



# “Easier or Harder, Depending on Who the Hearing Person Is”: Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status

Emma J. McDonnell  
University of Washington, USA  
ejm249@uw.edu

Soo Hyun Moon  
University of Washington, USA  
smoon88@uw.edu

Lucy Jiang  
Cornell University, USA  
lucjia@cs.cornell.edu

Steven M. Goodman  
University of Washington, USA  
smgoodmn@uw.edu

Raja Kushalnagar  
Gallaudet University, USA  
raja.kushalnagar@gallaudet.edu

Jon E. Froehlich  
University of Washington, USA  
jonf@cs.washington.edu

Leah Findlater  
University of Washington, USA  
leahkf@uw.edu

## ABSTRACT

With improvements in automated speech recognition and increased use of videoconferencing, real-time captioning has changed significantly. This shift toward broadly available but less accurate captioning invites exploration of the role hearing conversation partners play in shaping the accessibility of a conversation to d/Deaf and hard of hearing (DHH) captioning users. While recent work has explored DHH individuals’ videoconferencing experiences with captioning, we focus on established groups’ current practices and priorities for future tools to support more accessible online conversations. Our study consists of three codesign sessions, conducted with four groups (17 participants total, 10 DHH, 7 hearing). We found that established groups crafted social accessibility norms that met their relational contexts. We also identify promising directions for future captioning design, including the need to standardize speaker identification and customization, opportunities to provide behavioral feedback during a conversation, and ways that videoconferencing platforms could enable groups to set and share norms.

## CCS CONCEPTS

• **Human-centered computing** → Accessibility; Empirical studies in accessibility.

## KEYWORDS

Accessibility, d/Deaf and hard of hearing, Videoconferencing, Captioning

### ACM Reference Format:

Emma J. McDonnell, Soo Hyun Moon, Lucy Jiang, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2023. “Easier or Harder, Depending on Who the Hearing Person Is”: Codesigning Videoconferencing



This work is licensed under a Creative Commons Attribution-Share Alike International 4.0 License.

CHI '23, April 23–28, 2023, Hamburg, Germany  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9421-5/23/04.  
<https://doi.org/10.1145/3544548.3580809>

Tools for Small Groups with Mixed Hearing Status. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 15 pages.  
<https://doi.org/10.1145/3544548.3580809>

## 1 INTRODUCTION

For many d/Deaf and hard of hearing (DHH) people real-time captioning is an essential communication access tool. With advances in automatic speech recognition (ASR) [65] and the rise of videoconferencing following COVID-19 [3], platforms such as Zoom, Google Meet, and Microsoft Teams [38, 66, 67] now provide on-demand but imperfect captions. This new landscape introduces captioning options in informal contexts where CART<sup>1</sup> is typically unavailable, but ASR also provides less complete access [65], making the impact of group dynamics on conversation accessibility particularly relevant. Researchers have begun to explore how these group dynamics impact captioning experiences, identifying ways that hearing people’s adaptive or unsupportive behavior can shape conversation accessibility [43], that hearing people tend to speak in ways that run counter to DHH people’s captioning preferences [52–54], that the use of captioning needs to be socially structured and negotiated [29], and that, over time, DHH/hearing dyads can co-create accessible practices [61]. Furthermore, following the Deaf community’s long-standing argument for access approaches that decenter hearing norms [2, 13, 36] and disability justice activism’s framing of collective access [57]—wherein accessibility is a group rather than individual responsibility—we look to opportunities for captioning technology design that engages DHH and hearing people alike.

For videoconferencing specifically, prior work shows these platforms are not designed with DHH communicators in mind [50, 60] and adapting group social norms can be an effective but difficult-to-maintain approach to improve accessibility [35, 39, 55]. Certain aspects of videoconferencing (e.g., captions available on all conversants’ screens) may also be well-suited to engaging groups in taking on access work [43]. However, prior work has not examined

<sup>1</sup>CART, or Communication Access Real-Time Transcription, is a service wherein trained human transcribers provide captioning in real time

how small mixed-hearing groups<sup>2</sup>—common in workplace and educational contexts—negotiate and use captioning features together nor how they could be better supported. Working with established mixed-hearing groups, in particular, can provide insight into emergent social accessibility practices co-developed together over time and draw on their lived experiences and collective communication strategies to inform future caption tool design. Therefore, our work seeks to address:

1. How do established mixed DHH and hearing groups think about, interact with, and react to captions during online conversations?
2. When engaging in the codesign of future online captioning systems, what features do mixed-hearing ability groups desire, how would they design them, and why?

Over three codesign sessions, we explored how groups currently communicate while using captioning, brainstormed features that could be added to videoconferencing environments to better support accessible group communication, had participants individually sketch their ideas for those features, then shared and discussed participants' highest priority feature ideas. Researchers made video prototypes of each group's top three ideas, and groups both reviewed their ideas in depth and commented on other groups' feature ideas.

Our findings focus on groups' experiences communicating together and their proposed designs for future online captioning supports. Extending prior work that engaged only individual DHH people or dyads [42, 55, 61], we highlight the complex factors that shape individuals' current use of captioning (e.g., variable use depending on DHH people's reliance on audio), how participants' established relationships dictate their communication practices (e.g., relying on familial history or setting explicit norms at work) and considerations that go into group access norms (e.g., a hearing person leading norm enforcement to set an example for others). We then report on the codesigned features our participants developed throughout our study process, finding a continued, under-addressed need for captioning basics in videoconferencing platforms (e.g., speaker identification and customization), interest in features to give meeting participants feedback on factors that impact access (e.g., slow down warning, flagging confusing captions), opportunities to build sound recognition into captioning tools, and the desire to build access norms into videoconferencing infrastructure. Participants envisioned these features in the context of their social use, highlighting the need to center conversational dynamics when assessing the impact of captioning tools, which prior work has often assessed in controlled experiments or in terms of specific metrics (e.g., [5, 19, 33]).

In summary, this work contributes: (1) an empirical account of how mixed-hearing groups approach captioned conversations, (2) a participant-determined set of priority features for future accessible group communication supports, and (3) guiding principles for designing for behavior change in groups with mixed hearing abilities.

<sup>2</sup>By mixed hearing groups, we mean any group with both hearing and DHH members

## 2 RELATED WORK

We situate our work relative to disability and Deaf community approaches to accessibility, communication access for DHH people, the design and use of captioning tools, social dynamics and captioning, and videoconferencing accessibility.

### 2.1 Deaf and Disability Community Grounding

Our focus on captioning design for group use is rooted in Deaf and disability studies and activism. We take up disability justice activists' call to shift toward collective access, or the idea that "*we can share responsibility for our access needs*" [57]. Central to collective access is the idea that Deaf and disabled people should not be independently responsible for arranging access but that groups should interdependently make their interactions accessible [44]. Interdependence is the idea that everyone relies upon each other and that dependence is not a unique facet of disability [45, 46]. As interdependence is integrated into HCI accessibility research (e.g., [4, 39, 40]), this thinking creates opportunities to envision access as for communities, rather than solely individual disabled people. Additionally, Deaf scholarship and cultural politics emphasize an approach to communication that prioritizes Deaf, rather than hearing norms [2, 13, 36]. Often these approaches center signed conversation, but we adopt the assumption that hearing styles of communication do not have to be the standard for the design and use of captioning tools.

### 2.2 Communication Access for DHH People

DHH people use a range of tools and strategies for communication access, including sign language, speechreading, writing, gesture, hearing aids, cochlear implants, human-generated captioning (e.g., CART), and automatic captioning [64]. While these strategies are often used in concert, our paper focuses on real-time captioning, particularly automatic captioning technology and its increasing use during videoconferences. Despite years of development, recent evaluations found that, on high quality audio sources, popular automatic speech recognition (ASR) engines reach 88-95% accuracy [65] and range from 81-86% accuracy in less ideal conditions [49]. Moreover, assessments often use samples from hearing speakers and ASR performs markedly worse on people who speak with Deaf accents [16]. In contrast, CART writers must be able to caption speech at 180 words per minute with 96% accuracy [68]. While videoconferencing is an increasingly viable use case for ASR, its relatively high error rate and lack of human transcription judgment mean DHH people using automatic captioning for access face continued hurdles to comprehension.

### 2.3 The Design and Use of Captioning Tools

Captioning and its potential augmentations have been well-studied in HCI literature, with particular focus on the impact of speaker behavior, preferred displays, and ways to provide more contextual information.

Speaker behavior can impact captioning effectiveness. A foundational study of speech rate and captioning found that 145 words per minute (wpm) is an optimal, comfortable caption reading speed and that audience comprehension decreases above 170 wpm [27]. Further, communication issues can arise from overlapping speech

[15, 24], not seeing other speakers [24], ASR not understanding some speakers' accents [15, 30], and background noise [15, 47].

Prior work has investigated how DHH people prefer captions to be displayed. Despite many proposed alternatives, DHH viewers consistently prefer familiar captioning styles, such as standards used for TV and movies [5]. Captioning viewers also appreciate options to customize caption styles [18]. Further, DHH captioning users need to have captioning in the same visual field as other relevant information (e.g., lecture slides, speakers' faces, interpreters) [10, 11, 32–34, 62]. Strategies to minimize visual dispersion include collating distributed information to one interface [11], projecting captions to follow the speaker [33], and displaying captions underneath lecture slides [10]. However, Amin et al. [1] highlight that captions also must not occlude relevant information, such as faces.

