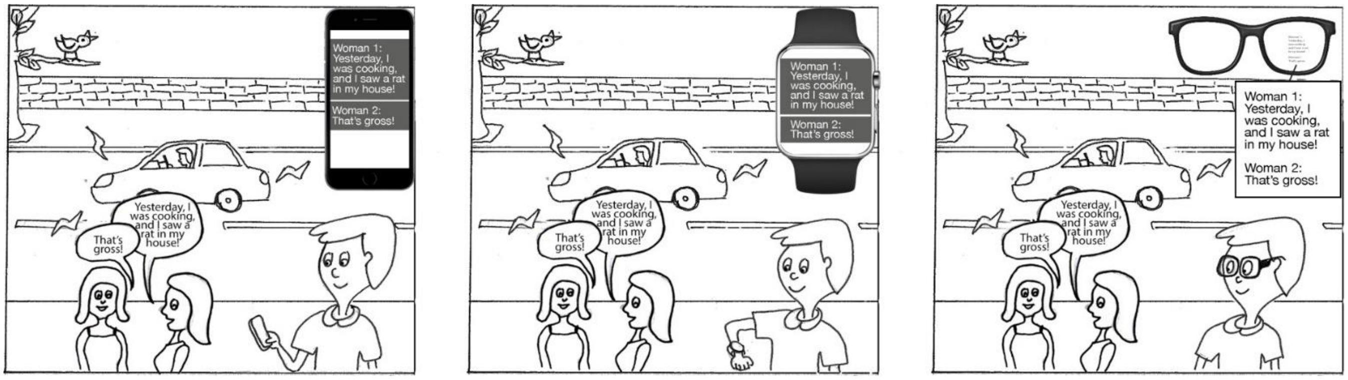

Figure 2: Images shown to the participants in Study 1, to guide them in choosing their preference among three potential real-time captioning devices: phone, smartwatch and HMD.

Data Analysis

We thematically analyzed [3] the interview transcripts that had been recorded in the shared Google Docs. One researcher first scanned the transcripts and identified 962 excerpts to be coded across all participants (M=80.2 excerpts/person, SD=19.0). Next, the first researcher and a second one defined three deductive codes based on prior research [4,5] (*adaptive strategies*, *maladaptive strategies*, *form factor comparison*). Both researchers then categorized the excerpts using the deductive codes and an affinity diagraming process to allow for inductive themes [1]. Through this process, we identified six inductive codes (*social*, *environmental*, and *personal challenges*, *technology use during moving conversations*, *characteristics of moving conversations*, and *user requirements*). Finally, a third researcher reviewed the codes assigned by the first and second researcher, agreeing in 93.6% of cases. Disagreements were resolved through consensus.

Findings

We report on findings related to moving conversations, technology use, social and environmental challenges, maladaptive and adaptive communication strategies, comparison of HMD, phone and smartwatch for captions, and design suggestions for future technologies.

Characteristics of Moving Conversations

Participants mentioned having conversations (spoken or via sign language) while walking to or from meetings, classes, and social activities (N=11), as well as on public transport (N=11), in cars (N=9), and during other recreational activities, such as hiking, playing pool, and kayaking (N=6). All participants had engaged in moving conversations in the preceding week, although the typical frequency varied from only once (P7) to 20 times (P12) a week (M=8.1, SD=5.1).

For conversational partners, most participants reported having moving conversations with both deaf and hearing classmates (N=11), followed by coworkers (N=6), family (N=4), and strangers (N=4). Further, P2, P3, and P4 stated that they never engage in mobile conversations with their hearing bosses or professors since, as P4 explains: *"Those conversations [with boss] are critical. We [sit and] always face each other when speaking."*

Technology Use in Moving Conversations

To communicate with hearing people during moving conversations, eight participants reported using a messaging or note taking app on their smartphone. Most (N=7) resort to using an app only when it is too difficult to hear the other person (N=5) or when the environment is noisy (N=2):

> *"I don't usually use technology other than hearing aids in moving conversations. I will occasionally use my phone to type something if it's impossible to hear."* (P10)

Participants use these apps on public transport (N=7), when riding in cars (N=4), and while walking (N=2). For example, *"BIG [a note taking app] makes it easier to communicate in the dark while driving"* (P6). Five participants reported that they rarely use technology in moving conversations because as P12 explained, *"I don't usually write [type on phone] and move at the same time because it is too challenging"*.

No participants reported using a real-time captioning system during moving conversations, but three participants had used transcription services while moving, including *Microsoft Translator* (P12), *Google Speech to Text* (P4), and voicemail transcriptions (to read rather than listen to voicemail, P8). P12, for example, discussed using a speech-to-text app:

> *"The Microsoft translator [speech-to-text app on phone] makes it easy to have a walking conversation because I can hold my phone in whatever orientation I need to see my surroundings [...]. [Further, using the group chat option] each person can talk/read on their own phone to have the conversation."* (P12)

P12 also explained some problems with this app:

> *"The MS Translator isn't perfect because it demands that I split my attention and [also] have one [hand] holding the phone."*

Social Challenges

Our second key finding relates to social challenges experienced during moving conversations.

