Understanding, Designing, and Theorizing Collective Access Approaches to Captioning-Mediated Communication

Emma J. McDonnell

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Leah Findlater, Chair

Daniela Rosner

Jon E. Froehlich

Program Authorized to Offer Degree:
Human Centered Design and Engineering

University of Washington

Abstract

Understanding, Designing, and Theorizing Collective Access Approaches to Captioning-Mediated Communication

Emma J. McDonnell

Chair of the Supervisory Committee:
Leah Findlater
HCDE

For the many people who cannot access audio content, perhaps because they are d/Deaf or hard of hearing, captions are a crucial accessibility tool. While a significant body of work has developed and studied captioning technologies, researchers have traditionally only considered d/Deaf and hard of hearing (DHH) people as captioning users. Yet, communication is inherently interactive, and Deaf and disabled scholars and activists increasingly emphasize that accessibility ought to be a group-level, not individual, concern. Treating DHH people as the sole users of captioning places all of the work of ensuring communication access on the group that faces access barriers. Further, when captioning tools are not designed to also engage conversation partners, a number of avenues to make conversation more accessible cannot be considered. In my dissertation I identify the impact of conversation partners on captioning use and design collective communication access approaches, reimagining how we conceptualize communication access.

My dissertation research uses a range of qualitative, theoretical, and design methods to understand the context that shapes caption use and to envision collective access technologies. I begin by outlining a theoretical framework for collective communication access, drawing from disability studies, Deaf studies, disability justice, and communication studies. I then identify factors that shape DHH people's experiences of real-time captioning in small groups and identify the potential for and interest in group captioning tools. Via a codesign study with mixed hearing ability groups, I identify promising practical directions for the design of collective access captioning tools. I then explore the

role of contextual factors and collective access in a different form of captioning – user-generated captions on TikTok. Finally. I review the past decade of captioning literature through a collective communication access lens, identifying that designing for the group, grounded in communication context, is a novel but promising approach to creating captioning technologies.

My dissertation makes empirical, theoretical, and design contributions, envisioning and grappling with the complexities of designing communication access technologies anchored in Deaf and disability scholarship and activism. I propose a future of accessibility practice that uses technology to guide nondisabled people toward more accessible norms and builds tools that can better match the ways they are used in practice.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

While my name is on this dissertation, it is a product of the care and support I have received from so many people over the last five years. I have been so fortunate to have such a big community, such that I cannot possibly list all of the people who got me here.

I want to begin by thanking my committee. Specifically, Daniela Rosner has generously and incisively challenged me to think critically about power in research processes and who is and is not included, making me a sharper researcher and critic. I have learned so much about designing and sharing research while collaborating with Jon Froehlich, and could not ask for a better cheerleader for my work and career. Mark Harniss has gone above and beyond his role as GSR, bringing a wealth of knowledge about disability studies and how to work interdisciplinarily. Leah Findlater has been exactly the advisor I needed. She has shown me how to balance high expectations of rigorous work with care and understanding for the people you do it with. May all students who find their way into disability during a PhD have the gift of grace and flexibility Leah has given me – thank you.

There are so many others at UW and in my broader academic community who have supported me into this moment. My cohort and peers across HCDE, Computer Science, Disability Studies, CREATE, UAW4121, and beyond have made this a truly collective effort. Specifically, but far from exclusively, thank you Anna Beers, Jay Cunningham, Akeiylah DeWitt, Caitie Lustig, Neilly Tan, Keri Mallari, Beth Dunbar, Sourojit Ghosh, Alainna Brennan Brown, Joey Schafer, McKane Andrus, Joice Tang, Venkatesh Potluri, Maitraye Das, Abir Saha, Tessa Eagle, Ellen Simpson, Francesca Colonnese, Sophie Hurwitz, Annie Ross, Daniel Revier, Taylor Gotfrid, Momona Yamagami, Mia Hoffman, and Calvin Liang. I have had the honor of working closely with my lab mates in the Inclusive Design Lab: Steven Goodman, Lotus Zhang, Gina Clepper, Lucy Jiang, Dhruv Jain, and Abby Stangl. To those who have taken me under their wing and shown me the ropes, especially Emily Wang, Megan Hofmann, and Cindy Bennett, thank you. I am so grateful to those who have

trusted me to mentor them – I learned at least as much while working with Ana Liu, Jazz Ang, Andy Moon, Pitch Sinlapanuntakul, Adrian Rodriguez, Lucy Jiang, and Gina Clepper. Thank you also to my collaborators, without whom I would be a much poorer researcher and far from done with this dissertation. I am grateful to faculty members from UW, Northwestern, and broader accessibility communities who have mentored and supported me, specifically: Joanne Woiak, Heather Evans, Jennifer Mankoff, Raja Kushalnagar, Christian Vogler, Kristen Shinohara, Garreth Tigwell, Katta Spiel, Robin Brewer, Sara Sood, Anne Marie Piper, Jennifer Nash, Darren Gergle, Sean Munson, Julie Kientz, Audrey Desjardins, and Jennifer Turns. Thank you to the administrators without whom my work would not happen: Jane Skau, Allen Lee, Kathleen Rascon, Pat Reilly, Alanna Robins Daniels, Dimitri Azadi, Stacia Green, Kathleen Quinn Voss, and those whose work is instrumental but not visible to me. To the decades of union organizers with UAW4121 who have fought so that I and generations of students to come are protected to do our best work: solidarity forever.

Thank you to all those who have participated in my research studies and to my broader disability communities. My driving force in research is to try to do right by the communities I work with and belong to—thank you for trusting me with your stories.

Part of being able to do this research is being able to be human in the process. I am indebted to the hundreds of books, tens of thousands of minutes of listening to the Mountain Goats on repeat, thousands of crossword puzzles, and the dozens of knitting projects that have gotten me through the last five years. The communities that have welcomed me keep me grounded – thank you to all of the friends and family who have learned perhaps more about a PhD than they bargained for, to the Acorn Street Knit Night, and the Seattle Fusion dance community. Lassie, Bear, and Daisy have been indispensable companions. Thank you to Jesse Martinez, Allie Kerr, and Aashaka Desai for making this place home. To Avery Mack, my partner in crime in and out of work, this could not have happened without you and I am so grateful. To my parents, Sandra and Joel McDonnell, your unconditional love and support from across the country and across the kitchen table are instrumental to everything I do.

Chapter 1

INTRODUCTION

The last five years have been a been a period of significant change for captioning technologies. The potential of automatic captioning had been of interest to researchers (e.g., [45, 93, 144, 27]) and d/Deaf and hard of hearing (DHH) communities [61] for years. Widespread commercial automatic captioning tools began to hit the market in spring 2019 with Google's launch of the Live Transcribe app [104] and addition of automated live captioning to Google Meet [217]. The global shift to videoconferencing in the wake of the COVID 19 pandemic also brought significant shifts in captioning technologies. Google Meet had 3 million daily users at the start of 2020 and made the platform free to all Google users (not just enterprise clients) in spring 2020 [34], the first widely-used free videoconferencing captions. By the start of 2024 Google Meet had 300 million monthly users [75]. Similarly, Live Transcribe, an app developed in collaboration with Gallaudet University to provide live captioning in mobile contexts, had 100 million downloads at the end of 2021 [104] and over one *billion* downloads by the end of 2023 [109]. Google has not been the only player innovating live captioning – Microsoft's Skype and PowerPoint rolled out live captions and subtitles at the end of 2018 [199] and Microsoft Teams launched automatic captioning in summer 2020 [155]. Zoom lagged behind, making automatic captioning available to all users in fall 2021 [306] following critique from disability communities around their prior launch of captioning as a paid feature [80].

Along with an evolution of captioning tools, how people caption and consume caption media has changed significantly over the last five years. Historically, the low quality of user-generated captions has made it hard for caption users to access user-generated video. Deaf YouTuber Rikki Poynter starting the #NoMoreCraptions campaign in late 2016 [76], and both YouTube and TikTok have been thoroughly critiqued for the state of captioning on their platform [173]. However, the state of user-generated captioning is changing. Amid its rise to dominance in user-generated video platforms [35], TikTok released automatic captioning in 2021 [121] and is now often reported on as

a site where captions are common and innovative [260, 57]. Google Chrome also released universal in-browser live automatic captioning in 2021 [187]. Furthermore, young people now use captions at an unprecedented rate: in summer 2023 18-29 year olds were watching TV with subtitles on 63% of the time [46], a trend that was just beginning to be reported on in 2019 [66].

Yet, the rise in automatically generated and user-generated captions is not a universal boon – even in highly controlled, ideal circumstances commercial automatic speech recognition (ASR) tools still reaches an average of 88-95% accuracy [290]. In contrast, to become a certified CART writer, individuals must achieve 96% accuracy at 180 words per minute [215]. There is also significant bias in these algorithms – ASR performs particularly poorly when on the voices of Black men using AAVE [153] and performance drops precipitously between hearing and Deaf speakers [101]. Bias and decreased performance are real and pressing concerns as automatic captioning threatens to replace human services in education [171] and other crucial services. However, as the NCRA[1] only reported 243 newly certified professionals in 2023, tools and guidance around how to navigate the increased availability of automatic captioning are necessary [214].

To understand how the changes in captioning technologies over the last five years change the day-to-day lives of captioning users, let's imagine what a hypothetical DHH college student might have experienced in 2019 (when I started my dissertation research), compared to their experiences in 2024. Take Alicia – she identifies as deaf, was raised by hearing parents, and mainstreamed throughout her K-12 education. She doesn't know ASL, but uses her hearing aids, speechreading, and captioning to communicate. She had real-time, human-generated captions (CART) provided for classes throughout high school but struggled to engage in extracurriculars – even if her school had provided a captioner after school, it would be a logistical mess to set up CART during basketball practice.

Now, imagine it's 2019 and Alicia is a junior in college. She can watch all the Netflix shows she and her friends love on her laptop – Netflix captions their content following a lawsuit by the National Association for the Deaf in 2012 [74]– but her roommate's boyfriend sometimes "jokes" that she watches TV like his grandma. The cooking videos on her Instagram feed are almost never captioned, and the YouTube clips her friends send her are often unintelligible via YouTube's auto-

[1]The certifying association for court reporters and captioners

matic “craptions”. When she goes to class, she has a CART captioner, but last minute meetings with group mates in her many project-based classes are draining. Alicia can’t schedule a captioner for a 6-9 PM working meeting planned that afternoon, so she gets home exhausted from trying to listen to and speechread people who don’t do a great job remembering to turn away from the whiteboard and look at her when they talk. She wants to try an app like LiveTranscribe, but Alicia has an iPhone and it is only released on Android.

What if, instead, Alicia was a junior in college in 2024. She rarely has to ask her friends to turn on captions for movie nights – most of them already have them on. Alicia’s TikTok feed is mostly captioned, and if she really wants to watch an uncaptioned video, Google Chrome’s integrated live captions do a pretty good job. She still uses CART in class but when she and her peers need to work on a project late into the night, often they hop on Zoom, Google Meet, or Facetime (all of which have integrated real-time automatic captioning). They all finished high school online due to COVID – they’re really used to video calls. When Alicia hangs out with friends at a noisy restaurant, she’ll sometimes open Otter.AI to have real-time captioning running, but she’s nervous to do so if she’s with new people. She still runs into access barriers – some of her closest friends are international students and automatic captions often break down on accented English, and she’s had a hard time communicating with new Deaf friends in her disability activist group because none of their voices are well-captioned by automatic tools. Alicia is still drained by trying to figure out if the captions she’s reading are accurate or not.

The evolution of captioning technologies makes a significant difference for someone like Alicia in 2019 compared to 2024. However, communication access has never just been about technology availability or performance. In both cases, social behavior and norms make a significant difference in her experiences, despite rarely being targeted as a way to improve captioning. Consider the different experiences of project meetings in 2019 versus 2024 – the switch toward broad use of videoconferencing tools allows Alicia to work with group mates in an environment that is far friendlier for captioning. Yet, in person conversations in 2024 remain difficult – while Alicia has more tools in her arsenal, they perform poorly in noisy environments and the expectation that she navigates access alone and social anxiety around using assistive technologies limit her options in practice.

1.1 The Current State of HCI Captioning Literature

Human-computer interaction (HCI) captioning research literature has enabled and responded to these technical advances. As automatic captioning has developed, researchers investigated how, if at all, conveying errors to captioning users aids them in interpreting error-laden captions [30, 28] and what impact those errors have on comprehension [145, 144]. A key concern when designing captioning tools is how to place captions in a way that minimizes visual dispersion, often addressed by developing new display configurations [50, 158] or new form factors, such as head-mounted displays (HMDs) [136, 228, 192]. Researchers have also explored how caption placement [8, 12, 10, 9, 41] and formatting [124, 28, 106] impacts how much information caption users can take in. Specific contexts for captioning use have also received significant attention, including classroom captioning [45, 50, 149, 159] and captioning online, user-generated video [175, 181, 258, 31, 260].