Because captioning is an inherently incomplete representation of a conversation, research has also explored ways to convey more information to captioning viewers. As ASR-based tools become ubiquitous, researchers have explored many ways to visualize algorithmic confidence and potential errors. Yet, DHH people consistently prefer captions with no markup, valuing non-distracting captions over error information [6, 56]. Rather than annotating captions, Harrington and Vanderheiden proposed a tool for meeting attendees to correct errors [22]. Others have explored caption formatting to convey prosodic information. ASR usability increases when pauses in speech are represented [19] and text is better punctuated [59]. DHH research participants had mixed reactions to conveying tone [37] and volume [23] through caption design but reacted positively to showing speaker identification through dynamic caption placement [23, 31, 37]. In addition to speech, prior studies have found that DHH people desire sound information to contextualize spoken conversations [24]. Yet, while implementing sound recognition for DHH people is an active research area (e.g., [7, 17, 26]), there has been limited focus on combining sound recognition with captioning [20, 69].

## 2.4 Social Dynamics and Captioning

Prior work has explored the dynamics of one-on-one captioned conversations between DHH and hearing people. In a lab study, DHH/hearing dyads reacted positively to an app supporting ASR-captioned speech and typed contributions [14] and Mallory et al. found preliminary evidence supporting this tool's usefulness in the workplace [41]. Seita et al. [53, 54] investigated DHH people's preferred communication style for in-person ASR use, with DHH participants responding to a hearing actor using varied speech behaviors (e.g., rate, intensity, enunciation, eye contact). They found, for a small set of in-person cases, that standard or exaggerated behaviors (e.g., over-enunciation) may be preferable to minimized behaviors (under-enunciation). In contrast, our study explores practices and preferences during online ASR caption use by small groups.

Most relevant to our study is research exploring the impact of group dynamics on captioning. Seita et al. [52] quantified the impact of ASR use on hearing people's speech during small-group interactions with DHH people, finding they speak louder, faster, and with non-standard articulation. McDonnell et al. [43] argue that captioning research must account for social, environmental,

and technical factors and report on DHH participants' positive reactions to design probes that target hearing people's behavior. We use their design probe findings to anchor our codesign activities, but we expand to working with established mixed DHH and hearing groups and engaging in codesign, rather than prescribing design ideas. Seita et al. [55] also used McDonnell et al.'s probes as a backdrop to determine the best methodological practices for codesign with DHH/hearing dyads. They briefly summarize dyads' designs; pairs suggested hearing people monitor and correct ASR errors and explored behavior notification design paradigms (e.g., icons, overlays). We focus on established small groups, rather than one-on-one conversation between strangers and, rather than method, focus on participants' desired designs and their social impact.

Other work has explored how to support groups with DHH members in accessible communication not mediated by captioning. Brandão et al. [8] built a well-received tool to help instructors regulate their lecture pace to be accessible to students using interpreters. Wang and Piper [61] explored the collaborative practices of established DHH/hearing dyads, finding that they co-created practices to communicate effectively without formal captioning or interpreting services. We extend their focus to the role of social dynamics in small group captioning use.

## 2.5 Videoconferencing Accessibility

Videoconferencing environments present unique considerations for DHH users, particularly after the software's surge in use during the COVID-19 pandemic [3]. These platforms pose challenges for many DHH communicators: Ang and Liu et al. [50] highlight limitations for signed communication while Vogler et al. [60] identify significant technical workarounds required for DHH-accessible hybrid meetings. Studies with remote DHH employees highlight high cognitive load and difficulty identifying speakers [39, 58]. To mitigate platform failings, Kushalnagar and Vogler [35] provide practical recommendations for DHH-accessible videoconferencing, including strong conversational guidelines and monitoring chat. Building from this work, we focus on videoconferencing as a unique, but difficult conversation environment and target designs to better support DHH people during video calls.

## 3 METHODS

This research employs a codesign methodology to explore the experiences of established mixed-hearing ability groups during captioned conversations and to design features to support more accessible online group conversations. We recruited groups of participants with at least one person who identifies as d/Deaf or hard of hearing and at least one who identifies as hearing and sought groups who had prior experience using captions when meeting together online. Each group participated in three codesign sessions over the course of Fall 2021 and Winter 2022. We piloted each of the three study sessions with a mixed-hearing group to refine the study protocols prior to meeting with participants.

### 3.1 Study Procedure

Over three sessions, we explored groups' current use of real-time captioning, their ideas for future captioning technology to support accessible group communication, and their reactions to video

prototypes of these ideas. Sessions were loosely structured and tailored to individual groups' needs, in line with codesign best practices [51]. Each study session was conducted by two members of the research team: the hearing lead researcher and, with one exception, a hard of hearing research assistant. Sessions included either CART captioning or ASL interpreting and automatic captioning, depending on the group's preferences. We conducted all study sessions over Zoom. See supplementary materials for session protocols.

**3.1.1 Session 1: Questioning Current Practices.** The first session explored how each group uses captions and ideas for technology that could support accessible group communication practices. After introductions, we played the game Twenty Questions<sup>3</sup> using automatic captioning<sup>4</sup> to get groups immediately using and thinking about captions in an engaged conversation and to observe their communication practices to inform questions later in the session. The game's frequent turn-taking and niche vocabulary increased the chances for participants to address captioning breakdowns. Participants played for an average of 17:24 minutes (range: 13:08–25:14), and each group played at least two rounds. Researchers took notes on notable interactions and generated questions about how the group approached conversation during the games.

After Twenty Questions, researchers led a group interview, with reflections on the game and tailored questions about specific aspects of in-game communication. This was followed by questions on the group's background communicating together and practices they use to communicate effectively. Next, we introduced participants to the broader focus of the study: improving the design of captioning tools to support groups in developing accessible communication practices. Researchers inquired about the impact of hearing people's behavior on communication access, and their reactions to the idea of using technology to give feedback about accessible communication practices. Finally, we used the five features proposed by McDonnell et al. [43]—speech rate, volume, caption lag monitoring, speaker identification, speaker overlap warning—to discuss potential captioning tool features and as a basis for collective brainstorming.

**3.1.2 Session 2: Feature Sketches.** The second session developed participants' proposed feature ideas via individual sketching and group discussion. Sessions began with groups reviewing, discussing, and expanding ideas they generated from Session 1. Next, participants spent five to seven minutes sketching out the ideas they felt were most compelling. When finished, participants sent pictures of their sketch(es) via text or email to the research team. We screen-shared these pictures and each participant presented the context and motivation for their ideas. Participants had the opportunity to respond and ask questions about others' ideas. After all group members had presented, the group reflected on the ideas and identified their top three shared design priorities.

After the session, the research team created video prototypes [63] of each group's top three ideas (12 features total) to provide a more tangible representation of each idea for critique and discussion (see Supplementary Materials for video prototypes). We elected to make video prototypes rather than functional prototypes as they allow participants to assess high-fidelity, dynamic implementations of their ideas before they are built, enabling low-cost iteration or abandonment of designs. The prototypes integrated participants' specific designs whenever possible. For each group, researchers made three unique videos that showed a feature's design elements and its usage in a simulated conversation. A final, fourth video showed all the features in use during a short round of Twenty Questions played by the research team.

**3.1.3 Session 3: Design Review.** Finally, in the last session, each group reviewed both their own and other groups' design ideas. To begin, we explained that video prototypes *"take new design ideas and animate them to demonstrate how these new ideas might look and function."* We stressed that video prototyping *"gives us a chance to view how designs work before they've been built, making them easy to iterate on"* and encouraged participants to share their *"honest opinions."* Each group watched their three video prototypes and shared reactions. We then played the final video showing all features in use during a conversation and invited final reflections. Afterwards, we reviewed other groups' video prototypes and feature sets, with discussion after each set. Finally, we concluded by asking participants to reflect on the study as a whole and to share any additional thoughts.

## 3.2 Participants

We recruited participants via mailing lists, social media posting, and snowball sampling. Our study goals focused on how mixed-hearing ability groups with experience using captioning together leverage their relational history to approach accessible communication. Thus, we sought groups of three to five people who knew each other; we required at least one hearing member and one d/Deaf or hard of hearing member, and experience using captioning while meeting each other online. These recruitment criteria were flexibly designed as we looked to learn from small groups as they are, rather than overdetermining the perspectives included in the study. We defined rough guidelines for group size and proportion of DHH and hearing participants rather than standardizing those factors, anticipating between-group variation. Codesign methods gain strength from focusing on the particulars of participants' lives and do not emphasize finding a uniform sample, but rather revealing in-depth insights [21, 51], the goal of our study. We invited three groups, 13 people total (six hearing, seven DHH), who fully met our inclusion criteria (Groups A–C). Because they could offer a complementary perspective, we also invited a group of four participants who preferred signing together and did not fully meet our inclusion criteria (Group D). In total, 17 people participated: ten DHH and seven hearing. We compensated each participant \$200 for their time and contributions. The following sections describe each group; names of participants have been replaced with pseudonyms matching with the letter of their group (e.g., Amelia is in Group A, Barbara in Group B, Colin in Group C).

<sup>3</sup>See supplementary materials for the game description and directions shared with participants.

<sup>4</sup>We told participants that we could revert to CART or interpreting if automatic captioning did not support communication, but no groups took that option.

### 3.2.1 Group A.

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use During Video Conferencing
Amelia	Hard of Hearing	Signing	Multiple times a day
Audrey	Hearing	Speaking	About once a month
Anna	Hearing	Speaking	A few times a year
Allison	Hearing	Speaking	A few times a year

Group A are cousins who live across the country from each other and communicate frequently using the video messaging app Marco Polo. They communicate orally when meeting synchronously (both in person and online) and use automatic captions for occasional online meetings with their entire family. All four group members identify as white and female, and their average age is 27.8 (range 25-30).