Moving conversations are superficial. Participants explained that moving conversations lack social connection and convey limited information compared to stationary conversations (N=9). Because of split attentional demands, moving conversations were described as being brief and shallow, consisting mainly of small talk (P3, P7), comments and jokes (P4), questions and answers (P11, P1), and "*syncing up information*" (P8). Consequently, the potential for social connection is compromised. For example, P1 reported feeling socially isolated, "*It's hard to make the [walking] conversation smooth enough to go deep… I feel like left out all the time*." Further, four participants mentioned that it is difficult to convey and receive complete information from a moving conversation. For example,

> *"[While driving,] often we are relying on smartphones, GPS, transportation, and many other distractions. [So,] conversations will go on strange tangents because my focus is divided."* (P8)
>
> *"It's really hard to walk and talk and lip read... The experience overall often can be negative because 90% of the time you're unsure if you conveyed/understood the conversation well."* (P7)

Contexts and communication method. Participants faced challenges with speechreading (*N*=5) or signing (*N*=3) based on the context. For example, because they have to focus on looking ahead, P1 and P10 face difficulty with speechreading in the car when they are driving, and P1 and P7 could not speechread while hiking or biking. Signed conversations were difficult mainly during driving because of the need to keep hands on the wheel (P5, P8, P12). In addition, P5 and P12 explained that understanding signs requires the listener's complete visual attention, which is difficult to provide as a driver, especially when the passenger is not fluent in signing.

Lack of participation from others. Seven participants said that hearing people do not understand and accommodate communication needs in moving conversations. Group conversations with hearing people are especially difficult since they do not stop talking when the deaf person needs to look away (*N*=5). As P10 explains,

> "*If I need to look away for some reason, a deaf person will automatically stop talking and resume when I'm ready. A spoken conversation doesn't have that type of natural stop and start…"*

Challenges during recreational activities. Recreational activities can be especially challenging (*N*=4) since they often require instruction and feedback during complex movements (*e.g.,* martial arts, yoga), greater physical exertion (*e.g.,* running, soccer), or chaotic environments (*e.g.,* white-water rafting). For example, P11 is sometimes unable to hear her yoga instructor calling out the next move and P12 reported missing instructions because he could not see his coach's signs during wrestling. These two participants also explained another challenge: their instructors found it difficult to demonstrate a movement and concurrently explain through signing. Pl2 said:

> *"[In] martial arts: you have an instructor showing how to move the arms, hands, body, etc. while talking to describe it. Well if they have to "talk" by signing, then how the [heck] do they also show you how to hold your arms in the proper position?"*

Environmental Challenges.

Participants also noted challenges related to the environment such as building layout, noise, and variable lighting.

Visual attention split. Seven participants felt that moving conversations were more challenging than stationary ones due to the need to divide attention between the conversation and environment (*e.g.,* cars, people, traffic lights, obstacles). This attention split is especially challenging on uneven pavement or hiking trails (*N*=7). For example, P5 said: *"When the sidewalk is bumpy, I have to focus more on looking ahead."* Nine participants were concerned that attention split during conversations posed a safety issue, as P8 explains, "*Safety and communication often compete for my attention in a walking conversation*." Four participants were particularly concerned about safety while driving.

Table 2. Participants' preference (if any) among three devices for captions (HMD, phone, watch) in different moving contexts.

	Walking	Transit (bus, car)	Recreational
HMD	All but P5	P1,P2,P4,P8,P9,P12	P2,P4,P7,P9,P12
Phone	P5	P3,P5,P10,P11	P3,P5,P10,P11
Watch		P6	

Space concerns. The design and composition of indoor space affected moving conversations. Large spaces and/or spaces with predictable layouts made it easier to move and face the other person (*e.g.,* grocery stores or museums, as mentioned by P8). In contrast, narrow spaces that required the speaker to be in front or behind the participant reduced understandability (*N*=3). Conversations in cars were also difficult when the speaker was not facing the participant (*N*=7). Another concern was the distance from the speaker; three participants reported not being able to hear speakers who were far away. There was also a social element to interpersonal distance, such as when P6 commented on the difficulty of conversing with strangers: "*it might be awkward to stand closer to them to hear their voice*."

Background noise. Moving conversations were impacted by background noise from announcements, bus engines, and other passengers' movement in public transit (*N*=4); rain and wind while walking and hiking (*N*=3); and other traffic while driving (*N*=2). Comparing walking and buses, P2 said:

> *"I do find conversations on buses easier than walking on a busy street. While there is background noise, it is consistent."*

P1 and P2 both found it easier to converse at night because *"environments tend to be quieter and less busy"* (P1).

Lighting concerns. In contrast to P1 and P2's comments above, nine participants felt it was easier to have moving conversations during the day when they could more easily see people's faces to lip read or see sign language. Indoor lighting could also negatively affect speechreading when it was too bright (P5, P11) or too dark (P8, P11).

Accommodation Strategies

To address the challenges noted above, participants employed a variety of adaptive and maladaptive strategies.