Notably, while caption quality has improved, many of the barriers Alicia faced in our hypothetical example came from the social, relational, and environmental factors that shape her daily life. There is a small body of captioning research that has identified the impact conversation partners have on the experience of someone reading captions [254, 256, 255] or identified the role conversation partners could play in developing more accessible communication norms [293, 257, 149, 183]. Traditionally, researchers have not conceived of captioning as a tool that groups use to communicate together – rather it is positioned as a tool DHH individuals[2] independently use to access a conversation. This oversimplification of caption use as a solitary endeavor has limited our capacity to understand the role conversational context has on the use of captioning technologies.

However, Deaf and disabled scholars and activists demonstrate that access ought to be seen as something people do together. Disability justice, a contemporary activist movement led by disabled people who are queer, trans, Black, Indigenous, or otherwise people of color, envisions a liberated, accessible future [205]. Disability justice theorizing centers interdependence, valuing mutual reliance over independence, and envisions collective access approaches, where access is done for and by the entire group, rather than individuals [131]. Disability studies scholars also focus on to target change towards social structures, rather than individuals [221, 26]. Disability studies scholar Devva

[2]captioning is a technology used to increase access for many groups other than DHH people, but that is seldom acknowledged in the captioning literature, for examples see [260, 194]

Kasnitz calls for a shift toward *'community based accommodations*', arguing that, for example, hearing non-signers need an interpreter just as much as their Deaf interlocutors do [148]. Furthermore, Deaf cultural concepts such as DeafSpace [81] and Deaf gain [19] prioritize a world that is built for and centers Deaf communication norms. The HCI research that has translated critical disability perspectives into the design of technology finds that designing for interdependence is a compelling way to better align access technologies with the worlds they're used within [25, 63, 64, 62]. My work brings Deaf and disability studies and activism to the design of captioning technologies, exploring the design ramifications of seeing captioning as something groups use together.

1.2 Research Questions and Thesis

My dissertation research investigates how social, relational, environmental, and technical factors combine to shape people's experiences of novel captioning tools, identifying the need to design future captioning technologies to serve groups, not just individuals. I am guided by the following research questions:

- DRQ1: How do socioenvironmental factors shape the experience and efficacy of caption use?
- DRQ2: How can captioning technologies be designed in a way that fosters collective access?

Over the course of the following chapters, I demonstrate my thesis statement:

Centering social, relational, and environmental factors as key determinants of communication accessibility while designing captioning technologies better matches captioning's realistic context of use and can advance disability and Deaf community politics.

1.3 Dissertation Overview

After situating my research relative to the many fields I draw from, I present the empirical and theoretical work that demonstrates my thesis statement and conclude with a discussion of the broader implications of my research.

In Chapter 2 I introduce and briefly discuss five key bodies of related research that inform my dissertation as a whole. First, I discuss how captioning technologies have evolved and how captions are provisioned, then I summarize the large body of work that has focused on how to best display

captions, followed by an overview of factors considered in the design of captions themselves, ending with an introduction to critical Deaf and disability perspectives.

Chapter 3 presents the first half of a theoretical literature review I conducted, forthcoming at ASSETS 2024 [195]. In this chapter, I develop and outline a theoretical framework for collective communication access. This framework crystallizes the years of thinking and discussion that both informed and were informed by the empirical work I present in this dissertation. I chose to separate the theoretical framework from the literature review it guided (see Chapter 7) to provide greater theoretical clarity on the approach I demonstrate in subsequent chapters. This chapter provides a theoretical lens to guide my answer to DRQ2.

In Chapter 4 I present the foundational empirical study that anchors my dissertation, published at CSCW 2021 [196]. The research in this chapter is informed by the semi-structured interviews and design probe activities I conducted with 15 DHH participants. The two major findings from this chapter are that captioning must be understood in the context of the social, environmental, and technical factors that shape its use and that designing captioning technologies for group use is seen as promising by DHH participants. This chapter directly investigates DRQ1 and motivates DRQ2, contributing to my conclusions that social and environmental factors must be considered in the design of technology and are necessary for understanding captioning's realistic context of use.

Chapter 5 reports on the three-phase codesign study I conducted with mixed hearing ability groups, published at CHI 2023 [198]. After recruiting established small groups with 3-6 members, at least one DHH and at least one hearing (17 participants across 4 groups), we conducted first, a formative session observing participants' current communication practices and introducing the idea of designing technology for group use, second, a design session wherein participants first individually and then collectively sketched their preferred captioning tools, and finally, had participants review video prototypes we developed of their ideas. From this study we document the role relationships play in shaping desired access practices and offer concrete design directions, developed by participants, for group captioning technologies. This chapter further investigates RQ1 and directly takes up RQ2, highlighting the role of relational factors in access practices and demonstrating how design practices grounded in contextual factors can enact a politics of collective communication access.

In Chapter 6 I explore a collective, contextual approach to user-generated captioning, specifically TikTok captions, published at CHI 2024 [194]. While nondisabled conversation participants do not

traditionally engage with real-time captions, content creators on TikTok are directly responsible for the presence and quality of captions on their videos. We conducted a content analysis of 300 TikTok videos, complemented by an interview study with 9 caption users, investigating how captions are created and shared in this specific environment. We found that, while TikTok is largely usable by caption users, future standards and greater knowledge and accountability among caption creators would improve experiences of user-generated captions. This chapter further explores both DRQ1 and DRQ2, articulating how the social and environmental context of TikTok shapes captioning practices and formulating user-generated captioning as a form of collective access.

Chapter 7 contains the final content of my dissertation research – the literature review that accompanies the theoretical framework presented in Chapter 3, published at ASSETS 2024 [195]. In this work I reviewed human-centered, accessibility-focused captioning research published between 2013 and 2023 using my framework for collective communication access. This chapter investigates DRQ1 and DRQ2 in the context of previously published literature, finding that work that engages contextual factors better matches captioning's realistic context of use and advances disability and Deaf politics.

In Chapter 8, I discuss the implications of my dissertation work, specifically considerations for an activist accessibility research practice, tensions in designing for collective communication access, and the future for interdisciplinary disability research. I also identify dissertation-level limitations and promising areas for future work

I conclude in Chapter 9 by revisiting my dissertation level contributions.

1.4 Contributions

At a high level, my dissertation contributes:

- Empirical evidence that social, relational, and environmental context are critical determinants of how captioning technologies mediate access (see Chapters 4, 5, 6, and 7)

- Design guidance identifying 1) practical considerations for captioning technology that shapes behavioral norms, 2) guidance for matching caption design to specific contexts, and 3) a design space of low technical complexity, high social impact captioning interventions (see Chapters 4, 5, and 6)

- A theoretical framework of collective communication access (see Chapters 3 and 7)

1.5 Positionality and Authorship

All of the research I present in this dissertation is work that I led, but none of it is work that I did alone. Collaborators Steven Goodman, Ping Liu, Soo Hyun Moon, Lucy Jiang, Tessa Eagle, Pitch Sinlapanuntakul, Kate Ringland, Raja Kushalnagar, and Jon Froehlich, and my advisor, Leah Findlater, are integral to this research's existence. Therefore, when writing sections that represent only my views, such as this introduction and Chapter 8, I use first-person singular pronouns. Throughout my collaborative work, however, I acknowledge my collaborators through the use of first-person *plural* pronouns.

This work is also deeply shaped by my positionality and politics. I write from a position of great privilege – I am white, with access to a world-class education at an American university, conducted in my first language, English. This power shapes how I interpret the world, interact with participants, and how my scholarship is received and shared.

Throughout my PhD I have lived and conducted my research on the lands of the Coast Salish peoples, specifically those of the Duwamish, Puyallup, Suquamish, Tulalip, and Muckleshoot nations, as well as the lands of the Peoria, Potawatomi, Myaamia, Sioux, Kaskaskia, Kickapoo, Anishinabewaki, and Metis peoples. Please visit `https://native-land.ca/` or a similar website to learn whose land you are on and consider donating to those nations.

My work is also shaped by an evolving disability identity and my perspective as a queer woman. Though I did not identify as disabled when I began my PhD, the last five years have taught me intimately what disability studies scholars know – disability is a category anyone can enter at any time. This experience, coupled with my status as a hearing person working with d/Deaf and hard of hearing communities, has driven me to think deeply about the limits of a politics based on shared identity. I see it as my responsibility as a researcher to learn deeply the politics and culture of a community I work with, advocate for their political goals in and beyond my research, mentor junior scholars who belong to those communities, and to stay critically reflective.

Specifically, while working on a dissertation primarily focused on access for DHH communities, I have learned ASL (reaching a conversational proficiency), studied Deaf culture, history, and

communication practices. I have also had opportunities to be mentored by Deaf experts in this field, to mentor burgeoning DHH scholars, and to work and learn alongside DHH colleagues and friends. I present this all from a place of humility, as a sincere effort to not replicate the significant harm hearing researchers have done to DHH communities.

My disciplinary background also has significant bearing on the work I do. I am graduating from the University of Washington with a Ph.D. and M.S. in Human Centered Design and Engineering as well as a graduate certificate in Disability Studies. I also have a B.S. in Computer Science and undergraduate minor in Gender and Sexuality Studies. I approach the design of technology with a range of critical, theoretical lenses, as a scholar formally trained in both domains. This unique, interdisciplinary education is central to the work I do.

Chapter 2

RELATED WORK

The bodies of work most related to my dissertation focus on how captioning is provisioned, understanding people's experiences using captions, how captions themselves are styled and displayed, critical Deaf and disability perspectives, and disability-studies informed HCI research.

2.1 Captioning Technologies

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 [52]. Captions are used for real-time spoken conversation, live television (e.g., news reporting), and prerecorded media, each format bringing their own accessibility considerations. Here I highlight the two forms of captioning I study in my dissertation work; real-time captioning, and captioning prerecorded media.

The range of available real-time captioning technologies each come with their own tradeoffs. CART—human-generated verbatim captioning—is the most popular and accurate. Certification for CART writers requires they be able to caption speech at 180 words per minute with 96% accuracy [215]. Yet, CART is expensive ($80-$100 an hour) and must be scheduled in advance [2]. An alternative, C-Print, summarizes content within sentences and uses a shorthand style, but is also costly ($60/hour) and must be pre-scheduled [85].

While human transcription remains the legally protected standard for captioning in the US and around the world [15], automatic captioning using automatic speech recognition (ASR) is increasingly used for informal interactions and when accommodations are not otherwise available [7]. Despite years of development, recent evaluations found that on high quality audio sources, popular automatic speech recognition (ASR) engines reach 88-95% accuracy [290] and ranges from 81-86% accuracy in less ideal conditions [244]. Additionally, ASR performance deteriorates in complex audio environments [282] and ASR technologies display algorithmic bias, including per-

forming poorly on Black speakers using African American Vernacular English (AAVE) [153] and speakers with Deaf accents [101]. Identifying when ASR-generated captions are unusably inaccurate poses a difficult problem, partly because long-established metrics do not adequately weight the types of errors that impact DHH viewers [144, 145]. Berke et al. [30] also caution that while some caption readers can indeed make sense of inaccurate captions, the ability to identify and make sense of errors depends on readers' literacy levels. Unlike human transcription, ASR does not convey non-speech information such as laughter or consider high-level context, such as a child trying to say a new word. Furthermore, many within the Deaf community oppose using automatic captioning in place of human transcription, considering it to be insufficient access [76].

When considering pre-recorded media, whether or not a video is mandated to be captioned depends on where it is aired. Closed captioning on American television dates back to the 1970s and became a legal standard [92, 91] via the 1990 American with Disabilities Act (ADA) and 1996 Telecommunications Act [138]. In 2010, the 21st Century Communications and Video Accessibility Act modernized these provisions to require that content aired on television with captions must remain captioned if uploaded online [302]. Legal standards vary internationally [238], changing style guidance on captioning aspects like color (e.g., [22]), and using different frameworks (e.g., W3C consortium guidelines) as the basis of law [84]. Despite some fledgling efforts to mandate captions for user-generated online video, they have not been widely implemented [1]. The rapid development of automatic speech recognition (ASR) has simplified the process of generating video captions and subsequently editing them, making captioning a far easier task [260, 175, 177]. However, automatically generated captions frequently remain unedited, which Deaf activists have highlighted as problematic [76].

2.2 Captioning Experiences

My work builds on the body of research that has sought to understand DHH people's experiences using captioning.

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 [139]. Further, communication issues

can arise from overlapping speech [99, 135], not seeing other speakers [135], ASR not understanding some speakers' accents [101, 153], and background noise [99, 228]. Seita et al. [254] studied automatically captioned small-group conversation between DHH and hearing people, finding that, in the presence of captions, hearing people altered speech characteristics, such as volume and rate. In two studies, they also explored the impact of a hearing researcher modulating their speech in several ways (i.e., speech rate, volume, eye contact), finding that modulations in intonation and enunciation statistically significantly impacted participants' satisfaction with the hearing person's behavior [256, 255].

Videoconferencing environments present unique considerations for DHH users, particularly after the software's surge in use during the COVID-19 pandemic [20]. These platforms pose challenges for most DHH communicators: Ang and Liu et al. [247] highlight limitations for signed communication while Vogler et al. [288] identify significant technical workarounds required for DHH-accessible hybrid meetings. Studies with remote DHH employees highlight high cognitive load and difficulty identifying speakers [183, 278]. To mitigate platform failings, Kushalnagar and Vogler [163] provide practical recommendations for DHH-accessible videoconferencing, including strong conversational guidelines and monitoring chat.