### 3.2.2 Group B.

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use During Video Conferencing
Barbara	Having hearing loss	Speaking	Multiple times a day
Brian	Hard of Hearing, having hearing loss	Speaking, writing	Multiple times a day
Blake	Hard of hearing, having hearing loss	Speaking, writing	A few times a week
Bea	Hard of Hearing	Speaking, writing	About once a month
Brenda	Hearing	Speaking	Frequently attends captioned meetings but doesn't use them personally
Bridget	Hearing	Speaking	About once a week

Group B meets weekly as colleagues; their work focuses on technology to support DHH people. They communicate orally with Zoom's automatic captions. While our recruitment materials sought groups with up to five members, we opted to include this group of six since they regularly meet. Five group members identify as female, one identifies as male, all are white, and the average age of the group is 53.7 (range 26-67).

### 3.2.3 Group C.

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use During Video Conferencing
Camille	Deaf	Signing	About once a day
Cad	Deaf	Signing	About once a month
Colin	Hearing	Speaking	About once a month

Group C are friends who know each other through the Deaf community. Camille and Cad are Deaf; Colin is a child of Deaf adults (CODA) and knows American Sign Language (ASL). However, all three have experience using captions while video conferencing. Colin could not attend Study Session 2 at the last minute, so some group discussion of design ideas occurred asynchronously via email. One group member identifies as female, two identify as male, all are white, and the average age of the group is 53.67 (range 42-61).

### 3.2.4 Group D.

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use During Video Conferencing
Daisy	Deaf	Writing, signing	A few times a month
Deanna	Deaf	Signing	Rarely
David	Deaf	Writing, signing	Multiple times a day
Dot	Hearing/Acquired hearing loss	Speaking	No prior experience

Group D formed around three friends (Daisy, Deanna, David) who know each other through the Deaf community; Dot joined as Daisy's mom, but she does not regularly chat with the others. Daisy, Deanna, and David primarily communicate via ASL; Dot knows some signed English but is not fluent in ASL. Dot at times identified herself as hearing and as having acquired hearing loss. In contrast to our recruitment criteria, not all members use captions when communicating with each other, and Dot had not previously experienced captioned video calls.

While Group D did not meet our recruitment guidelines, we accepted Mack et al.'s [40] invitation to adapt our study design to work with participants with valuable experience who may not fit all study criteria. Group D provides a useful perspective on how captioning might be used by people who prefer sign language and how to facilitate group communication across language barriers. However, we note areas where they may have different needs and requirements for technology than groups who frequently opt to use captioning. Due to technical

issues in Study Session 2, design feature sketching and discussion was conducted asynchronously via email. Three group members identify as female, one identifies as male. Three members are white and one is AfroLatina/x and South Asian. Their average age is 48.5 (range 36–65).

### 3.3 Analysis and Positionality

We analyzed our data using reflexive thematic analysis, as outlined by Braun and Clark [9]. We took a semantic and critical realist orientation to the data, with an inductive approach to groups' captioning practices and experiences and a deductive approach to their designs for future captioning systems. The first author led analysis, beginning by reading transcripts, taking notes of recurring patterns, then synthesizing notes into an initial codebook which they applied to data from groups A–C. Because researchers analyzed and integrated questions about communication during Twenty Questions in the moment, we did not directly analyze that data *post hoc*. Other authors reviewed a subset of coded transcripts, providing feedback and comments on the first author's coding. The final codebook consists of three versions for each session of the study procedure; codebooks for the second and third sessions incorporated all codes from the previous sessions as well as new, tailored codes. The final codebook is available in Supplementary Materials. Through discussion with coauthors, the lead researcher combined codes and data into the themes that now serve as findings subsections. Then data from Group D was coded and integrated into these themes, with an additional code to note when their experiences differed from the other groups. Members of the research team identify as hearing, hard of hearing, and Deaf. The first author, who facilitated study sessions and spearheaded analysis, is hearing and an ASL student.

## 4 FINDINGS

Below, we describe our findings: we first highlight participants' current practices using captioning to communicate online, then discuss participants' future design considerations and reactions to video prototypes. Throughout the findings we emphasize considerations around how captioning technology may impact social access strategies, both currently and in the future.

### 4.1 Current Practices

Drawing on our participants' established communication practices, we detail individuals' use of captioning online, examine how groups communicate, and explore the development of accessible communication norms.

**4.1.1 Individual Captioning Practices.** Individual participants' approaches to caption use online varied based on their hearing ability and communication practices. DHH participants balanced using captions alongside their residual hearing and speechreading skills. Some only used captions *"if the audio got spotty or something"* (Amelia) and used them *"part of the time"* (Brian), depending on the group's familiarity and if speakers' camera feeds were available. Barbara described needing to *"stare at the captions"* to follow a conversation, but that she compensated for captioning errors by using context and her residual hearing, such that her *"brain is constantly correcting and not paying attention to those corrections"* (Barbara). However, for Bea, motion from automatic captions *"would grab my*

*attention and distract me,"* making them something she only uses as a backup. For others, captions operate as *"the primary feed of information"* (Camille). Cad explained that relying entirely on visual information sources (e.g., captioning, facial expressions, signing, chat), makes video calls *"twice as hard for us as it is for hearing people."* Notably, participants approached captioning differently depending on their ability to participate in the conversation without it.

For participants in Group D who primarily communicate via sign language, captions were often a secondary source of information during interpreted conversations. Daisy leverages captions as a backup to ensure that, if interpreters encounter difficulties, she *"do[esn't] miss too much information"* and so that she can *"make sure that the interpreter is voicing what I say correctly."* David finds that captions alone are not sufficient for expressive communication, as he *"would use sign, I would not type, I would not chat."* However, he runs a third-party automatic captioning app during all video calls so that he does not miss key information if he looks away from a conversation. Despite these limitations, Deanna emphasized that captions do provide value, as when she's *"one Deaf person in a group of hearing people . . . captioning gives us a way to be involved."*

Hearing participants largely did not report using captions during video calls. Some participants noted that while they may check captions out of curiosity, they usually would not have them running on their screen (Anna, Audrey, Allison). Though Brenda serves as the moderator who works to ensure access during her group's conversations, she noted that she *"did not use the captions much—I mean certainly not for communication access."* Colin demonstrated a mode of being more attentive to captions: while he stated he was only *"paying attention a little bit"* he also described actively waiting to participate until captions had caught up and monitoring them for errors. Even in groups with active attention to accessible norms, it was not assumed that hearing people would be paying attention to or even viewing captions.

**4.1.2 Group Communication Practices.** Groups developed specific practices and norms to fit their conversational context and interpersonal relationships. We highlight examples of how each group's context shaped their communication, then identify key takeaways.

Group A highlights a form that access can take within families. Group members have been close since childhood but had not established explicit access norms. They explained, *"We never really talked about accessibility for a hard of hearing person because when you grow up with it, it's just — you already have your system down"* (Allison). Amelia affirmed that, while access needs to be actively considered with many others in her life, *"I'm very comfortable with these ladies and I'm able to understand them very well."* Further, they usually communicated via the app Marco Polo, where users send recorded videos back and forth, which Audrey described as *"super nice because you can't speak over each other."* Despite not having in-app captions, Amelia could access Marco Polo via the Live Transcribe functionality on her phone. Group A did not intentionally create accessible norms, but their established relationships left Amelia feeling well-supported.

Group B is comprised of colleagues who work in DHH spaces, and, in contrast to Group A, they actively enact accessibility practices. They have developed explicit conversation rules, including

clear turn-taking and monitoring for caption errors, which they share and teach when outsiders join their group. They developed their rules and habits through time and close collaboration. For instance, Brenda reflected on her communication with Barbara:

"I too watch faces . . . I've sort of learned when [Barbara's] looking at the captions and it looks like she's not understanding something, then I'm immediately reading the captions myself to see 'oh did it not get it right?'"

This team climate, where hearing and DHH members alike attend to access, was special for Blake: *"I was pretty emotional after the first meeting because it was just so inclusive. . . . It was a really dramatic difference having those set norms . . . that gives you time to catch up and it gives the captions time too."* Group B created an environment where all members can effectively collaborate through an active commitment to accessibility norms.

Group C reflected on how the expectations of conversation partners impacted access. For instance, when Cad spends time with his wife's non-signing family via Google Meet, *"they also have their issues with the captioning and we laugh about it."* Camille pointed out that this becomes possible in *"an environment that's more accepting of flaws."* Participants also speculated about how the study game of Twenty Questions would have differed with *"10 hearing people who knew nothing about Deaf people in the room . . . they would be all talking over each other"* (Colin). Camille emphasized that, if playing a game with those group dynamics, *"I'm sure I would just fade away and not even be a part of it."* However, in Group C, the hearing member, Colin, is a CODA, so he knows how to differentiate *"hearing norms, Deaf norms, hearing values in a meeting, Deaf values in a meeting"* (Camille). In Group C they *"didn't have to say the rules"* (Colin), and implicitly understood that Colin would adjust to Deaf norms, rather than Camille and Cad adjusting to hearing standards.

Group D focused on how their access provisions were shaped by their communication partners. For Daisy, captioning is a *"less commensurate method"* of communication access, but when without an interpreter, she communicated using *"a combination of gesturing, and signing, and captioning, and typing."* Deanna explained that her goal in communication was *"to be very accommodating and flexible to deal with whoever is there—the point is accessibility."* Yet, group members noted that, in their experience, *"hearing people . . . don't always have a lot of empathy or understanding"* (Dot). This led to frustrating interactions, such as a conversation between Deanna and a new family member who refused to write back and forth and insisted on inaccessibly voicing. Group D stressed that accessible communication, especially for signers who aren't fully supported by captioning, requires mutual flexibility.