Adaptive. Verbal adaptive strategies included asking the speaker to repeat (*N*=3), repeating what was heard back to the speaker (P2), asking the other person to speak louder (P3), explaining that they are hard of hearing (P6), and controlling the flow and length of conversations (*N*=2). The most common non-verbal adaptive strategies were adjusting one's seating position to be next to or across from the speaker on public and personal transport (*N*=7), walking side-by-side (*N*=4), turning to face the speaker (*N*=3), using mirrors to see speakers in the backseat of the car (*N*=2), and choosing a quiet path to the destination (*N*=2). P9 and P10 mentioned that they would sometimes stop moving to aid comprehension and participation:

> "*I stop everybody walking and resume the conversation until a point has been addressed, and then start walking again*" (P10).

Maladaptive. Maladaptive strategies included avoiding group conversations (P6) or talking to strangers (*N*=6), postponing a conversation (*N*=6), avoiding spoken conversations altogether (*N*=6), and briefly pausing a conversation (*N*=2). For example, P3 does not talk to Uber or Lyft drivers to avoid "*spending the energy on* it." P11 and P12 chose parts of a conversation to pay attention to rather than the whole conversation. Some participants (*N*=7) avoid conversations based on context. For example, P12 prefers to focus on the scenery while hiking than on conversations, P10 defers important conversations when walking, and P8 lets a phone call go to voicemail when walking.

Envisioning a Real Time Captioning System

We next discuss the social implications of a captioning system, and compare phone, smartwatch and HMD devices.

Social implications. After viewing the mockup in Figure 2, all participants said they would use real-time captioning in at least one moving conversation scenario (walking, transit or other recreational activity). However, seven participants wanted to employ captioning selectively because of how the captions may affect conversation quality. As P11 explains,

> "*I always prefer direct communication with hearing people. If technology or interpreters are involved, there is always a distance between me and the other person. It diminishes the quality of the human connection*."

Four participants were concerned that their communication practices and skills would change because of a captioning system. P1 wondered if a system would change the use of sign language, while P9 worried about a potential loss of communication skills by relying too much on a system. P8 wondered if *"a system would change my lifelong pre-disposition to look at people's faces when they talk*." P10 preferred to choose when to follow a conversation, which may be difficult if he had to *"[look] at captions all the time."*

Comparing devices. Eleven participants (except P5) preferred the HMD in at least one moving context (walking, transit, or recreational) (Table 2). The perceived main advantage of the HMD was that it would reduce attention split by positioning captions within the user's gaze (*N*=6). Four participants also felt that the HMD could improve both the quality of a moving conversation and social connections:

> *"[The HMD] would help me be more "connected" to people [compared to smartwatch or smartphone] since I can look at the people [while reading captions]." (P11)*

P9 wanted to use the HMD to overhear group conversations to increase a sense of social inclusion:

> "*I want to be unobtrusive, whether I'm part of the conversation or eavesdropping on the conversation (just like hearing people)."*

For high-contact sports, four participants preferred to use smartphone since they were concerned that the HMD "*would be knocked off easily"* (P11). These four also preferred to use smartphone rather than an HMD on a public or personal transportation because of the option to sit and focus on the display, a common behavior in these contexts. Only one participant preferred a smartwatch (for transit). The most common perceived disadvantage was that the small watch display would only accommodate a limited number of words (*N*=8), which would affect readability (*N*=5) and the ability to follow a conversation (*N*=3).

Participants also mentioned disadvantages of the HMD (*N*=4) and phone (*N*=7). Three participants were concerned that the HMD would be too visually distracting and overwhelming. P6 explained that the HMD might give her motion sickness, a recurring condition for her. Concerns for the phone included display size (*N*=5), needing to look down at one's hand (*N*=5) and the effortful process of holding up one's hand, especially during long conversations (*N*=4).

HoloLens for captioning. While we used the HoloLens to help ground discussion about HMD-based captioning and asked participants to envision a more futuristic, streamlined device, participants who tried HoloLens (*N*=11) expressed concerns about heaviness and bulkiness (*N*=8), comfort (*N*=6), and how it may hinder mobility (*N*=2). Six participants mentioned social acceptability concerns due to the large form-factor; all suggested using glasses or contact lenses instead to display captions. All eleven participants, however, appreciated HoloLens as a good prototyping tool for evaluating future HMD captioning devices: *"I know future devices would be smaller and [would] fit on my face better, but HoloLens works good for testing"* (P11).

Design suggestions for HMD captioning. When asked about future improvements for HMD captions, participants wanted the ability to go back through a conversation history if they missed something (*N*=3), adjust settings and caption attributes, such as position, contrast, and font (*N*=2), and convey additional information, including: who was speaking (*N*=3), the position or distance of the speaker (*N*=3), and noises in the environment (*e.g.*, door opening) (*N*=2). Five participants wanted captions to remain on top of the speaker regardless of the user's head movements since the caption movement was *"dizzy[ing]"* (P6) or *"distracting"* (P11).