When communicating with hearing people, prior work has identified the rich social context that shapes DHH people's experiences. Kawas et al.'s [149] analysis of real-time captioning in the classroom identifies that most hurdles students face are fundamentally sociotechnical, requiring technological, social, environmental, and policy solutions, Wang and Piper [293], interviewed and observed existing dyads of Deaf and hearing collaborators, focusing on interactions when accommodations are unavailable (i.e., not focused on captions). They found that, over time, these Deaf-hearing teams co-create accessible practices, including flexibly switching between spoken and written language, learning to prioritize shared visuals, and providing ad hoc, informal transcription and sign language interpretation. Seita et al. [257] explored methods for working with DHH/hearing dyads online, finding that dyads must negotiate how to coordinate visual attention to effectively collaborate. Additionally, Elliot et al. [87] tested an ASR/typing based app to facilitate 1:1 communication between DHH/hearing dyads and found initial positive reactions.

2.3 Captioning Styling and Display

While there are standards for captioning styling and placement, many have explored novel approaches in an effort to make captioning more prosodic and usable.

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 [28]. Captioning viewers also appreciate options to customize caption styles [105]. Researchers have explored various techniques, including color, animation, placement, and styling, to convey information such as volume [124, 198], emotion [238, 172, 224], and the quality of sound effects [303]. Butler terms these approaches 'aesthetic' or 'alternative' captions, contrasting them with 'integral' captions that prioritize access [44]. In a series of focus groups, she found that DHH people opposed highly aesthetic captions, but concluded that creative captions that "maintain accessible qualities" could be useful [44]. Research has also shown that more humble interventions, such as using punctuation to indicate pauses in automatic captions [106, 258] can positively impact caption readability.

Research into captioning displays has focused on how to make the visually-demanding task of reading captions easier. A key concern when designing captioning systems is limiting visual attention split, and researchers have explored myriad display configurations to enhance DHH people's ability to read captions while attending to other aspects of conversation. These include integrating captions into the environment, using head-mounted displays, and annotating captions [50, 135, 136, 157, 201, 228, 232, 45]. However, Amin et al. [10] highlight that captions displays must be designed as to not occlude relevant information, such as faces. Differentiating between speakers while using captioning remains a pervasive problem, and researchers have explored various solutions, such as adding animations that point to the current speaker [99, 158], moving captions next to the current speaker [124, 228, 126, 303, 198], and designing graphical displays that use a speaker's image and name for identification [289].

2.4 Critical Deaf and Disability Perspectives

My work build heavily on critical perspectives from Deaf studies, Deaf cultural politics, disability studies, and disability justice activist theorizing. I briefly overview the perspectives I draw on here.

Deaf communities have asserted Deafness as a cultural identity that extends beyond an audiological diagnosis [212, 168], and in doing so develop theoretical approaches to communication that do not center hearing people. Concepts such as Deafhood [166] and Deaf Gain [19] assert that Deaf people's way of being and knowing in the world is productive and valuable, rather than an unfortunate loss. Following this reframing, places such as Gallaudet University have developed architectural practices known as DeafSpace [81], using the built environment to assert that visual communication methods, not auditory ones are prioritized. Deaf and disability studies scholars Robinson and Henner put forth the framework of *'crip linguistics'*, calling on scholars to see Deaf and disabled languaging as not in need of correction and valuing communication in all of its forms [117]. Deaf community practices and scholarship demonstrate the value of non-dominant communication styles and show that a world where hearing communication norms are displaced is possible. I work to bring this sensibility into my exploration of captioning technologies.

Disability justice activists and disability study scholars also offer new ways to theorize and enact access. Turning to the basics of disability studies, using a social model (i.e., disability occurs when societies discriminate based on ability), rather than a medical model (i.e., disability is a fundamental flaw of some bodies) when approaching accessible technology design better identifies the target for change: sociotechnical systems, not disabled people [189]. Following this model shift, Kasnitz [148] proposes working under a model of 'community based accommodations', recognizing, for example, that hearing and Deaf people are equally reliant on an interpreter to communicate. I take up disability justice activists' call to shift toward collective access, or the idea that *"we can share responsibility for our access needs"* [131]. 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 [206]. Interdependence is the idea that everyone relies upon each other and that dependence is not a unique facet of disability [208, 220]. It has begun to be used within HCI accessibility research (e.g., [25, 183, 185]) and creates opportunities to envision access as for communities, rather than individual disabled people.

2.4.1 Critical Perspectives on Disability in Accessibility Research

I am not the first to translate critical perspectives on disability into the design of technology and draw from foundational work introducing critical disability research to HCI and from research that puts these ideas into practice.

Disability studies was first introduced into the space of HCI accessibility research by Mankoff et al. in 2010 [189]. Their paper introduces key disability studies concepts then applies them to assistive technology case studies, illustrating the potential for disability studies to transform current practice. Nearly a decade later, Bennett et al. [25] introduced interdependence to HCI, citing primarily disability activist sources to introduce and expand interdependence as an approach and mode of analysis.

In recent years, accessibility research has seen a significant turn toward disability studies and activist-informed work. For example, critical disability analysis has driven studies of how mixed-ability teams collaborate remotely [183], how blind and sighted teams co-write [63], and novel approaches to data visualization [125]. Critical perspectives on disability have also been central to calls for greater consideration of race in accessibility research [114], identifying ways to make methods more accessible [185], and frameworks for reckoning with and preventing ableist harm in research [297]. Autoethnographies by disabled researchers [186, 122] have also used disability studies and community perspectives to guide and contextualize their findings. Beyond the rise in published work that engages critical perspectives on disability, it has also been a central part of workshop conversations (e.g., [266, 272, 197]). I engage in this tradition of exploring new directions for technology design by beginning with critical disability perspectives.

Chapter 3

A COLLECTIVE COMMUNICATION ACCESS FRAMEWORK

In this chapter, I detail the framework for collective communication access I developed, part of *"Envisioning Collective Communication Access: A Theoretically-Grounded Literature Review of Captioning Literature from 2013-2023"*, coauthored with Leah Findlater and forthcoming at ASSETS '24. In the conference paper, this theoretical framework is accompanied by a literature review of the last decade of captioning research. I chose to separate the framework and review (to come in Chapter 7) as this framework provides greater theoretical context on my entire body of work, but the review discusses several other chapters in this dissertation. Therefore, I bookend my dissertation chapters with work from this singular conference paper.

This chapter articulates the theoretical approach that I have synthesized over the course of my dissertation research and doctoral studies. While not the kind of new theory one would derive from methods such as grounded theory, I argue that by putting disability studies, Deaf studies, disability justice, and communication studies together against the backdrop of communication access technology, I have generated a novel framework. Though I do not directly answer either DRQ1 or DRQ2 in this chapter, it motivates the need to attend to social, relational, and environmental factors and provides the theoretical background motivating the desire to build technologies that foster collective access.

3.1 Motivation

Communication accessibility sits at a unique nexus of related fields—fields that are infrequently used to inform communication technology research. Disability studies provides mechanisms to identify and realign how disability is conceived of and designed for [174, 221, 142, 202]. Deaf studies is deeply engaged in what it means to value and support marginalized communication styles (e.g., [81, 19, 269, 117]). Disability justice activists provide a new vision of how and with whom access ought to be arranged (e.g., [131, 207, 231]). Finally, the communication studies concept

of models of communication provides a more capacious sense of what comprises communication (e.g., [16]). In this chapter, we combine key concepts from these four fields to create a *framework of collective communication access*.

This chapter aims to answer the following question:

1. How can disability studies, Deaf studies, disability justice, and communication studies thinking inform a theoretical framework of communication accessibility?

The framework we develop here echoes across the following chapters in this dissertation, articulating the approach we brought to the empirical studies in chapters 4, 5, and 6 and driving the reflective theoretical review in Chapter 7.

3.2 Theoretical Roots of Collective Communication Access

In this section, we engage key sources of critical thinking around communication and accessibility: disability studies, Deaf studies, disability justice, and communication studies. We chose to engage these bodies of work because they provide foundational theorizing about disability and access (disability studies), specific insight into communication-focused access needs (Deaf studies), cutting edge thinking on access and community (disability justice), and a framework for understanding communication (communication studies). From this scholarship, we produce a framework—collective communication access—to guide the design of communication technologies aligned with critical Deaf and disabled perspectives on communication and access.

3.2.1 Disability Studies

Disability studies is the academic field interested in understanding disability as a social phenomenon. While disability studies often researches and draws from disability activist movements, we discuss disability justice theorizing in subsection 3.2.3.

Foundational Concepts

Disability studies provides several mechanisms for understanding and countering discrimination against disabled people. Lewis defines 'ableism', a *"a system of assigning value to people's bodies*

and minds based on societally constructed ideas of normalcy, productivity, desirability, intelligence, excellence, and fitness"[174]. Ableism emphasizes that discrimination on basis of ability is systematic and pervasive [3]. Another key analytical framework is Oliver's models of disability [221, 220]. The *medical model* describes the view that disability is a problem in individual bodies that ought to be fixed if possible, hidden otherwise, and eventually eliminated. Disability studies scholars propose the *social model* as a way of envisioning a better future [26]. Under the social model, disability is not a product of faulty bodies but is a social arrangement that fails to anticipate the natural diversity of human ability. Therefore, disability could be solved for by changing policy, attitudes, and infrastructure to build an accessible world. Recently, scholars have critiqued the idea that disability would disappear in a fully accessible society [102], proposing new models that maintain a focus on political causes of ableism but have a more embodied understanding of disability (e.g., [142]). Disability studies has also expanded its understanding of the kinds of access needs people have, moving beyond a narrow focus on physical and sensory access infrastructure [230] and developing theories such as crip time—the idea that disability often manifests in a different pace of life [142, 249]—and integrating perspectives from groups with access needs not well-met by current practices (e.g., psychiatric survivors [133], neurodivergent communities [259], chronically ill people [295]).

Disability Studies and Technology

Disability studies scholars argue that we must challenge the exploitation and eugenics that define historical disability technology. Medical model ideologies are often embedded in technologies, ranging from the explicitly eugenic drive to eliminate disability via gene editing and selective abortion [251] to the more subtle belief that disability is a shameful, private thing, which is translated into the design of 'discreet' assistive technologies [287]. Mills argues that there is a long history of technology development under an *"assistive pretext"*, wherein technologies (e.g., sound recording and visualization technologies) are developed with the stated purpose of improving access for a given group (e.g., DHH people), but that once developers identify a more lucrative general purpose for a tool, the disability application is abandoned [202]. Additionally, Hamraie highlights that while accessibility efforts have historically been central to universal and inclusive design movements, too often projects that claim to serve *all* users simply better obscure the many people they exclude

[110, 111].

Yet, activists have also identified ways that technology can advance disabled people's lives. The disability rights movement often saw creating accessible technologies as a key strategy for expanding rights [230]. Hamraie and Fritsch [112] propose 'crip technoscience', a framework for designing, implementing, and critiquing disability technologies that is aligned with disability justice principles. Crip technoscience elevates the design work disabled people do to survive in an inaccessible world and calls for an accessibility practice that sees design as a way to enact a political vision and uses technology to disrupt the status quo. This paradigm provides a way to think critically about the differences that exist between technologies that all fall under the umbrella of assistive or accessible.

From this review of disability studies thought, we identify the following imperatives that can guide the design of accessible communication technologies:

1. Interventions should work to change built environments and social worlds to encompass the current practices of disabled people, rather than supporting disabled people in assimilating to nondisabled habits

2. Accessible technologies ought to embody an actively anti-ableist politic, as doing otherwise risks reifying existing systems of oppression

3.2.2 Deaf Studies

Deaf studies is the academic field that studies and theorizes about Deaf people's lives and is highly intertwined with Deaf history, culture and signed languages [164]. Deaf studies research tracks how the Deaf community has come to proudly claim a cultural identity in the face of audism, or systemic discrimination on basis of hearing ability [18].

Foundational Concepts

Deaf studies scholars focus on access to language and the harms of language deprivation. Contemporary audism has its roots in the oralist movement of the late 19th and early 20th century, which saw international opposition to teaching Deaf people signed languages [211], causing irreparable

damage to sign language use and acquisition [242]. Sign languages were not recognized as full languages until the 1960s [269], and the fight to educate DHH people using signed languages is ongoing [271]. In the face of this history, Henner and Robinson propose *crip linguistics* [117], emphasizing that no way of communicating is disordered or wrong. Crip linguistics argues that all forms of communication ought to be respected because efforts to normalize language are steeped in ableism, audism, racism, and many other forms of prejudice.