While all four groups found that access becomes possible when conversation partners work to meet each other's needs, the form of those approaches varied significantly. Participants also described conversations with others outside their group that took different approaches, to varied degrees of success. Ultimately, this demonstrates that there is no one-size-fits-all approach to accessible conversation between DHH and hearing people. Rather tailored, contextual access evolves and is informed by conversation partners' relationships with each other.

**4.1.3 Developing Group Access Norms.** Groups reflected on the positive impacts and the challenges of developing access norms in mixed groups. Participants, DHH and hearing alike, had experiences where collective effort resulted in more comprehensible communication. Barbara found that most of the hearing people she communicates with *"understand how to make themselves understood to people and understand what captions are about."* Still, when outsiders join groups that have set rules there is a learning curve, as *"it was hard for [them] to slow down"* and led to *"a stop/start environment"* (Bridget). Brenda explained that she actively moderated those conversations, hoping that *"when a hearing person also makes those requests and reminders, that it just helps reinforce the need for making sure things are accessible."* Camille reflected on similar dynamics, concluding that conversations could be *"easier or harder, depending on who the hearing person is... Are they aware? Are they unaware?"* For her, successful, accessible communication requires *"sometimes not following the hearing pattern of turn-taking or communication"* (Camille). Desired behavior changes can also depend on the context. For example, if Amelia was the only DHH person in a conversation, she was less likely to ask others to adjust, because *"if I understand it, then it's fine."* On the other hand, in conversations that include other DHH people, she would *"try to make people aware"* of communication rules to *"support each other"* (Amelia).

However, participants noted that behavior change was not always a smooth process. Barbara reflected that it *"takes people time and experience to adopt those norms"* and that regular practice and reinforcement were critical. Attending to access often requires significant effort from hearing interlocutors. Colin described his experience monitoring captions for errors as he and others spoke: *"if it wasn't right, I wanted to fix it, but I also didn't want to jump in and fix it, I wanted to let you try to repair it for yourself if you wanted to. . . . There's a lot of cognitive load there, and I'm not even the Deaf person."* Likewise, Allison explained that she *"can't do two things at once"* and monitoring multiple sources of information while speaking was not feasible for her—despite her commitment to communicating accessibly. Therefore, when considering Cad's observation that *"there are hearing people who are experienced, who are cognizant and mindful, and there are those who are not,"* we must also recognize the complexity involved in learning and applying this expertise.

## 4.2 Future Designs

After understanding groups' current practices, we codesigned features to support them in having more accessible captioned conversations online. In this section, we first present the 12 ideas participants proposed (three per group), which our research team developed into video prototypes in between Sessions 2 and 3 (Table 1). Then we discuss all groups' envisioned motivations for and reactions to these feature prototypes, which fell into four major categories: 1) identifying speakers and overlap, 2) feedback systems to address conversation breakdown, 3) videoconferencing infrastructure, and 4) non-speech sound information. While many of these ideas have been previously explored (e.g., [43, 55]), as part of a commitment to codesign, we prioritized participants' enthusiasm for features over novelty. We also highlight new facets of these features by exploring how they may impact group dynamics—an area that has been



overlooked in previous research. This section focuses on future designs, but, when relevant, we include some discussion of current experiences that participants used to explain their reactions to or motivations surrounding particular features. While participants were not asked to focus on Zoom during their design process, many features were designed with Zoom as the starting point as it was the platform used throughout the study.

**4.2.1 Codesign Artifacts.** Table 1 shows key frames from each of the video prototypes that researchers developed based on participants' ideas. These prototypes reflected participants' discussions across their proposed ideas, searching for convergence in ideas and priorities. For example, Cad drove discussion with his group, saying *"I think we are all looking to make it easier to ID the speaker quickly and easily."* Groups assessed their video prototypes in depth and also reviewed other groups' designs.

**4.2.2 Speaker Identity and Overlap.** Quickly identifying speakers and automatically flagging when they overlap were clear priorities for participants, with Groups A, B, and C designing features (Table 1, A3, B2, C3) to address these information gaps. While speaker identity and overlap have both been explored in prior literature (e.g., [18, 23, 24, 35, 37, 39, 47, 58]), participants' experiences demonstrate that ambiguity around crosstalk and speaker identity has yet to be resolved, with key design nuances and social impacts for each feature in group videoconferences.

Though some human-generated (e.g., CART) and automated (e.g., Google Meet) services integrate speaker identity into captions, participants wanted it to be a universal feature. Many groups (A, B, C) favored conveying speaker identity in captioning by splitting captions up by speaker and using a visual indicator connected to their name (Table 1, A3). Groups initially proposed color-coding speakers, provided it met colorblindness and other visual accessibility standards, and later also considered using profile images. Group C envisioned other ways they might like to see speaker identity (by displaying captions under the active speaker's video, and with threads between the speaker and the captions), and therefore proposed providing multiple options that each user could select as desired. Separating captions by speaker and identifying the speaker's name were universally liked and showing threads between speakers and captions (as designed in our video prototype) was comprehensively disliked. Placing captions below the active speaker's video (Table 1, C3) garnered mixed reactions. Amelia liked that it would help consolidate information and limit instances where people *"don't even look at the captions because [they're] too busy looking at the person talking"* (Amelia). However, others felt that *"would be a lot to follow"* (Anna) and *"take more cognitive effort"* (Camille), particularly in meetings of five or more people.

Participants who frequently joined conversations with interpreters considered how speaker identity should work when an interpreter voices for multiple signers. They stressed that *"when there are a lot of Deaf people it's not really entirely clear who's doing the actual uttering"* (David) and that *"there needs to be some way of the computer knowing to connect that interpreter to me so that [speaker identification] comes to me when I'm the one speaking"* (Daisy). Deanna suggested enabling interpreters to click a button that would then *"show who exactly they're voicing for"* (Deanna).

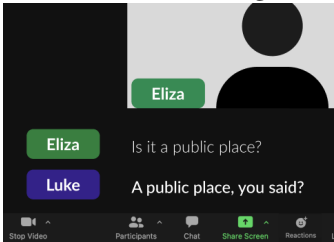
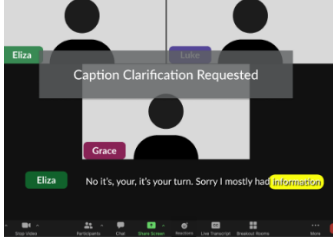
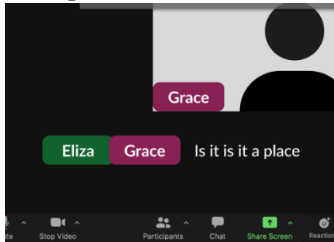
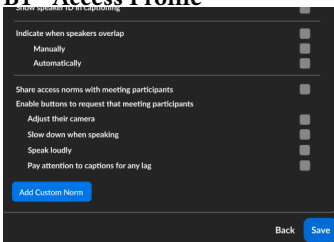
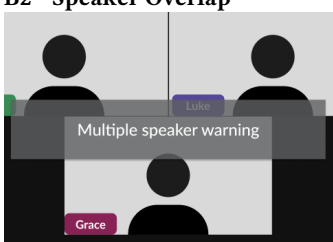
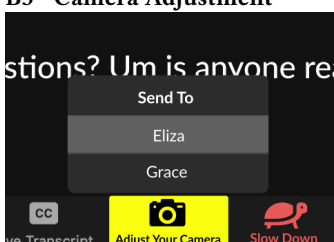
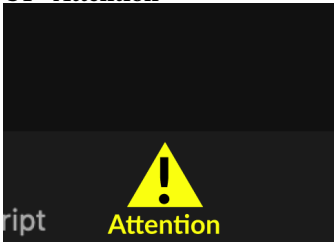
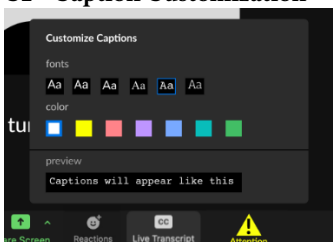
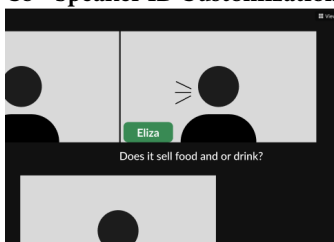
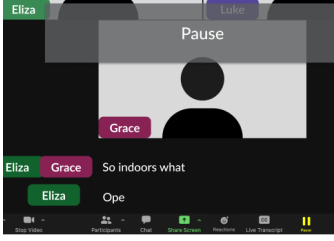
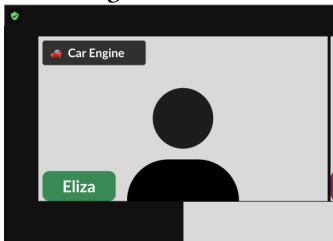
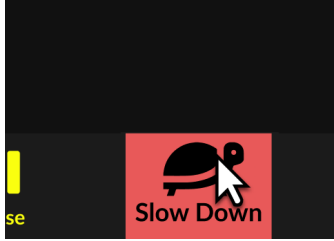
Building on speaker identification, participants highlighted the importance of identifying overlapping speech for both DHH and hearing conversants. Overlapping speech poses a significant access barrier that is not well-addressed by current captioning solutions, leaving captioning users out of conversation. Barbara motivated the need for a tool to limit overlap by explaining; *"[with] automatic captions, or even CART, there's no way to capture [overlap]... I'll just pull away emotionally or walk away physically."* Therefore, participants proposed both using speaker identification approaches to indicate overlap and sending conversation participants a pop-up when it occurs. When reacting to video prototypes of speaker overlap notifications (Table 1, B2), participants liked the baseline feature but had suggestions for nuances to build into future implementations. Participants stressed the importance of language choice in notifications, as having the pop up read 'multiple speaker warning' made participants feel like *"something bad's gonna come out of the screen and grab you or something"* (Barbara). Both displaying overlapping speakers' names in the captioning and sending an alert was valuable to some—one hearing participant noted: *"I would not have my captions on, so I do like that it pops up and makes it very, like, front and center"* (Anna). When considering the impact of getting an alert, Blake remarked *"[I liked] how annoying that is gonna [be] for teams that talk over each other a lot . . . because I think that creates an incentive . . . to create more accessible, inclusive conditions."* However, participants also envisioned possible negative impacts of penalizing overlapping speech: *"people that are more shy, more self-conscious . . . may start to feel afraid to say anything"* (Brian). Participants also worried that alerts could limit equitable turn-taking, as they could cause *"the dilemma of 'do I interrupt, or do I let this person take all the time?'"* (Blake). Groups A and B liked that speaker overlap alerts could guide behavior change, since *"it's something that happens over time with reminders"* (Brenda), but others worried that that *"most hearing people are used to being interrupted or talking over each other . . . so they don't want this visual alert"* (Colin).