Technology design sketches. When describing their ideas for future captioning technology, four participants also sketched designs. Three of them (P2, P4, P9) extended conventional glasses to include a small screen for displaying

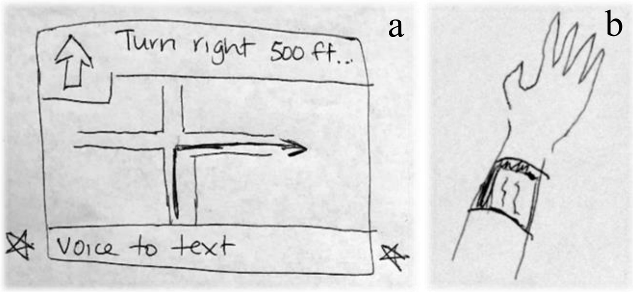

Figure 3: Future technology design sketches by participants. (a) P5 proposed integrating speech-to-text system with the car GPS to reduce information overload from multiple devices. (b) P2 proposed a wrist-worn screen to display captions. They said this would be more readable than a smartwatch and more portable than a smartphone or HMD.

captions to avoid wearing the heavy HoloLens. P5 sketched an integrated GPS and voice-to-text system for a car (Figure 3a); she was concerned about the technology burden posed on deaf people and wanted to avoid wearing a new technology while driving. P2 sketched an integrated display for contact lenses and proposed a wrist-worn display (Figure 3b). Finally, to share the technology burden with hearing people, P9 sketched a t-shirt design that displayed captions that could be read if the speaker is wearing that t-shirt.

PROOF-OF-CONCEPT ON MICROSOFT HOLOLENS

Informed by Study 1 findings, prior work [13,30], and our own experiences as persons with hearing loss, we built a proof-of-concept prototype to display real-time captions in 3D space on the Microsoft HoloLens (Figure 1). To inform the design of our prototype, we had two aims: (1) increase conversation accessibility and (2) reduce the visual attention split between the environment and captions. For accessibility, we chose to place the captions in the user's gaze at all times—even when the wearer is not facing the speaker(s). To reduce visual split, the captions were automatically projected onto surfaces (*e.g.,* walls, floor) using the HoloLens' environment mapping capability if desired by the user.

Users could configure the number of lines, length of each line, the font size, and the distance of captions from the eyes (2m, 4m, 8m, or projected onto surfaces; see Figure 1). To reduce jitter by head movement, captions stayed at the same location until the user's gaze exceeded 25 degrees. We used Streamtext [40], a remote online captioning software, to receive real-time captions from an on-site professional transcriptionist; however, future versions could incorporate ASR engines. Captions were rendered in white, Arial font.

To prepare for Study 2, our hard-of hearing lead author evaluated our prototype in a campus building via three walking sessions (*avg.* 44mins each). While moving, the author interacted with a total of eight people and evaluated different caption configurations. The lead author could understand speakers better using our prototype; however, wearing the heavy HoloLens device for long periods proved tiring. Thus, we decided to limit Study 2 sessions to 20mins. For the captions, our author preferred two lines of text, 60 characters per line, an angular font size of approximately 0.75 degree, and surface projection. These configurations were Study 2 defaults but users could adjust them if desired.

STUDY 2: EVALUATION OF PROOF-OF-CONCEPT

To validate our Study 1 findings and gain further insights into HMD-based captioning in mobile contexts, we evaluated our prototype in a real-life walking scenario. Our primary goals were to assess whether the use of HMD-based captions increased conversation accessibility and decreased attention split for walking conversations. We chose walking because walking and public transport were the most common moving conversation scenarios mentioned in Study 1 but walking requires more consistent visual attention.

Participants

We recruited 10 participants, including six participants who preferred the HMD idea from Study 1 and four new participants recruited through snowball sampling (Table 1). The four new participants (R7 to R10) were on average 35.2 years old (*SD*=10.9, range 28–54). All were female with profound hearing loss. R7 and R8 developed hearing loss at 2 years of age while R9 and R10 had congenital hearing loss. R9 and R10 used a hearing device (hearing aid). All four participants preferred sign language for communication, and three participants (except R7) employed speech-reading. While conversing verbally with hearing people, R7 could barely understand speech (<20%), R8 and R9 could understand 41-60%, and R10 21-40%. Finally, of all 10 participants, three (R3, R5, R8) used real-time captioning in daily life and the remaining used sign-language interpreters. Participants received $25 as compensation.

Method

The study procedure took on average 57 minutes (*SD*=10.3) and included a walking scenario (*M*=24 mins, *SD*=5.4) and an open-ended interview (*M*=30 mins, *SD*=12.4). The study began with participants briefly viewing our HoloLens prototype. A researcher adjusted the font size, number of lines of text, and the length of each line if the participants desired. After a quick test, the participants walked in a university building wearing our prototype. Two researchers accompanied the participant. Researcher 1 walked alongside the participant and initiated a conversation on casual topics, such as family, weather, food, or events in the city. Researcher 1 also wore a wireless microphone, which relayed the conversation to a remote professional transcriptionist to generate real-time captions. Researcher 2 observed the interaction and recorded notes. Participants could interact with other passersby if they so desired.

After the trial, we conducted an open-ended interview about the experience and solicited feedback about the prototype. Both researchers also asked follow-up questions based on Researcher 2's observations. For this interview component, participants communicated with the researchers by typing in a Google Doc (*N*=6) or verbally (*N*=4). Similar to Study 1, we also used the captions from a transcriptionist and sign language to facilitate communication. Finally, participants completed a short multiple-choice questionnaire asking about how much they depended on captions during the trial and how their speech understanding in daily life compared to that when using our prototype.