Responding to centuries of discrimination, Deaf communities have built a strong culture of self-determination and valuing Deafness. A key example of the move to replace hearing paternalism with Deaf self-determination is the 1988 *Deaf President Now* movement. Gallaudet students protested until the Deaf-serving university hired its first Deaf president, setting precedent for decades of Deaf leadership on issues that impact the Deaf community [79]. Many people identify as culturally Deaf, seeing membership in the Deaf community as akin to a linguistic minority, rather than a disability [168, 271]. Scholars increasingly question the impulse to separate Deafness from disability, noting its ableist and eugenic roots [242], and movements such as disability justice have taken a nuanced approach to building coalition with Deaf communities [131, 243]. Deaf studies scholars counter deficit approaches to Deafness via theoretical frameworks [19, 166, 168]. For example, Bauman and Murray's concept of Deaf gain [19], highlights the skills and community that Deaf people gain, countering narratives of hearing *loss*.

Deaf Studies and Technology

Deaf studies and Deaf culture have a contested relationship with technology developments 'for' the Deaf community. Technology was a core component of the oralist movement, and the modern inconspicuous, unobtrusive design of assistive listening devices stems from a belief that Deafness is undesirable and that emulating the behavior of a hearing person is ideal [287]. Cochlear implants, surgically implanted devices that promise better performance than hearing aids, have been highly controversial within the Deaf community, particularly as they are now approved to be implanted in young children who do not have the agency to consent to the procedure [54]. In contrast, Deaf-led design practices affirm Deaf ways of being in the world. For example, the architectural approach termed 'DeafSpace', pioneered at Gallaudet University, has guided the principled redesign

of campus spaces to support signed communication (e.g., prioritizing clear sight lines, choosing high contrast, low eye-strain paint colors, optimizing for natural light) [81].

A Deaf studies informed approach to communication accessibility requires understanding:

1. It is unethical to attempt to correct communication to a hearing and spoken norm

2. Deaf communities ought to be centered as leaders in technology design, as they are already experienced in designing Deaf worlds and there is a long history of destruction by hearing technologists

3.2.3 Disability Justice

Disability justice is an activist movement led by disabled people who are queer, trans, Black, indigenous, and/or people of color, building from and critiquing the disability rights movement of the 20th century. Rather than centering legal frameworks and emphasizing the needs of white, physically disabled men, disability justice focuses on intersectional cultural change, believing that the state will, ultimately, not save disabled people [131]. In organizing for a new world, disability justice activists, often working through the performance collective Sins Invalid[1], have done significant theoretical work in naming current practices within disability communities and imagining a new paradigm for disabled futures.

Principles of Disability Justice

Of the ten key principles of disability justice [130], two are particularly relevant to the design of accessible communication technologies: interdependence and collective access.

Theorists of interdependence emphasize that all people are dependent upon each other and that viewing dependence as characteristic of disability is a tactic of ableism [207]. Mingus argues that valuing and designing for mutual reliance is key to *"practic[ing] an accessibility that moves us closer to justice, not just inclusion or diversity"* [207]. Within HCI, Bennett et al. [25] emphasize that interdependent design challenges the idea that nondisabled people's approaches are necessarily

[1] https://www.sinsinvalid.org/

better and ought to be the default. Valuing interdependence can redirect the focus of design from independent solutions toward technologies people use together.

Collective access provides a framework for putting interdependence into practice. Sins Invalid defines what it means to take a collective access approach: *"access needs can be articulated and met privately, through a collective, or in community, depending upon an individual's needs, desires, and the capacity of the group"* [131]. In essence, while access needs are sometimes best met privately, groups should also seek ways to build access into their core practices. Disability justice organizers highlight how organizing collective access allows for a pragmatic balancing of access capacity and makes the experience of needing to negotiate accessibility feel communal, rather than isolating [231, 205]. The invitation to think about access as something that groups have a shared stake in—rather than something that is provided for an individual—produces new imagined users of accessible technologies.

Nuanced Understandings of Access

Disability justice thinking also dwells in the nuances inherent to the process of creating access in a group. Mingus coined the term 'access intimacy', described as *"that elusive, hard to describe feeling when someone else 'gets' your access needs"*. Mingus reflects that she, and the broader disability community have *"experienced access that has left us feeling like a burden, violated, or just plain shitty"*, but that there is an alternative—a way of negotiating access that feels freeing—that requires trust and mutual understanding to experience [206]. Access intimacy can be a key lens for understanding where collective access approaches succeed or fail. Yet, Piepzna Samarasinha reminds that the impulse to point to interdependence, access intimacy, or other forms of care as a panacea to address all forms of ableism, while understandable, is a naive view [167, 231]. It takes skill and a deliberate unlearning of ingrained social norms to make interdependent, collective access a viable approach—that nuance must also inform the design of future technologies.

While disability justice goals can never fully be achieved in the hegemonic space of the academy, the following imperatives can guide more just technology research:

1. Access is something that happens between people, and access solutions can be collective and interdependent

2. Collective access approaches rely on a deep attention to a group's specific context and broader power dynamics, and cannot be understood or produced in a vacuum

3.2.4 Communication Studies

Finally, though not focused specifically on accessibility, we draw on a foundational concept from communication studies—models of communication—to add nuance to how we discuss communication.

Models of Communication

Interpersonal communications scholars have created theoretical models of communication to understand and analyze instances of conversation. The initial model of communication, the linear model of communication, focused on the actions of a sender, communicating a message to a receiver. While the linear model conceptualizes asynchronous communication (e.g, email) well, scholars quickly realized that it was deeply inadequate for understanding synchronous communication [298]. After several intermediate iterations, Barnlund proposed the *transactional model of communication*. Barnlund begins with the premise that people cocreate meaning through communication and that communication is dynamic, continuous, circular, unrepeatable, and complex [16]. Central to the transactional model is the notion that interpersonal communication is not simply a process of trading information, but a complex, situated act that is fundamentally shaped by the interlocutors' social, relational, and cultural contexts [140]. Therefore, to study how people are communicating, it is inadequate to study only one party or assume that communication approaches are independent of communication partner—communication is highly contextual and all parties co-create meaning together. Bringing the transactional model of communication to the design of communication technologies, we see that conversational partners are a key stakeholder in determining the kind of communication that accessible technologies help mediate. Further, communicators' instinctual attunement to their interlocutors could be leveraged to improve communication access. Notably, these perspectives focus on the role of context and shared efforts to understand and be understood between interlocutors. Within communication studies, Hickman has highlighted the often invisibilized access work that goes into providing CART transcription [120]. Yet, the formal role of paid access provider differs

significantly from access efforts done by direct interlocutors, the focus of this framework.

The foundations of communication studies remind that:

1. The context communication occurs within and relationships between interlocutors fundamentally shape how people communicate

2. We should study and build access technologies that engage all communication participants

3.2.5 Collective Communication Access: A Theoretical Framework

Combined, these bodies of theory provide a comprehensive way of thinking about communication and access that can reshape how researchers approach technology design, which we articulate as a **framework for collective communication access.** We present this framework as both a practical set of considerations and a call to action, inviting future researchers to ground their communication access technology research in critical perspectives on disability and communication. This framework is oriented toward technology and intended to be useful in analyzing and guiding human-centered accessibility research, but draws entirely from other fields. The imperatives we derive from disability studies, Deaf studies, disability justice, and communication studies have significant overlap, and therefore we synthesize these imperatives into five key tenets of collective communication access, as follows:

1. **Research approaches access as a collective process, where all communicators engage in crafting access.** We combine both imperatives from disability justice and both imperatives from communication studies to highlight that communication is definitionally collective and, therefore, collective access approaches are particularly well-suited to communication access. When we design access technologies *only* for DHH communicators, we make the task of providing communication access harder than it needs to be by not considering key stakeholders – conversation partners.

2. **Interventions prioritize and protect Deaf and disabled people's norms, targeting change toward dominant social worlds and technical infrastructures.** Synthesizing imperative one from disability studies and both imperatives from Deaf studies, we highlight that the goal

of interventionist research into communication access ought to be to change social norms and technical infrastructures to prioritize Deaf and disabled people's existing communication styles, given the history of technical interventions that aim to redirect Deaf and disabled people toward nondisabled norms.

3. **Research sets out to counteract historical harms.** We combine our second imperatives from disability studies, Deaf studies, and disability justice—all emphasize that systemic discrimination impacts Deaf and disabled people's ability to access conversations, and that conversations about accessibility are deeply political. To conduct collective communication access technology research and not perpetuate the history of technology's harm toward Deaf and disabled communities, research should be *explicitly* oriented toward dismantling harm.

4. **Research actively centers the knowledge and expertise of impacted communities.** Imperative one from disability studies, impreative one from Deaf studies, and both imperatives from disability justice stress that the expertise of communities who face communication access barriers must be centered. While this may seem counterintuitive to a focus on collective access, we stress that an anchoring in Deaf and disabled people's expertise is crucial to ensuring that technologies support access practices that are substantive rather than performative.

5. **Research considers the role of relational, social, and environmental context in shaping access practices.** We combine imperative one from disability studies, both imperatives from disability justice, and both imperatives from communication studies to highlight the role that context plays in determining how accessible communication can be. Researchers need to study communication as situated and determined by its relational, social, and environmental context.

3.2.6 Conclusion

In this chapter, I provided brief introductions to four key bodies of theory that guide my work: disability studies, Deaf studies, disability justice, and communication studies. From each, I identify two key imperatives relevant to the design of communication access technologies. Given the signifi-

cant overlap between fields, I synthesize these imperatives into a theoretical framework of collective communication access with five key tenets.

Within-Dissertation Context

This framework articulates how I think about technology and accessibility throughout my dissertation research. At the beginning of my dissertation, I was still searching for the sophistication of language and theory to be able to articulate the ideas presented in this chapter, but the threads of this thinking are present throughout the following chapters. In this chapter I also clearly articulate the disability and Deaf community politics I aim to advance through the design of collective communication access technologies.

Chapter 4

SOCIAL, ENVIRONMENTAL, AND TECHNICAL: FACTORS AT PLAY IN THE CURRENT USE AND FUTURE DESIGN OF SMALL-GROUP CAPTIONING

This chapter contains work that was published at CSCW 2021, and titled *"Social Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning."* I developed and led this work, with collaborators Ping Liu, Steven Goodman, Raja Kushalnagar, Jon Froehlich, and Leah Findlater. This work was awarded an honorable mention and recognition for diversity and inclusion.

Within the broader scope of my dissertation, this chapter serves as the foundational empirical work establishing how DHH captioning users perceive the impact of their conversation partners and identifying interest in captioning technologies designed to engage hearing people in adopting more accessible behaviors. It provides a direct answer to DRQ1 and identifies priorities to explore in answering DRQ2.

4.1 Motivation

Real-time captioning provides vital spoken conversation access for many d/Deaf and hard of hearing (DHH) people. Both human-generated and automatic captions have received substantial attention from HCI and CSCW researchers, with a focus on captioning in classrooms or other formal environments (e.g., [45, 149, 169]). In contrast, captioning for more informal small-group and one-on-one interactions has received less attention, despite the fact that human captioning's high cost ($60-$200 an hour in the US) and need for advance scheduling create significant barriers in this context [2]. While automatic captioning is an increasingly viable alternative to human captioning [165, 241], its error rate varies from 9-37% across tools [165]. Unlike human captioners, automated techniques cannot convey non-speech context (i.e., visual references, emotion, emphasis) nor can they intervene to improve communication.

Moreover, small-group conversation's interactive nature, flexible social dynamics, and high level

of spontaneity further limit existing captioning services. Ultimately, captioning and other access tools (e.g., paper and pen, texting, notes apps) all come with limitations and do not fully support DHH people during small-group conversations [86].

Despite the sociotechnical nature of small-group captioning, most prior work has only examined technical considerations, such as how to convey uncertainty in automatic captioning through the use of simulated conversation in controlled experiments [29, 232, 239]. Seita et al. offer exceptions that explore how social interactions and behaviors impact captioning [254, 256, 255], providing quantitative evidence that hearing people's behaviors impact DHH people's experiences of one-on-one captioned and interpreted conversations. These findings motivate the need to more deeply understand DHH people's small group captioning experiences through a sociotechnical lens.

In this chapter we address the questions:

1. What social, environmental, and technical factors impact the use and usefulness of captioning in small groups?

2. What opportunities exist to design captions and caption displays in ways that support more accessible group communication practices?

To begin addressing these questions and to ground future small group captioning technologies in the needs and desires of DHH people, we conducted an interview and design probe study with 15 DHH participants.

Our findings highlight the myriad social (e.g., group norms, preferred communication modes), environmental (e.g., furniture configuration, online availability of a text chat), and technical (e.g., caption lag, built-in speaker identification) factors that shape real-time captioning, contributing an understanding of the context that surrounds captioned conversation. Particularly, we find that: (1) captioning's efficacy is highly determined by the group being captioned, (2) current captioning tools are often insufficient during interactive conversation, and (3) while the lack of visual and spatial information online create barriers, features of video conferencing also provide new opportunities to increase access. Participants' responses to the design probe activity also highlight the potential to create more captioning-friendly environments, both online and in person, and suggest that providing conversation feedback and warnings to guide captioning-friendly group norms is a promising

direction for future development. Based on these findings, we discuss the need to consider the intersection of social, environmental, and technical factors in captioning research, propose a reframing of captioning as a group technology, and put forth future design guidelines that center DHH peoples' needs.