**4.2.3 Support for Behavioral Feedback.** The next category of features involves a set of feedback mechanisms to alert the group to ways to make conversation more accessible. Participants proposed a variety of possible behaviors to alert conversation partners about, often using a similar pop-up implementation (Table 1, A2, B3, C1, D1, D3). Specifically, participants' designs focused on providing captioning error feedback, asking others to adjust their camera, a communication breakdown alert, asking for a pause in conversation, and asking other speakers to slow down. While these access barriers have been discussed in prior work (e.g., [8, 11, 25, 27, 33, 43, 52, 53]), we focus on how technical tools could help to mitigate them by guiding behavior change.

While Group A's design (Table 1, A2) included the opportunity to either flag or correct captioning errors in real time, participants only saw promise in flagging errors. Driving the focus on errors is the reality that, while DHH participants recounted many nonsensical errors in captions, such as *"calling 'site administrator' 'satan administrator'"* (Bea), hearing conversation partners rarely noticed or tried to address them. Despite *"lov[ing] the idea of being able to correct [captions] as we go"* (Daisy), participants concluded that doing so was *"a little too far-fetched"* (Anna) because it would be



**Table 1: Each group's top three ideas for captioning tools, including the feature's name, a key frame from the video prototype, and a description of the feature in a participant's own words. To view full versions of video prototypes, see supplementary materials.**

Group A	<b>A1 - Volume Monitoring</b>  <p><i>"Bigger or bolder font as they get louder"</i> – Anna</p>	<b>A2 - Error Correction</b>  <p><i>"Subtly [let] the speaker know if something was unclear"</i> – Anna</p>	<b>A3 - Speaker Identification</b>  <p><i>"[It lets you] intervene and say, 'Hey, like, turn-take here'"</i> – Amelia</p>
	<b>B1 - Access Profile</b>  <p>An accessible group "standard ... provided automatically" – Brenda</p>	<b>B2 - Speaker Overlap</b>  <p><i>"In the middle of the screen, in a way that it's not too jarring, but you can't ignore it"</i> – Blake</p>	<b>B3 - Camera Adjustment</b>  <p><i>"If I'm talking ... and I'm muted, ... that same kind of approach could be used"</i> – Bea</p>
Group C	<b>C1 - Attention</b>  <p><i>"Something is wrong' button [that] mean[s] several things; 'hold on', 'I'm lost' ..."</i> – Cad</p>	<b>C2 - Caption Customization</b>  <p><i>"Adjust the captions' size, colors, font from within the meeting"</i> – Colin</p>	<b>C3 - Speaker ID Customization</b>  <p><i>"If I don't know who's speaking, I don't have the contextual information."</i> – Camille</p>
Group D	<b>D1 - Pause</b>  <p>A way to "realize ... I need to wait" – Dot</p>	<b>D2 - Background Noise</b>  <p>Could display background noise like "barking in the background" – Daisy</p>	<b>D3 - Slow Down</b>  <p>An "alert for speaking too fast," – Daisy</p>

cognitively overwhelming and likely too delayed to be useful. However, participants saw social benefit in being able to flag caption errors anonymously, imagining that it would help *“that shy person who doesn’t want to interrupt”* (Blake) and allow users to ask for repeated clarification without feeling like there’s *“a target on you”* (Amelia). However, others worried that it could make *“the flow of the conversation stop”* (Dot). While the video prototype simply alerted that an error occurred, participants proposed that the alert should point out the confusing caption in context, since it was likely said *“10 seconds ago, and then you’re like, ‘Oh, well, what word was it?’”* (Colin). In summary, real-time error correction may not be feasible, but participants were enthusiastic about being able to call their conversation partners’ attention to errors that impacted comprehension.

For participants who relied on seeing conversation partners clearly, being able to discreetly ask someone to adjust their camera view was exciting, but discussion revealed social complexity in doing so. Many participants shared Amelia’s video conferencing experience of feeling *“like, I don’t know who’s talking . . . can you please turn on your camera.”* Participants saw the use case for being able to alert someone to adjust their camera – in fact, while discussing this video prototype (Table 1, B3), Daisy took the opportunity to tell David *“you’re kind of cut off at the neck . . . you gotta move.”* However, participants noted that just telling someone to adjust their camera without a reason or specific directions was too ambiguous, proposing that the alert could specifically mention *“someone can’t read your lips right now”* (Blake). However, participants also pointed out reasons why camera use was not always desired. For some *“the exhaustion of being on camera all the time”* (Bridget) is significant, and others found that only having active speakers or signers on camera could help minimize visual overload (Barbara). Participants also considered innocuous reasons why someone might not be visible on camera, such as when *“someone could be holding their puppy and it’s in front of their face and you can’t see their face or lips”* (Blake). Participants stressed the importance of thinking about the need for clear camera feeds within broader social context and cautioned that norms around using such an alert needed to account for nuance.

Participants also spoke to the need for a mechanism to tell conversation partners to slow down and were positive about Group D’s design (Table 1, D3). Speech rate alerts were considered in context of participants’ current strategies to get speakers to slow down. When considering how groups might address speech rate, Brian shared how he approaches new vendors at work who *“just talk too fast.”* He begins conversations by saying *“I would really appreciate it if you just slow down your voice, just a bit, so I can follow what you’re saying.” . . . They will at some point speed up again; [I] just kindly remind them.”* When assessing the speech rate video prototype, Amelia imagined that it would be *“a good way to teach people how to have good speaking skills”* by providing a mechanism to unobtrusively remind speakers when they speed up. Blake saw additional benefit in being able to get feedback on her speaking rate, saying, *“having hearing loss . . . we don’t really have to hear ourselves, I’m always someone who . . . tries to work on slowing down”*. While the prototype showed manual speech rate alerts, participants also proposed automated speech rate monitoring, either in the form of

auto-generated alerts (Brian) or adding a speedometer visualization for speakers (Daisy, Dot).

The final two behavioral feedback ideas, a pause (Table 1, D1) and attention (Table 1, D3) button, had similar motivations – identifying and encouraging groups to address conversational breakdown – and participants had mixed reactions to both. Participants from Groups C and D, who frequently communicate via interpreters, proposed a way to address the fact that delays in communication often mean that *“once I finally get to that point where I can actually add something [to the discussion] . . . now it’s the wrong time”* (Daisy). However, despite agreeing that this was a common problem that needs addressing, participants’ reactions to attention or pause pop up alerts focused on the need for stronger guidance and mixed feelings around halting conversations. Anna suggested that alerts provide more specific guidance than simply calling for ‘attention’ as she felt that notification would cause her to *“panic and . . . not know what to do from there.”* Most participants worried that a pause or attention button would be too disruptive to a conversation or become *“something that gets abused”* (Amelia). However, Blake *“got really excited”* about building tools into a system that could *“empower the person who’s maybe too timid to speak.”* Participants brainstormed ways to address the need to identify breakdown while minimizing disruption, proposing that it could be *“up to the person who’s pressing the alert button whether or not they want to send that alert just to the host, or to everyone”* (Colin). Participants were united on the importance of calling attention to conversation breakdowns but after watching their simulated use, concluded that disruptive alerts were not the right tool to address this need.

**4.2.4 Videoconferencing Infrastructure for Accessibility.** Another target for technology that could support groups in more accessible conversation was videoconferencing platform infrastructure itself. Participants focused on adding a new set of access norms in software settings that could build group norms into the platform (Table 1, B1) and desired greater customizability over current captioning interfaces (Table 1, C2). Customizability has been highlighted throughout prior work [11, 12, 18, 43], but the role of platform infrastructure on accessibility for DHH people has so far only been explored in the context of sign language use [50].

Group B’s access profile (Table 1, B1) allowed groups to enable desired features (such as the behavioral feedback tools discussed in Section 4.2.2) and share social norms. Across groups, participants were excited about the idea of an access profile and brainstormed ways to address the many complexities it introduces. Camille reflected that *“we want technology to solve things, but we realize that people have to modify, they have to change”* and imagined that building access norms into a system was a way to *“leverage technology to help.”* Participants highlighted the benefit of having preset but highly configurable options, as there are common issues that *“d/Deaf people agree are the pain points when attending an online meeting,”* but for individual groups, settings *“should be able to be customizable”* (Colin). Additionally, Blake considered how, often when joining a meeting, *“I’m going really fast and I’m not setting things up ahead of time”* and Barbara suggested that settings *“should be real easy for the consumer to turn on or off, even if you’re in the middle of a Zoom meeting.”* Bea proposed that users should be able to save context-specific presets, for *“big meetings, classes, small*

meetings." While participants valued having an anonymous way to request their needed access supports and norms, questions arose around misuse (e.g., malicious users). Broadly, participants were excited about the possibility to build accessible conversation norms into videoconferencing systems and continued to think through the nuanced factors that would make such a tool effective.