Data Analysis

We retained the professional transcripts and used them to conduct an iterative coding process on the interview responses [3]. One researcher scanned the responses, developed an initial codebook, then iteratively applied the codes to the transcripts, updating the codebook as necessary. The codes were applied to transcripts as a whole and not individual excerpts. The final codebook contained eight codes (*e.g.*, *caption placement*, *impact of the environment*). A second researcher then independently assigned codes to each transcript using the final codebook. Krippendorff's alpha across all codes was on average 0.65 (*SD*=0.24). Conflicting assignments were resolved through consensus.

Findings

We first describe overall reaction to the prototype, followed by specific themes that arose.

Overall reaction. When asked about the overall experience, all participants mentioned using our prototype to understand at least some part of the conversation while walking. For example, R6, who cannot comprehend oral speech, said:

> *"Being deaf, I can't have that [walking] conversation with a person without some assistance. I need another person who can sign, or another captioning device (Like MS Translator on the phone). Both of these would entail looking away from the speaker. So HoloLens might be better in that regard. [...] With the phone translation, I also need to dedicate a hand to holding the phone."*

On a scale of 1 (mostly unintelligible) to 5 (mostly intelligible), we asked participants to rate how well they typically understand their everyday walking conversations (sign language or oral). Participants reported an average of 2.4 (*SD*=1.17). When asked how well they understood conversation during the study session with the HMD, participants reported an average of 3.8 (*SD*=0.79) A pairwise t-test was significant; $t(18) = 3.13$, p=.006.

Four participants (R2, R4, R6, R7) appreciated the ability to follow the conversations while looking ahead (Figure 4a). R4 explained, *"The big thing with this is you can look where you want and still follow along with the conversation."* However, R3, R4, R5 and R8 found captions to be occasionally distracting. For example, R5 said:

> *"When I was trying to formulate my own responses, I would find the captions quite distracting and, in cases like that, I wish [...] that I could look away from [the captions], at my discretion."*

Caption preferences. Most participants (except R3, R5) preferred the default configuration for the captions (0.75 degree font, two lines, and 60 characters per line). R1 said:

> *"Two lines were good... I can look away from the captions for a little bit and then something that somebody said will still be there when I look back."*

Figure 4: Images illustrating findings from Study 2. (a) R7 followed the conversation while looking ahead. (b) R8 held the railing to guide her movement; it was difficult for her to walk on stairs wearing the HoloLens. (c) R4, who prefers to converse orally, looked at the speaker instead of captions.

R3 and R5 increased angular font size to 1 degree, and R5 used three text lines. Regardless of preference, all participants appreciated the customizability of the captions.

Apart from captions, participants wanted the HMD to display speaker identification cues (*e.g.*, name, location) (*N*=5) and environmental sounds (*N*=3). R6 wanted an indicator that someone was talking based on which he could *"decide to listen or not"*. He also thought it would be useful to display voice tone and volume:

> *"Tone of voice (and volume) is a big one, for example indicating volume by font size and tone by font. You could also have something else visually indicating it (e.g. sometimes tv/movies put a little music symbol up for music)."*

To reduce information overload, R9 proposed that we:

> *"include everything [additional sound cues] and maybe have options for people to filter out because I can imagine some people going crazy with all the information overload."*

Visual attention split. To understand the conversation, all participants used both speechreading and captions. The post-trial questionnaire showed average dependence on captions of 3.2 (*SD*=1.03) on a scale of 1 ("*I did not look at captions*") to 5 ("*I only looked at captions*").

Participants who prefer to converse orally (R3, R4 and R5; Table 1) reported looking at speakers more than captions (average dependence on captions: 2.3; *SD*=0.58; Figure 4c). They used captions only to fill in missed parts of the speech or confirm their understanding of the speech. For example, while involved in a group conversation with three passersby, R3 and R4 missed speech during speaker transition and used captions. The remaining participants, who preferred sign language for communication, focused on captions more than speakers (average dependence on captions: 3.6; *SD*=0.98). For example, *"[I was] mostly focused on captions [and did] not really [look at] faces"* (R6).

When not actively engaged in a conversation, participants alternated between looking at their surroundings and captions. R4, R6 and R8 responded that they focused more on their environment, while R2 focused more on captions to see if somebody was speaking. For the other six participants, we could not ascertain a clear indication of attention split between captions and environment from their responses. But based on the post-trial questionnaire, all six indicated that they looked at their surroundings much less than in a typical moving conversation in their everyday lives.