In this chapter, we contribute (1) an empirical account of DHH participants' experiences of small-group captioning which highlights how social, environmental, and technical factors impact its use and efficacy, (2) an exploration of design opportunities to support small-group captioned conversations and future design guidelines, (3) an understanding for both (1) and (2) of how online environments—a historically little-studied captioning context—shape captioning experiences and preferences, and (4) reflections on reframing captioning as a group technology.

4.2 Method

To understand DHH people's experiences using captions in small-group scenarios and their preferences for future captioning systems, we conducted individual qualitative study sessions with 15 DHH participants. We intentionally recruited only DHH participants for this research because our study design is shaped by a commitment to placing the power to shape design recommendations for future captioning tools in the experiences, desires, and needs of the DHH community. The study was conducted remotely via videoconferencing and had three components: a pre-session survey, a semi-structured interview, and a design activity.

4.2.1 Participants

Participants were recruited via email lists at two U.S. universities, social media, and snowball sampling. We required that participants be 18 years or older, able to participate in a Zoom call, self-identify as d/Deaf or hard of hearing, and frequently use real-time captioning—either automated or via a human transcriptionist—for conversation access. We recruited 15 DHH participants (4 men, 10 women, 1 non-binary person), a sample size in line with community norms for similar studies and appropriate for reflexive thematic analysis [39, 37, 47]. On average, participants were 44.8 years old (SD=17.9, range=21-71)—see Table 4.1. Participants had a wide range of "preferred communication methods": sign language (60.0%), oral (40.0%), and written (46.7%) communication (par-

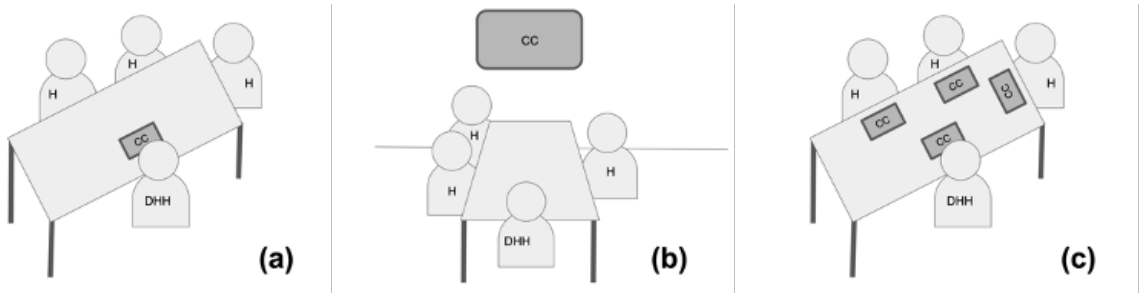


Figure 4.1: Participants discussed captioning displayed for only DHH people via personal devices (a), or for all meeting attendees via a large screen/projector (b) or personal devices (c).

ticipants could select multiple communication preferences). They also had differing experiences with spoken conversations, with participants communicating orally all the time (26.7%), most of the time (20.0%), some of the time (26.7%), infrequently (6.7%), and never (20.0%). Frequency of captioning use ranged from multiple times a day (46.7%) to 2-3 times a month (13.3%).

4.2.2 Procedure

All interviews were conducted remotely due to the COVID-19 pandemic during the summer of 2020. Prior to meeting with researchers, participants completed a 20-minute online survey providing demographic information, background on their experience with captioning, context on how they access captioning services, and their perspectives on the technical, environmental, and social factors that impact their experiences using captioning. The study session took 90 minutes, beginning with a semi-structured interview (35 minutes) followed by a set of design activities (55 minutes). All study sessions were conducted by the hearing first author and facilitated via participants' preferred accommodations: eight participants chose ASL interpreters, five chose CART, one chose automatic captioning, and one chose neither interpreting nor captioning. The researcher screen-shared a slide deck with the study instructions, questions, and design probes, both to be able to discuss design ideas remotely and to allow for multiple ways to access study materials.

The semi-structured interview focused on how participants use captioning in their daily lives and how social factors shape their experiences. Questions covered experiences with different captioning services (e.g., CART, automatic captioning), when captioning works well or poorly, when cap-

Table 4.1: Summary of participant demographics, as reported in the pre-study session survey. P15 chose not to disclose her age

ID	Age	Gender	Identity	Preferred Method of Communication	Frequency of Captioning Use	Frequency of Oral Communication
P1	54	M	deaf, having hearing loss	Oral, Written	A few times a week	All the time
P2	26	F	deaf, Hard of Hearing	Sign, Oral, Written	A few times a week	Most of the time
P3	44	F	Deaf	Sign	2-3 times a month	Never
P4	34	F	Deaf	Sign, Written	Multiple times a day	Never
P5	71	F	deaf	Written	Multiple times a day	Some of the time
P6	24	F	Deaf	Sign	Multiple times a day	Some of the time
P7	47	M	Deaf	Sign	Multiple times a day	Most of the time
P8	30	F	Deaf	Sign	Multiple times a day	Some of the time
P9	69	M	Deaf	Oral	About once a day	Most of the time
P10	53	F	Deaf	Sign, Written, Texting	About once a day	Never
P11	70	F	Hard of Hearing	Oral, Written	Multiple times a day	All the time
P12	21	NB	deaf	Oral	Multiple times a day	All the time
P13	56	F	Deaf	Sign	About once a day	Some of the time
P14	28	M	deaf	Sign	Multiple times a day	Infrequently
P15	–	F	Deaf	Oral	2-3 times a month	All the time

tions are unavailable but would be helpful, how hearing people help or hinder captioning, and how comfortable they are asking hearing people to adopt new communication practices. Throughout, the researcher asked participants to reflect on captioning use both in person and remotely. Following the interview, the researcher facilitated a design probe activity with each participant, inspired by the use of this method in other papers, including accessibility work with DHH participants [96, 103, 135]. Design probe investigations afford light-weight investigation of future technologies and allow researchers to get participant input before committing to a specific design [129], making this method well-suited to our research questions.

The activity included three sets of probes, which were designed to act as a starting point for discussion about future captioning setups, including ideating on potential new features and caption correction systems to be used during small-group captioned conversation. Specifically, we introduced the design probe activity by asking participants *"to try to envision captioning in the future"* and clarifying that *"we don't have to be limited to how technology currently works."* Throughout, we grounded the discussion in the context of being the sole DHH person *"using automatic captioning during an in-person meeting with a small group of hearing people."* We also asked participants to contrast their in-person responses to online contexts. This activity included probes exploring the following:

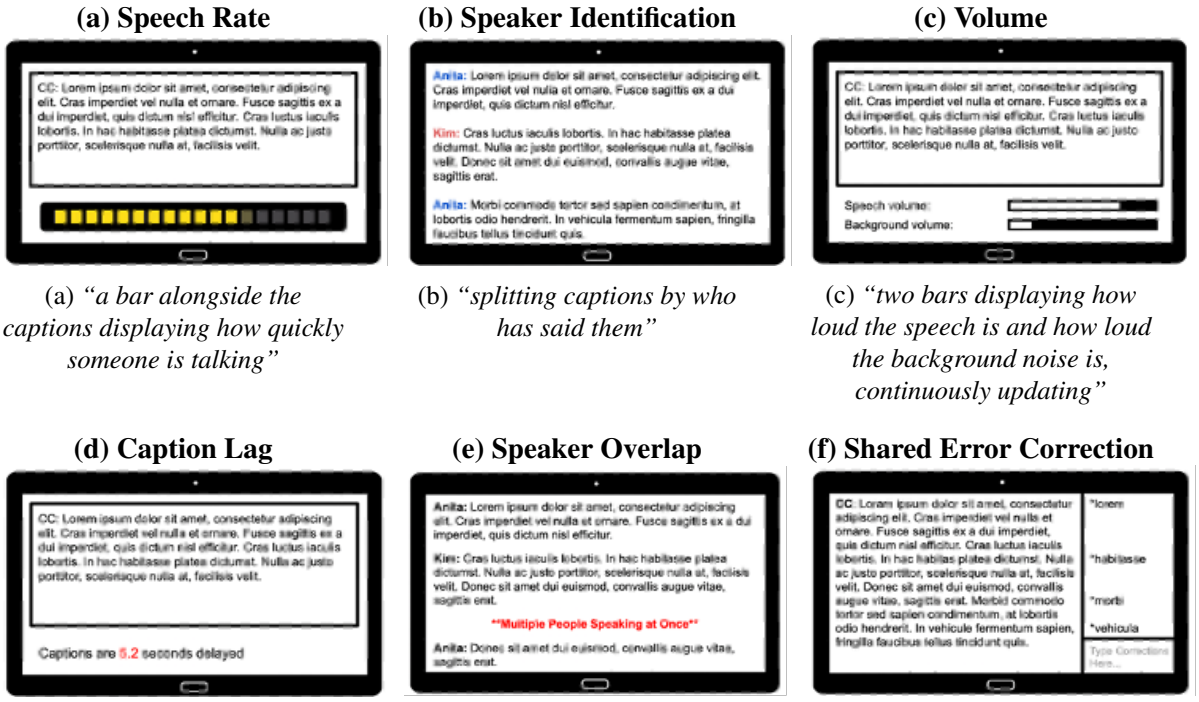
- **Caption visibility.** The first probe centered on participants' preferred method for viewing captions. To introduce this probe, we asked participants to describe their ideal captioning set up for the scenario, prompted with *"for example, you might think about the room setup, where the captions are displayed, who sees them, and anything else."* We then showed captioning setups that varied in who could see them: captioning on the personal device of only the DHH person, captioning on a large screen/projector that all meeting attendees could see, and captioning on the personal devices of all meeting attendees (see Figure 4.1). We had participants consider the advantages and disadvantages of each setup, and how they would feel about analogous setups for remote meetings.

- **Additional features.** The second probe focused on adding information to the captioning display. Following a general introduction to this focus, we described five potential display

features: speech rate (Figure 4.2a), speaker identity (Figure 4.2b), volume of a speaker's voice and of background noise (Figure 4.2c), caption lag (Figure 4.2e), and a multiple concurrent speaker warning (Figure 4.2e). We selected each probe based on current captioning practices as of June 2020, prior work, and knowledge from our team of Deaf and hearing researchers: human captioners often convey speaker identity and overlap, our team identified lag as a significant consideration during small-group captioned conversation, and prior work has identified speech rate [139, 256] and volume [124, 256] as of interest. For each feature, the researcher first introduced the idea (e.g., *"Speech Rate: Show how fast the speaker (you or others) is talking"*) and asked participants what they thought about showing this information in some way during in-person meetings. To make the idea more concrete, the researcher then showed a specific design mockup (Figure 4.2), which we described as a *"rough example of how this feedback could look,"* and asked participants: (1) to share any other ideas they had about how the information could be shown, (2) who (if anyone) they would want to see that information, and (3) how they would feel about this type of information being included in online vs. in-person meetings. After viewing all five sound qualities, participants ranked them and had the opportunity to suggest other information. Note that when creating the design mockups, we opted to display information directly rather than via abstraction (e.g., showing caption lag in seconds delayed rather than as a warning to wait for captions to catch up), to act only as a starting point for discussing how the information could ultimately be displayed.

- **Caption corrections.** The third probe focused on allowing meeting attendees to correct captioning errors in real time. To elicit conversation around this idea, we first introduced the concept and discussed it in the abstract before showing a mockup of a system where meeting attendees could type corrections for errors they notice (Figure 4.2f). We intentionally kept the specifics of this mockup vague to explore how participants would imagine such a system could work. We asked participants for feedback on the idea for both in-person and online meetings.

Finally, participants sketched out their ideal captioning setup for both in-person and online meetings using a pen and paper. Participants shared their sketches by holding them up to the camera, describing them, and sharing ideas they could not capture in the drawing itself. To close the session,



(a) *"a bar alongside the captions displaying how quickly someone is talking"*

(b) *"splitting captions by who has said them"*

(c) *"two bars displaying how loud the speech is and how loud the background noise is, continuously updating"*

(d) *"a sentence informing how delayed captions are, updating as people keep speaking"*

(e) *"having a warning pop up whenever the system detects multiple speakers"*

(f) *"people could type corrections into a chat box that would show up alongside the captions"*

Figure 4.2: As a design probe, participants were shown mockups that we described as *"a rough idea of how"* each potential feature could be implemented: (a) speech rate, (b) speaker identification, (c) volume, (d) caption lag, (e) speaker overlap warning, and (f) shared error correction.

participants were given the opportunity to ask any questions, were compensated with a $50 Amazon gift card, and were asked to email their sketches to the researcher.

4.2.3 Analysis and Positionality

All interview data was transcribed, either directly using the CART transcripts for sessions where a human captioner was able to caption both the researcher and participant (N=4/15) or post hoc by either the first author or a transcription service. We analyzed the transcripts using reflexive thematic analysis as outlined by Braun and Clarke [37, 38], in combination with a summary of participants'

reactions to new designs, which helps identify cross-cutting themes as well as synthesize concrete design recommendations. Our thematic analysis is semantic and realist, with a mixed inductive and deductive approach to the data. Thematic analysis emphasizes that findings are not waiting to be discovered but are actively shaped by the research team and their own biases. The first author, who ran all interviews and led analysis, is hearing and an ASL beginner. Some authors, who were involved in study design, analysis, and writing, are Deaf.