Highly customizable captioning displays were an additional area of videoconferencing infrastructure with unified support. Customization is not a new concept – in fact Amelia responded to this prototype (Table 1, C2) by "*check[ing] the subtitle settings. I was, like, 'do they not have settings for captions?' And they don't. I was really surprised.*" In light of the lack of control in current tools, participants highlighted the specific dimensions that were important to them. Customizable colors were critical, as Cad explained that for "*DeafBlind people who have some vision but need some accommodation*" there can be some "*contrast of colors [or] particular colors that are better than others.*" Other features included resizing the "*short little box*" (Camille) used to display captions and supporting users' preferred setups by letting "*captions show up in a separate browser tab*" (Colin). While not necessarily a novel technological innovation, the control that platforms do or don't allow their users significantly shapes who can use captions effectively, and participants highlighted the need for greater control and customizability.

**4.2.5 Sound Information.** In the final category of designs, participants proposed providing more information about sound in addition to transcription. Designs included visualizing speaker's volume (Table 1, A1) and identifying non-speech sounds (Table 1, D2). Though Google has integrated sound recognition into the Live Transcribe app [69], it is not available within commercial videoconferencing tools. Volume visualization has only been explored in the context of pre-recorded captions [23].

Group A's proposal to visualize speaker volume in captions (Table 1, A1) seemed promising during their brainstorming but, in viewing it, participants identified more problems than benefits. Amelia initially was motivated to display volume as a proxy for tone, which is "*very easily misunderstood just reading the captions*" (Amelia). However, many questioned if simply showing volume could lead to misunderstandings. For instance, knowing that a person is speaking quietly could pose the question: are they "*unsure or meek? Or is it just that they [a] quiet talker?*" (Colin). Additionally, participants considered that volume may not always vary significantly, such as with Audrey's family "*who have one volume, and it's yelling . . . it would just be bold the whole time.*" While participants liked the idea of volume displays that were "*dynamic without being disruptive*" (Daisy), they concluded that this implementation would not be useful.

In addition to captions, participants were interested in identifying and displaying background noise during a conversation. Daisy described why background noise notification is needed: "*I can't tell you how many times I'll be talking to someone . . . and the hearing person suddenly looks off into the horizon . . . I'm like 'hey, what's going on? Why, you know, why is the conversation being disrupted? I can't hear that.'*" DHH participants valued the idea of displaying background noise within a videoconferencing tool with both an emoji and text description of the sound (Table 1, D2), though hearing

people questioned its necessity. Participants favored using emojis paired with background noise descriptions to "*just get a quick bit of information*" (Daisy), and the colorful nature of emojis makes alerts "*really bright and easy to capture*" (Deanna). However, rather than placing background noise alerts in video feeds, participants suggested they would be "*better in a bottom corner*" (Dot) or "*between sentences in the captioning*" (Deanna) to avoid splitting users' visual attention. When considering the social implications of this design, some hearing participants believed that background noise happening around them is "*not necessary for other people to know this, it's really more for the speaker*" (Bridget). However, DHH participants stressed that knowing a noise is happening "*just gives us clarification of why you're pausing*" (Amelia), and that it would be useful for people with some residual hearing to know "*is [a noise] me or is it someone else?*" (Barbara). Participants also wondered how to determine "*what the threshold is*" (Brenda) for identifying sounds—as Colin put it, "*some hearing people hear background noise and either intentionally or unintentionally ignore it . . . and other times it's like 'whoa I heard that fire alarm.'*" Being aware of background noise is critical for DHH conversants, and participants brainstormed how to best communicate that over videoconferencing platforms.

**4.2.6 Summary.** Our participants first designed features to address their most salient communication hurdles and then, by engaging with video prototypes, surfaced aspects of their designs that would need to be carefully considered to fit the social norms of mixed-hearing ability group conversations. Participants identified features that have already been implemented in captioning in the past, namely speaker identification and caption customization, as high-priority and high-impact to universally build into videoconferencing tools. Guiding groups to be more aware of speaker overlap, speech rate, comprehension-critical caption errors, and the need to adjust their camera via pop-up notifications are promising features for future development and innovation. Participants also wanted to be able to set, share, and customize access practices within videoconferencing platforms, and this is a promising new paradigm. Finally, products like Live Transcribe [69] have integrated sound recognition into their ASR apps, and our findings indicate that this would be valuable during videoconferences as well. Participants also identified features they did not want. While increasing group awareness of conversation breakdowns was important, pop-up alerts did not prove to be an appropriate approach. Additionally, displaying raw volume information by styling the captions themselves was seen as distracting and unclear.

## 5 DISCUSSION

In this paper we report on our codesign practice with established groups of DHH and hearing people. In Section 4.1 we find that people actively negotiate and build accessible group practices on top of captioning use and, in Section 4.2, identify participant-driven priorities for future videoconferencing features to support accessible group communication strategies, leveraging their past experience negotiating access together. We now situate our findings in related work, identify key priorities for future videoconferencing design, and reflect on approaching captioning design with a collective access lens.

## 5.1 Designing with Established Groups

Though involving existing groups for a multi-session study is challenging and effortful for the participants and research team alike, this method was crucial in shaping our findings. Through code-signing with our participant groups, we were able to observe and ask about their established practices, gain insight from the questions participants had for each other throughout the process, and learn from multiple perspectives on the same experiences. Further, learning from established groups reveals the communication access problems and social interventions that persist after people move past surface-level interactions or learn the basics of communicating with DHH people. Beginning with this deeper understanding of possible approaches to communication access could lead to richer tools for new groups (e.g., students working on a class project, new colleagues) or people interacting briefly (e.g., interactions with a telehealth nurse). We argue that paying attention to the rich relational context of established groups allows HCI researchers to identify pressing problems and promising avenues to address them in future captioning tools.

Our findings offer new insights on communication practices between DHH and hearing people and the design of captioning and videoconferencing tools. First, we document the accessibility practices of established groups, including hearing people and groups with multiple DHH people that have varied access needs. Prior work focused on the captioning experiences of individuals [29, 43] and communication practices of established DHH/hearing pairs [61], whereas we highlight varied ways that existing groups with mixed-hearing abilities engage with captions (e.g., work colleagues established formal rules while friends relied on established Deaf community values). Additionally, the variation between groups' practices highlights the extent to which communication practices and preferences are shaped by the specific people present for a conversation. While prior work has broadly explored the impact of hearing people's behavior on a conversation [52–54], we suggest that this behavior must be contextualized by the relationships between hearing and DHH conversants because its impact is not consistent across conversations (e.g., norms between strangers, family, and disability activists will likely differ). While understanding how to support DHH and hearing people communicating together is critical, we argue that there is not one universal solution waiting to be built. However, we see great promise in building a tool that can be customized to support cousins who have been communicating together since childhood and colleagues working on communication technology for DHH people alike in negotiating and sustaining a group commitment to conversation accessibility.

## 5.2 Implications for the Design of Future Videoconferencing

While prior work has identified the value of conversation norms (e.g., [15, 29, 43, 61]) and Seita et al. [55] briefly discussed new features to guide conversation, our work proposes participant selected and designed features and assesses ideas in context of their social impact. Specifically, participants desired videoconferencing platforms to support established captioning features (e.g., speaker identification and customization), wanted new ways to make conversants aware of speaker overlap, speech rate, comprehension

critical errors, and camera feed quality, saw opportunities to build tools for mixed-hearing ability groups into platform infrastructure, and wanted to also be aware of background sounds.

Beyond specific feature designs, our codesign sessions surfaced broader considerations for mixed-hearing ability groups. Participants emphasized that, while it is tempting to imagine solving for conversation access with technology only (e.g., ASR that can perfectly caption overlapping speech), communication access is fundamentally social. However, they consistently highlighted the potential value of technology that helps set and enforce group norms and guidelines. With this paradigm shift in mind, we identify guiding principles for future videoconferencing technology.

- **Low technical complexity, high social impact.** Many of the features our participants identified are not technically complex and leverage existing video conferencing system functionality (e.g., participants' design of a 'slow down' button). However, these tools could be critical in helping shape accessible conversation dynamics over multiple layers of language, communication, and social interaction. We encourage designers and researchers to explore these avenues that may be less obviously novel but desired and socially impactful.
- **Configurability in all facets.** Current videoconferencing platforms do not allow users to control many aspects of the appearance and placement of captioning, despite customization being clearly preferred in prior work (e.g. [11, 12, 18, 43]) and by our participants. DHH captioning users' myriad contexts of use (e.g., with interpreters, at work) and varied abilities (e.g., hard of hearing, DeafBlind) make it so that the ideal captioning style and display is one with high configurability.
- **Design to minimize cognitive load.** Prior work has established that following a conversation with captions is cognitively taxing [33] and our findings affirmed that this is a key factor in assessing the viability of captioning supports. Despite many attempts to convey information through caption styles (e.g., [5, 6, 18, 19, 23, 28, 48, 52, 56]), our participants' agreed that reading captions requires too much cognitive load to make on-caption visualizations useful [6, 56]. Designers should consider ways to augment captions without overloading users, such as placing critical information in consolidated regions of the screen and avoiding distracting edits to the captions themselves.
- **Maximize contextual information.** Notification systems that identified a problem but not a solution (e.g., 'Attention!'), did not adequately describe the steps to resolve a problem (e.g., 'Adjust your camera'), or did not provide full context on the problem (e.g., not identifying which caption was unclear) were not satisfactory interventions. Technology to guide behavior change should succinctly and specifically identify what breakdown has occurred and point to the resolution of that behavior, without assuming that users will be viewing the captioning.
- **Automatic or manual notifications.** While automatic notification systems may lessen cognitive load and outperform human reaction time (e.g., overlapping speaker alerts), participants highlighted the nuanced social context that informs even seemingly obvious cases for intervention (e.g., pointing

the camera at a new puppy). The decision of which features to automate must be done with careful consideration of behavioral nuances.