Caption placement. Six participants (all but R4, R5, R7, R10) appreciated our idea of captions staying in the user's field of view. For example,

> *"I liked that the captions are still going even when I am not looking at you in case you are talking to me when I am not looking."* (R2)

However, five of those six participants wanted to be able to turn-off captions. For example, R3 noted:

> *"[When] I'm cooking something and sometimes I don't want to see captions, so turn[ing] [captions] on and off would be a good option... it would give you a break from seeing them all the time."*

In contrast, R9 commented:

> *"[I] would just leave it on all the time because I can easily ignore [the captions] if I don't want to pay attention."*

R3 also wanted the option to *"move captions down here [below the display] or on the side [periphery]"* to focus more on the surroundings. R6 wanted to move the captions closer to the speakers in his field of view: *"I can see [the benefit of] moving the captions a few degrees towards the speaker."*

The remaining four participants (R4, R5, R7, R10) wanted the captions to be positioned above speakers (like speech bubbles) so the only way they could see captions is if they were looking at speakers. R4 also mentioned another advantage of speech bubbles:

> *"It would allow for the captions to be like a whole paragraph, so somebody could speak, I could look away, but they would have like a whole backlog of things, so I could follow along."*

However, when asked how they would notice speakers outside their field of view (*e.g.,* behind them), R4 and R7 wanted to be able to *"toggle between the two"* caption positioning options, *i.e.,* moving with the user's head or speech bubbles (R4).

Impact of the environment. Six participants found it harder to walk on stairs since they had to split their visual attention between the captions and the stairs (Figure 4b). For example,

> *"When I walked on stairs I needed to look at the stairs. I had to also pay attention on captions. So I was a little nervous that I might step wrong."* (R1)

R2, however, reported that after an initial period, *"I was more aware of how [the HoloLens] works, so I adapted a little bit."* Additionally, five participants who walked in broad daylight had trouble looking at the captions. For example, R3 said: *"the HoloLens display is not bright enough to accommodate for natural lighting."*

HoloLens device. As in Study 1, participants reported that the HoloLens is heavy (N=7), has a limited field of view (N=4), is not fully transparent (N=3), and draws attention (N=3). Three additional insights related to the device emerged. First, as mentioned, our captions moved only when the angular deviation of the user's head exceeded 25 degrees. R4 found this stabilizing technique *"laggy,"* but R5 liked it and described it as a *"cool stabilizing technique."* Second, as a person walks, the HoloLens device learns and adjusts to the new environment, which made the captions appear jittery for a moment. R5, noticing this, said: *"it [captions] kind of flickered a little bit when I was staring at architecture that [are] like rounded columns."* Finally, R5 commented that the automatic depth adjustment of captions in 3D space was a *"really cool feature."* However, R4, R7 and R8 wanted the captions placed at a fixed distance *"coz otherwise my eyes would have to constantly adjust [to different depths]"* (R7).

DISCUSSION

We now discuss our findings in the context of past work, open questions for future exploration, design implications for future technology, and limitations of our studies.

The Mobile Context – A New Space

Prior work has explored communication challenges for DHH people, such as background noise [5], inability to hear a speaker without visual contact [10], and visual attentional demands in signing conversations [19]. Our Study 1 findings indicate that these challenges also manifest in moving contexts though with greater severity. Mobile contexts impose greater attentional demands than stationary contexts, due to greater visual attention split and context changes, which include topographical, spatial, and noise variation. We also uncovered new challenges; mobile conversations for DHH people are generally brief and shallow, greatly affected by environmental and spatial characteristics, and have limited technology support and use.

Of particular significance is the context of recreational activities such as yoga, dance, or wrestling, which require that individuals receive time-critical information. As P12 explained, verbal instruction for physical movements are most effectively conveyed simultaneously with a demonstration of the movement. If instruction and demonstration are sequential rather than parallel, the meaning is *"diluted."* For non-signing users who are focused on the activity, paying attention to vocal information provided by instructors is difficult. For signing users, their hands might be otherwise occupied.

Due to attentional and mobility demands, mobile context technologies need to be carefully designed. Captioning technologies, in particular, need to be portable, should adapt to changing contexts, and potentially employ automatic speech recognition. As is mentioned in past work [23,35] and also corroborated in our findings, phones and smartwatches are not typically preferred for mobile contexts since these devices demand split visual attention, a dedicated hand to carry, or are too small for displaying captions. HMDs have the potential to reduce this attention split but need to be lightweight, comfortable, and unobtrusive for broad acceptance. As our participants suggest, other form-factors like glasses and contacts could be leveraged for captions in the future. Displaying captions on non-wearable artifacts, such as in-car GPS systems (P5), would further reduce visual dispersion and the need to carry a personal device.

Design Implications for HMD Captioning

Based on our findings, we propose the following design recommendations for HMD-based captions, which can be investigated and validated in future work:

Text alignment. Captions should automatically align close to the speaker or background to reduce the visual attention split between captions and the environment.

Adapt to changing context. Caption color and background should automatically change based on lighting conditions.

Wearer's voice. The HMD should have an option to disable the wearer's voice to minimize information on the screen.

Contextual information. Besides captions, HMDs should convey information such as speaker name and location, speech tone, and environmental cues (*e.g.*, door opening).

User customizability. HMDs should allow customization of caption position, contrast, font, and background. Prior studies also support the need for customizability [24,25].

Future Work

Our initial findings show that HMD-based captions can support communication access in mobile contexts. While the use of HMD-based captions seem to improve the attentional balance between the speaker(s) and navigating the environment, future work should explore this in depth. One potential solution may be to limit the amount of real-time text by displaying keywords or a summary of text; however, extracting this text automatically remains a difficult challenge. Another potential solution may be to position captions directly above speakers and use visualizations to identify speakers outside the field of view (*e.g.*, see [14]). As in [13], users should be able to control the placement of captions and to temporarily or permanently silence them as necessary.