4.3 Findings

Guided by our research questions, we (1) highlight themes we identified in participants' experiences of small group captioning, emphasizing the influence of technical, environmental, and social factors, and (2) report on participants' reaction to our design probes, providing design considerations around better supporting DHH captioning users and engaging their hearing interlocutors in making small-group conversation more accessible. As the study took place during the COVID-19 pandemic, we report on experiences using captioning for online and in-person small-group conversation.

4.3.1 Current Experiences of Captioning

To understand how intersecting social, technical, and environmental factors shape DHH people's use of captioning, our analysis highlights: (1) the role of interlocutors in making conversations (in)accessible, (2) mismatches between the capacity of current technology and the demands of interactive small-group discussion, and (3) specific considerations for captioning online conversation.

Social Impacts on Captioning.

Our participants described their experiences of captioning as highly determined by social dynamics: some groups develop collective, adaptive norms around captioning, while others may limit conversation access through unwittingly inaccessible behavior or explicit judgement of DHH captioning users.

Adaptive group practices. Both during informal interactions with automatic captioning and formal meetings with CART, participants described the benefits of hearing collaborators' willingness to alter communication styles. Participants who prefer not to voice (8 of 15 participants) explained that

to use captions during interactive conversation they write out their contributions while the person they are communicating with is captioned, often using speech-to-text on the DHH person's phone. While transitioning between typing and captioning has a different rhythm than most spoken conversation, P10 found that hearing interlocutors *"get more to the point rather than wandering [in their] speech. I think that's an advantage."*

Participants who voice also benefited from interlocutors who were willing to change norms during spoken conversation. P1 found that his manager, who is *"extremely sensitive"* to the access gaps that persist while using CART, has made work meetings more accessible by taking advantage of online environments to correct captions in the meeting chat and intentionally pausing between topics so that he has a chance to jump in, socially adjusting for the technical limitations of captioning. P2's workplace has even more extensive group practices—to avoid interrupting speakers, they use a set of hand gestures to communicate when to slow down, speak up, or spell out uncommon words. P2 explains that her colleagues do so because having *"captioning available [is] not always going to be enough for someone. Your culture needs to change in order for the captioner to be more effective."* When all parties in a conversation are willing to adopt new social norms around communication, they create a distinct, more accessible solution.

Unsupportive communicators. Participants further highlighted the importance of behavior when describing interlocutors who actively or unwittingly made conversations less accessible. Several participants mentioned disengaging from conversations when hearing people speak over one another or speak too quickly to be captioned. For example, during P12's discussion-based classes, *"there have been a couple times where it's just been like, I don't understand this conversation so I'm just going to go home and wait for the transcript."* Participants also described how moments of acute judgement from others altered how they felt about using captioning tools thereafter. For instance, P11 had to verbally communicate with her notetaker during a group meeting and *"somebody stood up once and said, 'why are they in the corner talking?' I had to say because I can't hear. And it was just like, why was that even necessary? So after that, I just kind of wanted to do my own thing."* While P11's colleague may have been oblivious to the ramifications of their comment, P15 described the impact of active judgement and rigidity around conversation norms. She now joins work meetings via text relay, despite preferring to voice and having tendonitis in her hands, because *"my manager doesn't like to hear my voice. She told me the coworkers said I talk loud. [. . .] It*

makes me feel insecure." Though participants described benefits when hearing people figured out how to communicate accessibly, others explained that sometimes hearing people's instincts, such as slowing down, speaking loudly, or overenunciating, are not effective. For example, P12 commented that these adaptations can be *"done with good intentions but that's not always helpful [because] it makes me feel like you're not treating me as an equal sometimes."* P4 explains that in her experience, hearing people *"do care, it's just that they don't necessarily think about deaf people,"* which means that the burden, or *"constant scourge,"* of creating a captioning-friendly environment typically falls on her.

Technical Considerations While Captioning Interactive Conversation.

Regardless of social support, captioning experiences differ with interactive conversation as compared to one-way communication (e.g., a lecture or seminar). Participants describe how technical aspects of captioning are ill-suited to the particular social dynamics of interactive conversation and the ways they use captioning as one of many access strategies.

Technical mismatches in interactive contexts. Our participants outlined aspects of interactive conversation that are not well-supported by real-time captioning. Participating in small-group interactive conversations requires being able to jump in during brief pauses and P1 explains that the delay inherent to captioning makes this difficult: *"everyone's still talking according to the screen, but they have finished. There's probably like an eight second lag. And sometimes I'm really anxious to say something or correct somebody and then I find that I'm interrupting somebody."* On top of temporal mismatches, captioning does not capture the speaker's tone, which P9 considered invaluable to avoid interrupting: *"I'm assuming some intonation [but] there is nothing on their face that indicates that they are going to complete their sentence."* Furthermore, while P2 found she could engage in captioned conversation with a small number of people, *"if the group gets bigger and other people are talking at the same time it's really hard to follow a conversation and it's also just as hard for the captioner."*

For participants who preferred not to voice, captioning alone does not adequately support interactive conversation. While captions worked when P14 did not need to reply, he explained that *"when I'm trying to say something, captioning doesn't really function for me in that capacity at*

all." However, as automatic captioning has become more widespread, participants who would otherwise use interpreters reported experimenting with the technology. When P4's workplace suddenly shifted online due to COVID, delays with remote sign language interpreting services caused her to join meetings using automatic captions and text chat. P4 explained that, while not ideal, she was *"happy [she'd] found more than one solution,"* one that was enabled by the online environment, the technical capacity to turn on automatic captions, and social expectations that her typed contributions would be integrated into work meetings.

Concurrent access strategies. To manage the limitations of captioning for interactive conversation, participants often used other communication strategies in tandem. Some participants could mostly follow a conversation using their speechreading skills, residual hearing, and assistive listening devices, and they described using captions to augment their understanding, rather than as the primary way of accessing a conversation. For instance, P12 used Google's Live Transcribe when conducting interviews for a class project: *"if I couldn't understand what the other person was saying, [. . .] I would ask them to repeat it first. If I still didn't understand, I would just look down [at the app]."* Three participants described their use of captioning to augment sign language, such as P8's experience in discussion classes where *"some students had really lousy signings. So, I was able to look at CART instead."* The preference for flexible access strategies was shared by P15, who gets frustrated by CART writers who make her look at their captions, stating *"I have a right to lipread or look at the screen. It's my choice."*

The Environmental Affordances of Online Captioning.

The environmental shift from in-person communication to online video conferencing introduces new social norms, a unique set of possible interactions (e.g., text chat), and different design and technical needs, all of which shape the experience of DHH captioning users. As this study was conducted in summer 2020, participants reflected on the sudden shift to online communication driven by the COVID-19 pandemic.

Spatial and environmental considerations. Many DHH people rely heavily on visual and spatial cues to follow and participate in conversation, and participants described challenges and gains that came with moving to online, two-dimensional space. P3 missed being able to spatially connect

captions to a speaker like she would in person, while P13 lost out on being able to follow the gaze of *"the captioner [who] is going to probably look towards the person that's speaking."* However, P9 explained that while he struggled to match captions to the speaker on most platforms, Google Meet's speaker identification was *"fingers and eyes above everybody else that's doing captions."* Other participants echoed P2's experience, that online *"it's actually a lot easier to identify the speaker and easier to capture whoever is speaking at a time,"* and P4 and P11 mentioned the benefit of features that highlight the active speaker in online meetings.

Toward inclusive conversation access. Overall, participants found that moving life online brought with it features (including speaker identification) that have provided greater access. P1 stated that *"now that we can't go into the office it's been much more of an equalizing factor"* because his hearing colleagues are more motivated to limit overlapping speech and are also juggling lagging and malfunctioning technology. New online interaction paradigms also served to equalize conversation; for example, P8 discovered that automatic captions and an active text chat made it so that *"lots of people in the audience don't realize that I'm Deaf because we're all running on the same system at that point."* Several participants said automatic captions allowed them to access online meetings or social gatherings that would have otherwise been difficult to join in person. Furthermore, having text chat available at all times has created new opportunities: many participants' hearing conversation partners used the chat to correct mis-captioned jargon, P3 was able to use Microsoft Teams' messaging features to clarify confusing captions mid-meeting, and P2's friends used private chat to provide a transcript for her during uncaptioned Bible studies. These emergent social practices are enabled by the unique affordances of online environments.

Summary and Implications.

Our findings show that DHH participants' experiences of captioned conversations are deeply shaped by social, environmental, and technical context. Participants' accounts of the impact of their hearing interlocutors demonstrate that captioning is a highly social technology and that the people being captioned are key stakeholders in determining conversational accessibility. These findings affirm Seita et al.'s focus on the interplay between DHH and hearing people [254, 256, 255] but suggest that relational contexts, which may not be captured in controlled lab settings, are crucial to negotiat-

ing accessibility when using captions. Wang and Piper [293] outlined how Deaf and hearing dyads adapt when communicating without accommodations and we find that collective adaptation remains critical even after captions are turned on. Additionally, the challenges participants described during interactive conversation show that captioning alone does not guarantee access, particularly for DHH people who do not voice. While some of these hurdles could be lessened with better technology, they are also fundamentally social. There is a growing body of work on one-on-one automatically captioned interactions in which DHH people type their contributions [87, 86, 188, 254] and future work could further explore emergent social norms during these interactions. Further, understanding that captioning is used in parallel with other access strategies for interactive conversation prompts consideration of how future captioning displays could better match their contexts of use. Finally, our participants' experiences suggest that online captioned conversations are occurring in an environment with fundamentally different affordances than in-person conversation. Some of these affordances, such as missing spatial information, pose new access barriers which designers have begun to address [163]. Yet many other aspects of online communication may be well-suited to captioned conversation and features, such as text chat, have been little explored but hold great potential for future accessibility.

4.3.2 Design Probe Findings

While the previous section reported on participants' experiences using captioning in small groups, here we turn to ideas and responses that arose during the design probe activity. As described in the Method section, these probes were meant to prompt participants to envision a range of possible future captioning designs. The probes were described simply as *"a rough example"* of how a particular idea could be instantiated in a captioning setup and were shown only after an initial conversation about each feature. We quantify positive/negative reactions as well as provide qualitative summaries.

Caption Visibility.

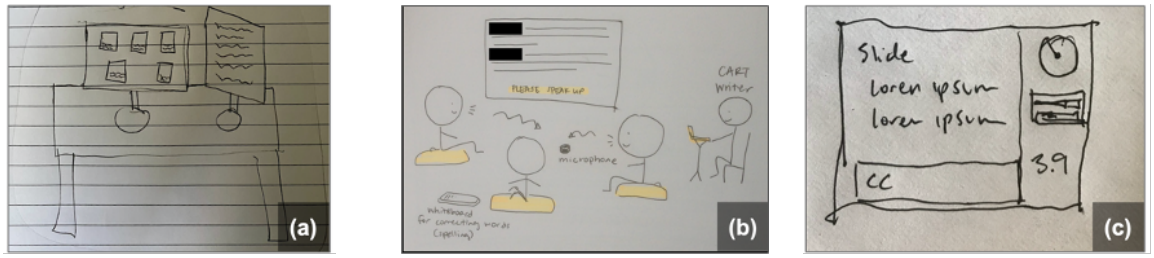
We asked participants to reflect on their ideal captioning setup and to consider the advantages and disadvantages of the following three in-person captioning setups and their digital analogs (Fig-

ure 4.1): captions available (1) only on the DHH person's device, (2) projected on a shared screen, and (3) available on the personal devices of all conversation participants. We posed these specific probes to gain a deeper understanding of how the type of display shapes captioning environments, to assess participants' feelings around making captions visible to their interlocutors, and to better understand how physical environments impact social dynamics.

Personal device only. Participants had mixed reactions toward having captions available to only themselves: some valued the autonomy and privacy of this setup, while others disliked it, describing feeling ostracized. For example, when considering how the display would impact conversation, three participants felt a personal display would be minimally disruptive because *"one person is usually pretty good at flicking between looking down and looking up,"* (P12) but four others disagreed, arguing *"the hearing people would be able to see each other [. . .] but the deaf person is glued to the screen"* (P9). Four participants took issue with the assumptions built into personal displays, arguing these assumptions suggest that *"the deaf or hard of hearing person is the problem that needs to be fixed"* (P10).