- **Anonymity in feedback.** Considerations of power dynamics, personality, and frequency of reminders led participants to conclude that they would be more likely to ask for a change they need if they can ask anonymously. Platform designers should consider when an anonymous feature could minimize embarrassment or social judgment to both the requester and recipient of feedback, but also must weigh its potential for misuse or harassment.

### 5.3 Captioning for Collective Access

We situate our findings in the context of disability studies, Deaf studies, and the disability justice principle of collective access. Disability studies and activism focuses on how accessibility is to be addressed in community, rather than on an individual basis [46]. Disability Justice activists have furthered this thinking and operationalized it in their principle of collective access. We found that participants' conversation practices and design priorities for captioning tools frequently demonstrated collective access. Norms were co-created by groups, and all group members were responsible for upholding them. Groups tailored best practices to match interpersonal relationships—trusting that members could and would resolve issues as needed. While these dynamics have been characterized in prior work [43, 61], we argue that the lens of collective access is necessary for a complete understanding of the factors that drive the use of communication technology. Additionally, many Deaf community ideals and practices reject hearing norms, including the architectural practice of DeafSpace [13, 70], which aligns aspects of the physical built environment with Deaf communication norms (e.g., avoiding pillars to maintain clear sight lines). In envisioning the future of videoconferencing tools, we propose building toward a form of digital DeafSpace where DHH communication norms get prioritized and embedded into the platform, rather than designing these tools in ways that are frequently hostile to DHH communicators [50].

### 5.4 Limitations

As a qualitative, codesign study we recruited a relatively small number of participants and those participants had a wide range of experiences. While this allowed us to explore their experiences in depth, we do not claim that these findings are generalizable. We also did not assess design features relative to a consistent set of factors (e.g., the proportion of DHH and hearing group members), and therefore do not speculate about what may have driven some participants' reactions. Additionally, our study was conducted using Zoom, potentially biasing designs and reflections. While we believe that findings are applicable beyond this single platform, future work may want to explicitly explore the ways different platforms impact conversation. Next, our participants were overwhelmingly white (16/17) and all based in the US, which limits our perspective. Finally, while video prototyping allowed us to understand what would or would not support mixed-hearing ability groups in more depth than paper prototyping and without the costs of software development, implementing tools to support mixed-hearing ability groups is a

crucial next step. We see this as an exciting area for future work that builds on our design guidelines.

## 6 CONCLUSION

As expanded use of videoconferencing and ASR reshapes groups' communication practices, we sought to understand how mixed groups of DHH and hearing people negotiate online captioned conversation. By conducting a three-phase codesign study with 4 groups (17 participants total, 10 DHH, 7 hearing), we found that groups develop specific social practices to increase accessibility and identify exciting features for future videoconferencing design that engage DHH and hearing conversation partners alike.

## ACKNOWLEDGMENTS

We thank our colleagues Kelly Mack, Aashaka Desai, Taylor Schenone, Lotus Zhang, Dhruv Jain, and Abigale Stangl for their support and input over the course of this study. Our work was made possible by many captioners and interpreters, and the University of Washington's Deaf and hard of hearing services coordinator, Dimitri Azadi. We also thank our participants for their thoughtful engagement in a long-term, complex study. This work was supported by the National Science Foundation under Grant No. IIS-1763199, the National Science Foundation Graduate Research Fellowships Program under Grant No. DGE-2140004, and by the University of Washington's CREATE center.

## REFERENCES

- [1] Akhter Al Amin, Saad Hassan, Sooyeon Lee, and Matt Huenerfauth. 2022. Watch It, Don't Imagine It: Creating a Better Caption-Occlusion Metric by Collecting More Ecologically Valid Judgments from DHH Viewers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22), 1–14. <https://doi.org/10.1145/3491102.3517681>
- [2] H.-Dirksen L. Bauman and Joseph J. Murray. 2014. *Deaf Gain: Raising the Stakes for Human Diversity*. U of Minnesota Press.
- [3] Macy Bayern. 2020. Zoom grew by 574% in less than two months, but Skype for Business reigns supreme. *TechRepublic*. Retrieved September 5, 2022 from <https://www.techrepublic.com/article/zoom-grew-by-574-in-less-than-two-months-but-skype-for-business-reigns-supreme/>
- [4] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*, 161–173. <https://doi.org/10.1145/3234695.3236348>
- [5] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. 2019. Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3290607.3312921>
- [6] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 155–164. <https://doi.org/10.1145/3132525.3132541>
- [7] 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. <https://doi.org/10.1145/2982142.2982171>
- [8] Alessandra Brandão, Hugo Nicolau, Shreya Tadas, and Vicki L. Hanson. 2016. SlidePacer: A Presentation Delivery Tool for Instructors of Deaf and Hard of Hearing Students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '16), 25–32. <https://doi.org/10.1145/2982142.2982177>
- [9] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide*. SAGE.
- [10] Janine Butler, Brian Trager, and Byron Behm. 2019. Exploration of Automatic Speech Recognition for Deaf and Hard of Hearing Students in Higher Education Classes. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '19), 32–42. <https://doi.org/10.1145/3308561.3353772>

- [11] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 2009. ClassInFocus: enabling improved visual attention strategies for deaf and hard of hearing students. In *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility - ASSETS '09*, 67. <https://doi.org/10.1145/1639642.1639656>
- [12] Becca Dingman, Garreth W. Tigwell, and Kristen Shinohara. 2021. Designing a Podcast Platform for Deaf and Hard of Hearing Users. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*, 1–4. <https://doi.org/10.1145/3441852.3476523>
- [13] Claire Edwards and Gill Harold. 2014. DeafSpace and the principles of universal design. *Disability and Rehabilitation* 36, 16: 1350–1359. <https://doi.org/10.3109/09638288.2014.913710>
- [14] Lisa B. Elliot, Michael Stinson, Syed Ahmed, and Donna Easton. 2017. User Experiences When Testing a Messaging App for Communication Between Individuals who are Hearing and Deaf or Hard of Hearing. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 405–406. <https://doi.org/10.1145/3132525.3134798>
- [15] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 2017. Deaf, Hard of Hearing, and Hearing Perspectives on Using Automatic Speech Recognition in Conversation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 427–432. <https://doi.org/10.1145/3132525.3134781>
- [16] Abraham T. Glasser, Kesavan R. Kushalnagar, and Raja S. Kushalnagar. 2017. Feasibility of Using Automatic Speech Recognition with Voices of Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 373–374. <https://doi.org/10.1145/3132525.3134819>
- [17] Steven M. Goodman, Ping Liu, Dhruv Jain, Emma J. McDonnell, Jon E. Froehlich, and Leah Findlater. 2021. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2: 63:1–63:23. <https://doi.org/10.1145/3463501>
- [18] Benjamin M. Gorman, Michael Crabb, and Michael Armstrong. 2021. Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, 1–11. <https://doi.org/10.1145/3411764.3445509>
- [19] Michael Gower, Brent Shiver, Charu Pandhi, and Shari Trewin. 2018. Leveraging Pauses to Improve Video Captions. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*, 414–416. <https://doi.org/10.1145/3234695.3241023>
- [20] Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-mounted Display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, 1–4. <https://doi.org/10.1145/3373625.3418031>
- [21] Christina Harrington and Tawanna R Dillahunt. 2021. Eliciting Tech Futures Among Black Young Adults: A Case Study of Remote Speculative Co-Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, 1–15. <https://doi.org/10.1145/3411764.3445723>
- [22] 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 (ASSETS '13)*, 1–2. <https://doi.org/10.1145/2513383.2513413>
- [23] 23. Richang Hong, Meng Wang, Mengdi Xu, Shuicheng Yan, and Tat-Seng Chua. 2010. Dynamic captioning: video accessibility enhancement for hearing impairment. In *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*, 421–430. <https://doi.org/10.1145/1873951.1874013>
- [24] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 241–250. <https://doi.org/10.1145/2702123.2702393>
- [25] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 81–92. <https://doi.org/10.1145/3234695.3236362>
- [26] Dhruv Jain, Khoa Huynh Anh Nguyen, Steven M. Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehlich. 2022. ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users. In *CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3491102.3502020>
- [27] Carl Jensema. 1998. Viewer Reaction to Different Television Captioning Speeds. *American Annals of the Deaf* 143, 4: 318–324. <https://doi.org/10.1353/aad.2012.0073>
- [28] Sushant Kafle, Becca Dingman, and Matt Huenerfauth. 2021. Deaf and Hard-of-hearing Users Evaluating Designs for Highlighting Key Words in Educational Lecture Videos. *ACM Transactions on Accessible Computing* 14, 4: 20:1–20:24. <https://doi.org/10.1145/3470651>
- [29] Saba Kaway, George Karalis, Tzu Wen, and Richard E. Ladner. 2016. Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 15–23. <https://doi.org/10.1145/2982142.2982164>
- [30] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 117, 14: 7684–7689. <https://doi.org/10.1073/pnas.1915768117>
- [31] Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, 1–12. <https://doi.org/10.1145/3313831.3376266>
- [32] Raja Kushalnagar and Poorna Kushalnagar. 2014. Collaborative Gaze Cues and Replay for Deaf and Hard of Hearing Students. In *Computers Helping People with Special Needs (Lecture Notes in Computer Science)*, 415–422. [https://doi.org/10.1007/978-3-319-08599-9\\_63](https://doi.org/10.1007/978-3-319-08599-9_63)
- [33] Raja S. Kushalnagar, Gary W. Behm, Aaron W. Kelstone, and Shareef Ali. 2015. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Real-Time Speech-To-Text. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '15)*, 223–230. <https://doi.org/10.1145/2700648.2809843>
- [34] Raja S. Kushalnagar, Gary W. Behm, Joseph S. Stanislaw, and Vasu Gupta. 2014. Enhancing caption accessibility through simultaneous multimodal information: visual-tactile captions. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility (ASSETS '14)*, 185–192. <https://doi.org/10.1145/2661334.2661381>
- [35] Raja S. Kushalnagar and Christian Vogler. 2020. Teleconference Accessibility and Guidelines for Deaf and Hard of Hearing Users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, 1–6. <https://doi.org/10.1145/3373625.3417299>
- [36] Paddy Ladd. 2005. Deafhood: A concept stressing possibilities, not deficits. *Scandinavian Journal of Public Health* 33, 66\_suppl: 12–17. <https://doi.org/10.1080/14034950510033318>
- [37] Daniel G. Lee, Deborah I. Fels, and John Patrick Udo. 2007. Emotive captioning. *Computers in Entertainment* 5, 2: 11. <https://doi.org/10.1145/1279540.1279551>
- [38] Kim Lyons. 2021. Zoom now has auto-generated captions available for free accounts. *The Verge*. Retrieved September 13, 2022 from <https://www.theverge.com/2021/10/25/22744704/zoom-auto-generated-captions-available-free-accounts-accessibility>
- [39] Kelly Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, John Tang, Andrew Begel, Erin Beneteau, Josh Urban Davis, Abraham Glasser, Joon Sung Park, and Venkatesh Potluri. 2021. Mixed Abilities and Varied Experiences: a group autoethnography of a virtual summer internship. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*, 1–13. <https://doi.org/10.1145/3441852.3471199>
- [40] Kelly Mack, Emma McDonnell, Venkatesh Potluri, Maggie Xu, Jailyn Zabala, Jeffrey Bigham, Jennifer Mankoff, and Cynthia Bennett. 2022. Anticipate and Adjust: Cultivating Access in Human-Centered Methods. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, 1–18. <https://doi.org/10.1145/3491102.3501882>
- [41] James R. Mallory, Michael Stinson, Lisa Elliot, and Donna Easton. 2017. Personal Perspectives on Using Automatic Speech Recognition to Facilitate Communication between Deaf Students and Hearing Customers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 419–421. <https://doi.org/10.1145/3132525.3134779>
- [42] Emma McDonnell. 2022. Understanding Social and Environmental Factors to Enable Collective Access Approaches to the Design of Captioning Technology. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*, 1–8. <https://doi.org/10.1145/3517428.3550417>
- [43] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2: 434:1–434:25. <https://doi.org/10.1145/3479578>
- [44] Mia Mingus. 2011. Changing the Framework: Disability Justice. *Leaving Evidence*. Retrieved March 20, 2020 from <https://leavingevidence.wordpress.com/2011/02/12/changing-the-framework-disability-justice/>
- [45] Mia Mingus. 2017. Access Intimacy, Interdependence and Disability Justice. *Leaving Evidence*. Retrieved February 4, 2020 from <https://leavingevidence.wordpress.com/2017/04/12/access-intimacy-interdependence-and-disability-justice/>
- [46] Michael Oliver. 1983. *Social Work with Disabled People*. Macmillan.
- [47] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 1–10. <https://doi.org/10.1145/3173574.3173867>