Future work should also explore the role of tactile feedback as a complementary information channel. While tactile information is lower bandwidth than visual information, tactile information often imposes a lower attentional demand [6] and may be useful for some tasks (*e.g.*, notifications, indicating direction of sound).

Finally, some assistive technology efforts have been criticized in the deaf community as "manifestations of audist beliefs" (*e.g.*, [7]), where the technology burden is imposed on deaf people to accommodate hearing communication standards. We acknowledge this important concern. Indeed, an example from our data that attempts to counteract this imbalance is P9's idea that all conversation participants (deaf and hearing) should display real-time captions on their shirts. Captions are predominantly visual, and therefore potentially provide functional equivalence to deaf and hearing users who are comfortable reading captioned broadcast and online programs daily. We emphasize that our studies and technology designs were informed by our own experiences as DHH individuals, our previous work with DHH participants and wearable sound awareness technologies [13,14], and perspectives drawn from the literature [22,30]; however, we are a team composed of technologists. We also emphasize the diversity among deaf viewers–some prefer captioned videos, and others prefer signed videos. Future work should continue to engage with the DHH community to ensure that we are asking the right questions and pursing appropriate solutions.

Study Limitations

Our work has four primary limitations. First, our findings on attention split relied on self-report. Future work should conduct a comparative gaze tracking study with and without HMD captions to more accurately determine how users' visual attention shifts in moving conversations. Second, in Study 2, we used lapel microphones and a professional transcriptionist for the real-time captions. Future work should explore tradeoffs in transcription quality, lag, and the impact on the conversational experience in mobile contexts that may come with automated captions. Third, Study 2 evaluated walking for a short period in a single building. Future work should consider longitudinal deployments in a variety of contexts. Finally, we largely relied on a live two-way Google Doc to communicate with participants during interviews (similar to [14]); however, some participants may have been more fluent in sign language than English (we did not collect data on this).

CONCLUSION

In this paper, we presented two studies examining the needs of DHH people in moving contexts and the potential for mobile captions to support those needs (*e.g.*, rendered via a watch, phone, or HMD). Our formative study with 12 DHH participants (Study 1) identified social (*e.g.*, limited social connection) and environmental challenges (*e.g.*, limited visual contact in narrow spaces) unique to or exacerbated by moving contexts. All but one participant preferred an HMD for mobile captions. Informed by these findings, we designed a proof-of-concept HMD prototype on the Microsoft HoloLens, which was evaluated in a semi-controlled study with 10 DHH participants (6 from Study 1, 4 new participants). Our findings demonstrate the promise of always-available captions rendered on an HMD and also help identify important areas for future work such as the incorporation of non-speech sounds (*e.g.*, speaker location, environmental sounds).

ACKNOWLEDGMENT

We thank Shuxu Tian, Bonnie Chinh, Liang He, James Fogarty, Jessie Schroeder, Lee Stearns, and Lisa Hutchinson. This work was supported by a Google Faculty Research Award, a University of Washington Reality Lab grant, and a HoloLens equipment donation from Microsoft.

REFERENCES

1. Hugh Beyer and Karen Holtzblatt. 1997. *Contextual design: defining customer-centered systems*. Elsevier.
2. Danielle Bragg, Nicholas Huynh, and Richard E. Ladner. 2016. A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '16*, 3–13.
3. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2: 77–101.
4. Marilyn E. Demorest and Sue Ann Erdman. 1986. Scale Composition and Item Analysis of the Communication Profile for the Hearing Impaired. *Journal of Speech Language and Hearing Research* 29, 4: 515–535.
5. Marilyn E. Demorest and Sue Ann Erdman. 1987. Development of the Communication Profile for the Hearing Impaired. *Journal of Speech and Hearing Disorders* 52, 2: 129–143.
6. Johan Engström, Nina Åberg, Emma Johansson, and Jakob Hammarbäck. 2005. Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems.
7. Michael Erard. 2017. Why Sign-Language Gloves Don't Help Deaf People. *The Atlantic*. https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/
8. Karyn L. Galvin, Jan Ginis, Robert S. Cowan, Peter J. Blamey, and Graeme M. Clark. 2001. A Comparison of a New Prototype Tickle Talker™ with the Tactaid 7. *Australian and New Zealand Journal of Audiology* 23, 1: 18–36.
9. Benjamin M Gorman. 2014. VisAural:: a wearable sound-localisation device for people with impaired hearing. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, 337–338.
10. Richard S Hallam and Roslyn Corney. 2014. Conversation tactics in persons with normal hearing and hearing-impairment. *International journal of audiology* 53, 3: 174–81.
11. Rebecca Perkins Harrington and Gregg C Vanderheiden. 2013. Crowd caption correction (ccc). In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 45.
12. F Wai-ling Ho-Ching, Jennifer Mankoff, and James A Landay. 2003. Can You See What I Hear?: The Design and Evaluation of a Peripheral Sound Display for the Deaf. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), 161–168.
13. 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 *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (DIS'18 Companion), 4–7.
14. 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 *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 241–250.
15. Charlene A Johnson. 2010. Articulation of Deaf and Hearing Spaces Using Deaf Space Design Guidelines: A Community Based Participatory Research with the Albuquerque Sign Language Academy.
16. Michael Jones, M Jeannette Lawler, Eric Hintz, Nathan Bench, Fred Mangrubang, and Mallory Trullender. 2014. Head Mounted Displays and Deaf Children: Facilitating Sign Language in Challenging Learning Environments. In *Proceedings of the 2014 Conference on Interaction Design and Children* (IDC '14), 317–320.
17. Y Kaneko, Inho Chung, and K Suzuki. 2013. Light-Emitting Device for Supporting Auditory Awareness of Hearing-Impaired People during Group Conversations. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, 3567–3572.
18. Raja. S Kushalnagar, Walter S Lasecki, and Jeffrey P Bigham. 2014. Accessibility Evaluation of Classroom Captions. *ACM Transactions on Accessible Computing* 5, 3: 1–24.
19. Jim G Kyle and Bencie Woll. 1988. *Sign language: The study of deaf people and their language*. Cambridge University Press.
20. Walter S Lasecki, Christopher D Miller, Raja S Kushalnagar, and Jeffrey P Bigham. 2013. Legion Scribe: Real-Time Captioning by the Non-Experts. In *10th International Cross-Disclipinary Conference on Web Accessibility (W4A)*.
21. Susan M Letourneau and Teresa V Mitchell. 2011. Gaze patterns during identity and emotion judgments in hearing adults and deaf users of American Sign