Shared caption display. Participants largely saw value in a prominent, shared caption display but some worried that it would reshape the conversation environment in a way that negatively impacts social dynamics. Several participants (N=6) explained that setting up a shared display felt like an effort to equalize the conversation: *"Rather than remaining in that dominant space where they normally do, everyone is a little bit more aware of what life can look like for us"* (P8). Participants further identified benefits of shared captions: four referenced past experiences having their hearing interlocutors notice and correct caption inaccuracies and three considered that their hearing conversation partners may also want captions, especially those who are learning English or have audio processing disabilities. Others, however, had concerns, including difficulty managing captions and presentation slides (P1) and a loss of eye contact with the speaker (N=3): P13 explained that needing to look up at the screen means she misses *"the human connection part, that's important."*

Captions on all personal devices. Many of the advantages of a shared group display also applied to the third setup, displaying captions on everyone's personal devices. Almost half (N=7) of the participants saw distributed captioning as an equalizing force, with P3 favoring it because everyone has the *"same thing going on and it helps hearing people feel like they're part of the deaf individual's team."* Participants also saw technical benefits of this setup, such as allowing customization (P2)



(a) (b) (c)

Figure 4.3: Captioning sketches from of (a) an online captioning setup with captions shown under each speaker and on a separate display (P14), (b) an in-person setup with a shared display and automated feedback for hearing speakers (e.g., "please speak up") (P2), and (c) a customizable captioning interface with speech rate, volume, and lag (P3).

and potentially leveraging device microphones to improve audio quality for captioners (P7). Others (N=5) worried that setting up captions for all would not be socially feasible: *"I don't see my friends using captioning devices."* (P1) Concern over sightlines persisted for four participants and P9 raised that *"seeing facial expression, seeing if they are angry, upset, happy—you can't get that from captions."*

Online contexts. While participants had varied reactions to each in-person setup, they overwhelmingly preferred making captions available to all during online conversations. Fourteen participants echoed P14's statement, *"if there's five percent of the captioning that is beneficial to other people, then why not?"* Only one participant (P11) was more hesitant. She explained that if no one else needed captions, she would prefer they were only available to her, *"simply because it's just sort of a personal thing."* Nine participants were excited by the fact that, with online interfaces, all participants could configure captions to meet their personal preferences and access needs.

Participants' preferred setups. In addition to the captioning visibility probes we showed, we invited participants to sketch out and reflect on their ideal captioning setups. Certain aspects of space were repeatedly mentioned—four participants stressed the importance of proper lighting and six wanted to be seated at a round table because with *"square, lateral type edges it's harder to look severely to my right or severely to my left. But if it's a more oval shape people are seated in a way that I can see them" (P13).* The form factors of captioning displays were also important to building

connection with interlocutors, and participants proposed various novel approaches. P2 considered how to make captioning environments feel cozy (Figure 4.3b), P7 wished for drone-based captions that hovered over speakers' heads, and three participants imagined the value of captioning glasses, including P9 who pictured hanging out in his living room with friends, *"able to lean back, lean forward, not have to have implements in front of them that makes them focus on one thing instead of looking at everybody."* P11 provided a counterpoint, explaining that her ideal experience is the one she is used to, seated next to a CART writer with the screen *"just between us,"* which she values because *"it's like you are connecting with a person and [. . .] there is something very human about it."* Participants highlighted how captioning environments and technologies dictate what social interactions will occur and how accessible they can be.

Adding Features to Captioning Displays.

We discussed six potential features to add to captioning displays: speech rate, speaker identification, volume, caption lag, overlapping speech, and error correction. For each feature, we introduced and had participants respond to the idea in general before showing a design probe and encouraging the participant to consider a variety of design possibilities. While speaker identification and overlapping speech were somewhat familiar to participants, the other features, as well as considerations as to how participants might like their hearing interlocutors to interact with them, are not captioning standards. We used the probes to explore what additional features participants perceived as potentially useful, to gauge reactions to having hearing interlocutors engage with captions, and to synthesize concrete design takeaways.

Speech rate (Figure 4.2a). Roughly half of the participants (N=8) thought speech rate feedback would be valuable for their hearing interlocutors to see, and three participants felt it would also be useful for themselves. P12 suggested that speech rate monitoring *"could be helpful if [a speaker] know[s] that [they] tend to talk really fast and that makes it hard for people to understand [them]"* and P15 saw value in shifting the social burden of telling people to slow down to technology. However, P9 was not optimistic, believing that inflexible social norms lead hearing people to pay attention to *"the content of what they are going to say, not whether they are talking slow or fast."* A key concern raised by six participants was that displaying speech rate could be distracting (our

probe included animation) and thus impact comprehension: *"The whole goal is to have things as least distracting as possible in order to maximize the ability to read"* (P14). Future designs could, as three participants suggested, only warn viewers when people speak too quickly or, as P10 proposed, display more caption lines when people speak rapidly so that the reader can catch up.

Speaker identification(Figure 4.2b). Reflecting earlier findings (section 4.3.1), participants unanimously wanted speaker identification in both automatic and human-generated captions, with seven participants underscoring that speaker identification in their current captioning tools is inadequate. Participants perceived speaker identification as more relevant for themselves than their hearing collaborators, though no one took issue with universal access and five saw it as actively beneficial: *"it would help them be more aware of what it is like when a person cannot hear"* (P5). When implementing speaker identification, ten participants emphasized that color-coding speakers (a feature included in our design probe) is a useful visual shortcut.

Volume (Figure 4.2c). Displaying the current speaker's volume and background noise levels was relatively popular; ten participants were interested, three uninterested and two had mixed feelings. While some participants (N=3) wanted to know how loudly they were speaking so that they could self-regulate, more (N=7) were interested in providing volume feedback to the group, though four worried the display would be distracting. Three participants hoped that displaying background noise levels might lessen hearing people's tendency to ignore it, lead to a quieter, better setup for captioning—an example of wanting to use the captioning technology to shape social norms and alter the environment. P14 also remarked *"I think hearing people would benefit too because then they would know where the noise is coming from too."* Ten participants independently suggested that they would get more value out of sound identity than volume levels, and desired sound classification integrated with captioning.

Caption lag (Figure 4.2d). DHH participants were interested in conveying how delayed captions are to their hearing interlocutors, though largely did not consider lag to be personally useful. While six participants suggested that seeing the lag would help them make sense of confusing captions, the other nine expressed that they always assumed captions were delayed and therefore did not want feedback. However, eight participants believed that highlighting caption lag for hearing people could support a shared attention to how captions function in practice. P1 stated, *"it's great because they might understand like why I might be jumping in later than I was supposed to,"* and P8 hoped

"this might actually help people put a little bit more buffer time into their speaking." While our probe conveyed lag in terms of seconds delayed, participants brainstormed other ideas: P4 proposed a *"number of sentences delayed"* metric and P12 imagined a more visual representation: *"It could be like dots indicating every vowel or important recognizable-as-speech sound, [. . .] something that transforms into the word as the captioning service catches up."*

Speaker overlap warning (Figure 4.2e). Reflecting the fact that captions are not able to capture multiple speakers at once, participants were overwhelmingly interested in an overlapping speaker warning, both for personal (N=14) and group (N=10) use. P12 was the sole participant uninterested in a built-in speaker overlap warning, explaining, *"this feels like more of a social norm thing rather than something that the programming should account for."* Ten participants wanted this information shared with their conversation partners because people *"just get really really excited and start speaking up over each other"* (P2). The other five participants, however, had reservations about the social impact of this technical intervention, with P9 believing *"hearing people want power, so they say, 'Ah, well everybody else can stop talking, I am going to continue"'* and P8 worrying the warnings could have the side effect of shutting down the casual conversations she loves participating in. When considering implementing overlap warnings, participants imagined different roles for this technology: for instance, P15 appreciated that an automated warning could be perceived as less socially disruptive, but P4 proposed adding a blaring siren so that her interlocutors *"all go, 'Oh crap, I need to stop."'*

Shared error correction (Figure 4.2f). Managing captioning errors has received significant attention in prior work (e.g., [27, 101]), but we sought to explore how participants felt about engaging their direct interlocutors to address errors in real time. Due to time constraints, we only discussed a feature to allow conversation partners to correct inaccurate captions with thirteen participants. Ten participants were interested, and they imagined many benefits of crowd-sourced corrections, such as addressing domain-specific acronyms (P1) or captioning a multilingual workplace: *"[If] someone comes from a similar culture as the speaker, they might be able to input those vocabulary words"* (P2). However, three participants did not think that a group could provide error corrections quickly, and P4 postulated, *"I don't know if you can listen to people speaking, and then also listen to yourself, and also make corrections to captions. I think you need to have someone there dedicated to doing the corrections."* Beginning with our basic mockup, participants brainstormed ways

to make error correction useful and readable in real-time, with six people independently suggesting that color could link corrections with their place in the transcript. This process made clear that while participants are interested in shared error correction, it is a complex social and technical problem.

Ideal interfaces.

Alongside environmental configuration preferences, the sketching exercise we completed with participants highlighted their ideal captioning interface designs. The majority of participants focused on interfaces for online communication, though some considered in-person interface design. Several participants wanted to have access to features we had discussed, such as P1, whose ideal setup (Figure 4.3c) included speech rate, volume, and lag monitoring, which could be "individually customizable to display or not, depending on the end user's preferences." The desire for customization was shared by P8, who posed that it would also be useful to "choose which features I want and are relevant to me depending on the situation." Other participants proposed new feature designs: five people independently suggested displaying captions next to each speaker's online video feed (Figure 4.3a) because it *"eliminate[s] the need to identify the person speaking, if each of them ha[s] their own individual caption"* (P5). P2 wanted to engage her hearing conversation partners in making captioning more effective by using online meeting software outfitted with *"different buttons to say slow down, speak softly, speak up, speak loudly, talk faster, please spell the word."* Participants imagined technical setups that leveraged the unique environment to build new social interactions and feedback for themselves and their interlocutors.

Summary and Implications.

Participants' responses to our design probes provide considerations for captioning designers and highlight the interrelated factors that shape the utility of captioning tools. When considering how to display captions, participants focused on the tension between shared displays' potential to negatively alter in person conversation dynamics and the isolation and information loss that can come with being the only person accessing captioning. However, this tension largely disappeared when participants considered captioning online conversation, suggesting that videoconferencing is a unique environment which could support socially acceptable, lightweight technical group captioning inter-

ventions. Participants were not uniformly excited about all of our design probes. When considering how to best improve DHH peoples' captioning experiences, providing speaker identification and overlap warnings are clear priorities. However, while the majority of participants were not interested in personally using speech rate, volume, or lag feedback, most imagined that they could be useful in guiding hearing people toward more accessible behavior. This, along with interest in shared error correction suggests that technology that shapes group social norms around captioning is worth pursuing. Additionally, participants' responses highlight that preference for caption configuration or new features are dependent on the interactions between social, environmental, and technical factors. Fully accounting for the context that shapes participants' experiences and preferences surrounding captioning can open new avenues for design. For instance, while prior work on visual dispersion has attended to the importance of captioning form factors, [50, 136, 157, 170, 200, 228], future captioning designers could integrate tenets of DeafSpace [81] to consider how to create in a way that matches Deaf environmental and sociocultural norms. Finally, our participants highlight the complex social dynamics these tools could impact, and while many were excited by and interested in trying feedback tools, others remind that not all hearing interlocutors are equally amenable to changing their behavior.

4.4 Discussion

The findings and implications presented above emphasize the social, technical, and environmental factors impacting small group captioning. We have provided an empirical account of DHH participants' experiences and their perspectives on future captioning design. In the discussion below, we synthesize the sociotechnical nature of small group interactive captioning. Further, we reflect on captioning as a group responsibility, the design of future captioning systems, and our study's limitations.

4.4.1 Social, Environmental, and Technical Influences on Small Group Captioning

Throughout our data, participants consistently explained their experiences with small group captioning as shaped by the interaction between social (e.g., DHH people's communication styles, hearing people's mal/adaptive behaviors), environmental (e.g., furniture configurations, features of video-

conferencing software), and technical (e.g., delay and accuracy of captions, captioning interface design) factors. Considering all of these factors together provides a more complete understanding of the use and efficacy of captioning technology. For example, our findings show that participants' preferences for captioning form factors are irreducibly determined by the social interactions they permit or prevent and how they shape environments (e.g., needing to have a room with a projector set up as opposed to using a personal laptop to view captions). When considering our participants' experiences of videoconferencing, we found that phenomena such as hearing people correcting captions in the chat can be more completely understood when considering the affordances of online environments that allow for real-time corrections, the social relationships that lead some hearing people to take on caption correction, and the technical failings of captioning that necessitate corrections.

Recognizing that these factors must be considered together to fully contextualize the use of captioning technology has implications for how we as HCI and CSCW researchers work. When formulating research questions, designing studies, analyzing data, and reviewing papers, researchers should consider and seek to account for social, environmental, and technical influences on captioning technology. Many proposed captioning designs have been evaluated out of context (e.g., [29, 106]) or in terms of a narrowly defined outcome (e.g., improved comprehension [170] or performance [50]), and future work could complement these analyses with a focus on their social, environmental, and technical contexts. As researchers move to consider the role that hearing people play in conversational accessibility, findings from controlled experiments, such as work done by Seita et al. [254, 255, 256], could be contextualized by qualitative work focused on social relationships (e.g., [293]) and the environments in which technology is used (e.g., [149]).

4.4.2 Toward Shared Responsibility for Small Group Captioning Success

Traditionally, researchers have identified DHH people as the primary users of captioning (e.g. [29, 50, 136]). However, building from our participants' accounts of the impact their hearing interlocutors have on captioning's efficacy, we propose treating captioning as a technology used by all members of a group, including hearing people and not solely DHH individuals. Hearing and DHH people both rely on captioning to understand and be understood, but, as participants explained, hear-

ing people often do not recognize their stake in captioning's success. While we believe captioning research should continue to center DHH people—because if captioning does not work, it is DHH people who will lose access—we also seek to reframe captioning as a community-based accommodation [147]. This reframing opens up possibilities for captioning technology designed to support group interdependence [25] by acknowledging that captions cannot work unless people are willing to work with them.