- [48] Agnès Piquard-Kipffer, Odile Mella, Jérémy Miranda, Denis Jouvet, and Luiza Orosanu. 2015. Qualitative investigation of the display of speech recognition results for communication with deaf people. In *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 36–41. <https://doi.org/10.18653/v1/W15-5107>
- [49] Anwesha Roy. 2021. How Reliable is Speech-to-Text in 2021? *CX Today*. Retrieved September 13, 2022 from <https://www.cxtoday.com/speech-analytics/how-reliable-is-speech-to-text-in-2021/>
- [50] Jazz Rui Xia Ang, Ping Liu, Emma McDonnell, and Sarah Coppola. 2022. "In this online environment, we're limited": Exploring Inclusive Video Conferencing Design for Signers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22), 1–16. <https://doi.org/10.1145/3491102.3517488>
- [51] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *CoDesign* 4, 1: 5–18. <https://doi.org/10.1080/15710880701875068>
- [52] Matthew Seita, Khaled Albusays, Sushant Kafle, Michael Stinson, and Matt Huenerfauth. 2018. Behavioral Changes in Speakers who are Automatically Captioned in Meetings with Deaf or Hard-of-Hearing Peers. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '18), 68–80. <https://doi.org/10.1145/3234695.3236355>
- [53] Matthew Seita, Sarah Andrew, and Matt Huenerfauth. 2021. Deaf and hard-of-hearing users' preferences for hearing speakers' behavior during technology-mediated in-person and remote conversations. In *Proceedings of the 18th International Web for All Conference*, 1–12. <https://doi.org/10.1145/3430263.3452430>
- [54] Matthew Seita and Matt Huenerfauth. 2020. Deaf Individuals' Views on Speaking Behaviors of Hearing Peers when Using an Automatic Captioning App. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI EA '20), 1–8. <https://doi.org/10.1145/3334480.3383083>
- [55] Matthew Seita, Sooyeon Lee, Sarah Andrew, Kristen Shinohara, and Matt Huenerfauth. 2022. Remotely Co-Designing Features for Communication Applications using Automatic Captioning with Deaf and Hearing Pairs. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22), 1–13. <https://doi.org/10.1145/3491102.3501843>
- [56] Brent N. Shiver and Rosalee J. Wolfe. 2015. Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*, 231–238. <https://doi.org/10.1145/2700648.2809857>
- [57] Sins Invalid. 2019. Skin Tooth and Bone: The Basis of Movement is Our People, a Disability Justice Primer. Sins Invalid.
- [58] John Tang. 2021. Understanding the Telework Experience of People with Disabilities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1: 30:1–30:27. <https://doi.org/10.1145/3449104>
- [59] Máté Ákos Tündik, György Szaszák, G. Gosztolya, and A. Beke. 2018. User-centric Evaluation of Automatic Punctuation in ASR Closed Captioning. In *INTERSPEECH*. <https://doi.org/10.21437/Interspeech.2018-1352>
- [60] Christian Vogler, Paula Tucker, and Norman Williams. 2013. Mixed local and remote participation in teleconferences from a deaf and hard of hearing perspective. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '13), 1–5. <https://doi.org/10.1145/2513383.2517035>
- [61] Emily Q. Wang and Anne Marie Piper. 2018. Accessibility in Action: Co-Located Collaboration among Deaf and Hearing Professionals. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW: 180:1–180:25. <https://doi.org/10.1145/3274449>
- [62] Kenta Yamamoto, Ippei Suzuki, Akihisa Shitara, and Yoichi Ochiai. 2021. See-Through Captions: Real-Time Captioning on Transparent Display for Deaf and Hard-of-Hearing People. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '21), 1–4. <https://doi.org/10.1145/3441852.3476551>
- [63] Matthijs Zwinderman, Rinze Leenheer, Azadeh Shirzad, Nikolay Chupriyanov, Glenn Veugen, Biyong Zhang, and Panos Markopoulos. 2013. Using Video Prototypes for Evaluating Design Concepts with Users: A Comparison to Usability Testing. In *Human-Computer Interaction - INTERACT 2013* (Lecture Notes in Computer Science), 774–781. [https://doi.org/10.1007/978-3-642-40480-1\\_55](https://doi.org/10.1007/978-3-642-40480-1_55)
- [64] 2017. Communicating With Deaf Individuals. *National Deaf Center*. Retrieved September 14, 2022 from <https://www.nationaldeafcenter.org/resource/communicating-deaf-individuals>
- [65] Lost in Transcription: Auto-Captions Often Fall Short on Zoom, Facebook, Google Meet, and YouTube. *Consumer Reports*. Retrieved September 13, 2022 from <https://www.consumerreports.org/disability-rights/auto-captions-often-fall-short-on-zoom-facebook-and-others-a9742392879/>
- [66] Use live captions in a Teams meeting. Retrieved September 13, 2022 from <https://support.microsoft.com/en-us/office/use-live-captions-in-a-teams-meeting-4be2d304-f675-4b57-8347-cbd000a21260>
- [67] Use captions in a meeting - Computer - Google Meet Help. Retrieved September 13, 2022 from [https://support.google.com/meet/answer/9300310?hl=\\$&en&co\\$=\\$GENIE.Platform%3DDesktop](https://support.google.com/meet/answer/9300310?hl=$&en&co$=$GENIE.Platform%3DDesktop)
- [68] Certified Realtime Captioner (CRC) | NCRA. Retrieved July 11, 2022 from <https://www.ncra.org/certification/NCRA-Certifications/certified-realtime-captioner>
- [69] Live Transcribe | Speech to Text App. *Android*. Retrieved September 8, 2022 from <https://www.android.com/accessibility/live-transcribe/>
- [70] DeafSpace. Retrieved September 16, 2020 from <https://www.gallaudet.edu/campus-design-and-planning/deafspace>