Language. *Perception* 40, 5: 563–75.

22. Tara Matthews. 2006. Designing and Evaluating Glanceable Peripheral Displays. In *Proceedings of the 6th Conference on Designing Interactive Systems* (DIS '06), 343–345.

23. Tara Matthews, Scott Carter, Carol Pai, Janette Fong, and Jennifer Mankoff. 2006. Scribe4Me: Evaluating a Mobile Sound Transcription Tool for the Deaf. In *Proceedings of Ubiquitous Computing (UbiComp)*, Paul Dourish and Adrian Friday (eds.). Springer Berlin Heidelberg, 159–176.

24. Tara Matthews, Janette Fong, F. Wai-Ling Ho-Ching, and Jennifer Mankoff. 2006. Evaluating non-speech sound visualizations for the deaf. *Behaviour & Information Technology* 25, 4: 333–351.

25. Tara Matthews, Janette Fong, and Jennifer Mankoff. 2005. Visualizing non-speech sounds for the deaf. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and Accessibility - ASSETS '05*, 52–59.

26. Matthias Mielke and Rainer Brueck. 2015. Design and evaluation of a smartphone application for non-speech sound awareness for people with hearing loss. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 5008–5011.

27. Ashley Miller, Joan Malasig, Brenda Castro, Vicki L Hanson, Hugo Nicolau, and Alessandra Brandão. 2017. The Use of Smart Glasses for Lecture Comprehension by Deaf and Hard of Hearing Students. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17), 1909–1915.

28. Mohammad Reza Mirzaei, Seyed Ghorshi, and Mohammad Mortazavi. 2012. Using Augmented Reality and Automatic Speech Recognition Techniques to Help Deaf and Hard of Hearing People. In *Proceedings of the 2012 Virtual Reality International Conference* (VRIC '12), 5:1–5:4.

29. National Association of the Deaf (NAD). Communication Access Realtime Translation. https://www.nad.org/resources/technology/captioning-for-access/communication-access-realtime-translation/

30. Yi-Hao Peng, Ming-Wei Hsu, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '18).*

31. National Technical Institute for the Deaf Rochester Institute of Technology. The C-Print® System. https://www.rit.edu/ntid/cprint/

32. Kazuki Suemitsu, Keiichi Zempo, Koichi Mizutani, and Naoto Wakatsuki. 2015. Caption support system for complementary dialogical information using see-through head mounted display. In *Consumer Electronics (GCCE), 2015 IEEE 4th Global Conference on*, 368–371.

33. Kristian Tambs. 2004. Moderate effects of hearing loss on mental health and subjective well-being: results from the Nord-Trøndelag Hearing Loss Study. *Psychosomatic medicine* 66, 5: 776–782.

34. Mike Wald. 2006. Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time. In *Proceedings of the 10th International Conference on Computers Helping People with Special Needs* (ICCHP'06), 683–690.

35. Samuel White. 2010. Audiowiz: Nearly Real-time Audio Transcriptions. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '10), 307–308.

36. Eddy Yeung, Arthur Boothroyd, and Cecil Redmond. 1988. A Wearable Multichannel Tactile Display of Voice Fundamental Frequency. *Ear and Hearing* 9, 6: 342–350.

37. Hanfeng Yuan, Charlotte M. Reed, and Nathaniel I. Durlach. 2005. Tactual display of consonant voicing as a supplement to lipreading. *The Journal of the Acoustical Society of America* 118, 2: 1003.

38. TextHear Speech To Text Technologies for the Hearing Impaired. https://texthear.com/

39. What is real-time captioning? | UW DO-IT. https://www.washington.edu/doit/what-real-time-captioning

40. StreamText.Net. http://www.streamtext.net/