This shift provides opportunities to de-center the hearing world norms that are often present in assistive technology design. Many of our participants saw promise for more equitable interactions by introducing hearing collaborators to their world rather than staying in the *"dominant space"* (P8). Participants' proposals to make this shift included simple changes, such as displaying captions for the entire group, and more extreme interventions, including playing loud sirens when hearing people break captioning-friendly norms. Furthermore, participants described the benefits of hearing people learning more accessible communication styles, such as combining typing and ASR captions for casual interactions but aired frustrations around consistently needing to teach these approaches. These sentiments extend Wang and Piper's [293] findings around Deaf/hearing collaboration without accommodations. If rooted in Deaf epistemologies [243], future captioning systems could both teach and reinforce captioning-friendly behaviors to shift labor away from DHH people.

However, we resist embracing captioning for the group without considering potential challenges and opposition. As many of our participants illustrated, there can be high costs to using captioning in an audist world, ranging from social discomfort to workplace barriers, and some remained skeptical that hearing people would ever change inaccessible behaviors. Future work will need to explore the social factors that led some participants to work extensively with hearing interlocutors to collaboratively improve access while others desired captioning solutions that minimized hearing people's involvement. Additionally, some participants were uninterested in changing how they communicate and, as designers of accessible technology, we must respect that technological intervention is not always appropriate or desired. Furthermore, captioning inherently centers spoken conversation. While many of our participants were oral, worked in predominantly hearing workplaces, and socialized with hearing people, some chose to orient their lives around the Deaf community. Lane [168] argues that trying to redirect Deaf people from this Deaf-World is unethical. Regardless of communication mode, however, there are unavoidable interactions in hearing spaces (such as stores

and restaurants) where captioning could be a useful tool. Therefore, we must balance building tools to support these interactions without implicitly or explicitly situating oral conversation with hearing people as superior to Deaf-World norms.

4.4.3 Reflections on Future Captioning Design

Building from our call to integrate social, environmental, and technical factors into captioning research and our reframing of captioning as a group technology, we provide concrete design considerations. Specifically, (1) approaches to providing real-time feedback during captioned conversation, (2) opportunities for online communication to advance captioning technology, and (3) discussion of how to design captioning technology for all DHH users.

Our findings suggest that adding real-time feedback and error correction to shared captioning displays are promising areas for future exploration. Participants identified several features that could be especially useful in guiding hearing people's understanding of captioning-friendly behavior (e.g., speech rate, lag), while other features may be more useful in providing DHH people with context on the captioned conversation (e.g., speaker identification, overlapping speaker warnings). How to specifically design and implement each of these features, however, is an open question. A range of options exists such as displaying information directly (e.g., raw decibel levels), integrating information into caption design (e.g., visualizing words that have been spoken but not captioned), or only providing warnings when captioning-friendly norms are breached (e.g., a warning when speakers talk too quickly). Further, there are likely pros and cons to conveying any additional feedback via a shared interface to all conversation participants versus providing individual displays with differentiated feedback—perhaps based on hearing status. Group-generated error correction also merits further study, but many social and technical considerations remain, such as who does the corrections, how can it happen efficiently, and how are corrections integrated into captions. Implementing and testing these features with DHH and hearing users is an important next step to explore questions such as how effective the feedback is at driving behavior change, how to appropriately bring attention to captioning without overwhelming participants, and how receiving feedback impacts all conversation participants' experiences of captioned conversation.

Our study highlights opportunities to evolve captioning tools for online conversations, with a

unique capacity to build group-oriented tools. For example, participants perceived that their hearing interlocutors face new constraints online, such as a single audio channel and technical delays, which align with more caption-friendly communication. Online systems could be designed to strengthen these social gains, leveraging the technology-mediated nature of online environments to more easily implement interventions. Other unique online affordances, such as the omnipresence of text chat, the social and technical ease of turning on captions, easily automated speaker identification, and less-settled social norms, could be leveraged to address the distinct disadvantage DHH people face without visual and spatial cues online. Currently many of these features are difficult to implement in person, even with customized hardware such as microphone arrays, and exploring their impact online could help drive priorities in software and hardware development to support captioning users as in-person conversation becomes feasible again.

Finally, future captioning systems should explore ways to allow full participation for all DHH people, regardless of their communication preferences, during captioned conversation. Our study participants who prefer not to voice (8 out of 15), stressed that captions do not support their contributions to a conversation, a concern relevant to the estimated 100,000–500,000 Americans who primarily communicate in ASL [209]. As our findings demonstrate, there are many captioning use cases for people who prefer to sign, particularly as automatic captions become widespread, and it is critical to consider people who do not voice when designing for interactive captioned conversations. This extends beyond simply making it possible to type, as designers must consider how to socially integrate typed contributions into the flow of conversation and account for differences in typing speed as compared to speaking or signing (50 vs. 160 wpm typing on a touchscreen vs. speaking and signing [23, 245]). Future work could explore, for example, allowing people to hold their conversational turn while they type, ways to stream typed contributions as they are generated, and how to help change social expectations around the pace of conversation.

4.5 Limitations

Our study has four primary limitations. First, while online recruiting allowed us to expand our geographic reach, we conducted this study during global health, political, and economic unrest, which limited recruitment to those who could spend 90 minutes participating in an online research study.

Our 15 participants were all U.S.-based professionals with high-speed internet access. Second, as we explain in section 4.2, we chose to conduct this research with DHH participants only, and we do not claim to advance knowledge about hearing people's experiences during captioned conversation. Instead, we explore how DHH people's captioning experiences are impacted by their hearing interlocutors and their preferences for future engagement from hearing people, intentionally giving the power to dictate future design directions to DHH participants only. Future work exploring group captioning tools should involve both DHH and hearing participants. Third, we focused our design probes on contexts where a single DHH person communicates with a group of hearing people, and we do not claim that our findings extend past this scenario. As some participants explained, the conversational dynamic can change when multiple DHH people are in conversation with hearing people. Finally, while DHH people are a large portion of real-time captioning users, they are not the only group that uses captioning as an access tool. We outline findings specific to DHH people, but future work could investigate to what extent these findings are relevant to other captioning users.

4.6 Conclusion

In reporting on a formative study with 15 DHH participants, we present an empirical account of DHH people's experiences of captioning during small-group conversation, highlighting the social, environmental, and technical factors that shape the use and usefulness of real-time captioning. Additionally, we outline participant's preferences for the design of future captioning systems, providing design implications regarding captioning as a group technology. Throughout, we discuss participants' experiences of and design preferences for online communication, recognizing it as an environment with unique affordances and considerations for captioning. Our discussion highlights the need to consider social, environmental, and technical context when undertaking captioning research, proposes a shift toward treating captioning as a technology used by groups, and outlines future design considerations. Guided by Deaf and disability studies, we look to a future where DHH and hearing groups use captioning as one of many tools to negotiate conversation accessibility that questions the hearing world's norms.

4.6.1 Within-Dissertation Context

Beginning my dissertation with this empirical work set the direction to continue exploring interactions between DHH and hearing people, anchored in the types of experiences and preferences participants expressed in this study. The preceeding and following chapters are in converation with this work, identifying the importance of first paying attention to contextual factors in order to design for collective access work (what I term in this chapter 'group captioning'). Though this chapter largely echoes the tenets of the collective communication access framework I develop in Chapter 3, my theoretical background was not as developed at this point in my PhD, and therefore this work draws entirely from disability and Deaf studies, not including disability justice frameworks. I have been excited to see the social/environmental/technical framing I established in this chapter taken up by other captioning researchers in the years since it has been published (e.g., [69, 257]).

Chapter 5

CODESIGNING VIDEOCONFERENCING TOOLS FOR SMALL GROUPS WITH MIXED HEARING STATUS

This chapter contains work that was published at CHI 2023, titled *"'Easier or Harder, Depending on Who the Hearing Person Is': Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status"*. I led this work with collaborators Soo Hyun Moon, Lucy Jiang, Steven Goodman, Raja Kushalnagar, Jon Froehlich, and Leah Findater.

Within the broader scope of my dissertation, this chapter takes up the notion of collective communication access and conducts codesign activities with participants to identify what kinds of technology designs could foster collective approaches to commmunication while remaining practical and desired. It is most concerned with answering DRQ2, but is anchored by DRQ1 and adds conversation partners' relationhips (e.g., relational context) to the set of factors that I have demonstrated impact captioning experiences.

5.1 Motivation

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) [290] and the rise of videoconferencing following COVID-19 [20], platforms such as Zoom, Google Meet, and Microsoft Teams [180, 155] now provide on-demand but imperfect captions. However, ASR inaccuracy [290] makes the impact of group dynamics on conversation accessibility particularly relevant (e.g., [196, 254, 255, 256, 183]). Furthermore, following the Deaf community's long-standing argument for access approaches that decenter hearing norms [19, 81, 166] and disability justice activism's framing of collective access [131], 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 [247, 288], and adapting group social norms can be an effective

but difficult-to-maintain approach to improve accessibility [163, 183, 257]. However, prior work has not learned from small groups about how they currently address these challenges nor how they could be better supported. Working with established mixed hearing groups can provide insight into emergent social accessibility practices co-developed together over time, and draw upon their lived experiences and collective communication strategies to inform future caption tool design. Therefore, this work seeks to answer the following questions:

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?

To examine these questions, we conducted a three-phase codesign study with four established small groups: a total of 17 participants (10 DHH, 7 hearing). 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 [196, 257, 293], 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 versus 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., [28, 106, 157]).

In summary, this work contributes: (1) an empirical account of how mixed hearing status 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.

5.2 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 between 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.

5.2.1 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 [250]. 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.

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 using automatic captioning 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 odds that participants

would have 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, beginning 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. [196]—speech rate, volume, caption lag monitoring, speaker identification, speaker overlap warning—to discuss potential captioning tool features and as a basis for collective brainstorming.

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 [308] of each group's top three ideas (12 features total) to provide a more tangible representation of each idea for critique and discussion. 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.

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 used in the context of 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.

5.2.2 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. We thus 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 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 [113, 250], 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 to-

Table 5.1: Group A Demographics

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use
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

gether and thus 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).

Group A

Group A (see Table 5.1) 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).

Group B

Group B (see Table 5.2) 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).

Table 5.2: Group B Demographics

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use
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

Table 5.3: Group C Demographics

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

Table 5.4: Group D Demographics

Pseudonym	Hearing Status	Preferred Communication Style	Frequency of Captioning Use
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 C

Group C (see Table 5.3) 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. 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).

Group D

Group D (see Table 5.4) 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 [185] 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).

5.2.3 Analysis and Positionality

We analyzed our data using reflexive thematic analysis, as outlined by Braun and Clark [37, 39]. 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, and because the game transcript consists of yes/no questions, we did not directly analyze that data. Other authors reviewed a subset of coded transcripts, providing feedback and comments on the first author's coding. The final codebook contained three versions for each session of the study procedure; codes for the second and third sessions included all codes from the previous sessions as well as new, tailored codes. Through discussion with coauthors, the lead researcher combined codes and data into the themes that now serve as findings subsections. After doing so, 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.

5.3 Findings

Below, we describe our findings: we first highlight existing socio-technical practices with video captioning and emergent access strategies before discussing design ideas and video prototype reactions.

5.3.1 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.

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 would only use 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, 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 they did 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.

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 shaped 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 are not 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.

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."* 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.

5.3.2 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 between Sessions 2 and 3 (Figure 5.1). Then we discuss all groups' 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., [196, 257]), 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. 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.

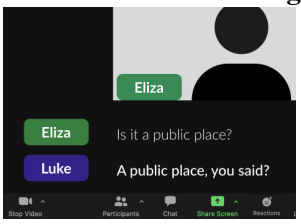
Codesign Artifacts

Figure 5.1 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.

Speaker Identity and Overlap

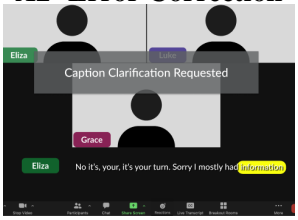
Quickly identifying speakers and automatically flagging when they overlap were clear priorities for participants, with Groups A, B, and C designing features (Figure 5.1c, Figure 5.1e, Figure 5.1i) to address these information gaps. While speaker identity and overlap have both been explored in prior literature (e.g., [105, 124, 135, 163, 172, 183, 228, 278], participants' experiences demonstrate that crosstalk and ambiguous speakers have yet to be resolved, with key design nuances and social

A1 - Volume Monitoring



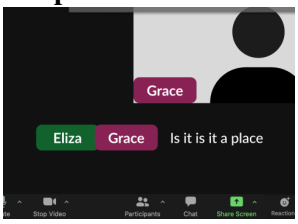
(a) *"Bigger or bolder font as they get louder"*- Anna

A2- Error Correction



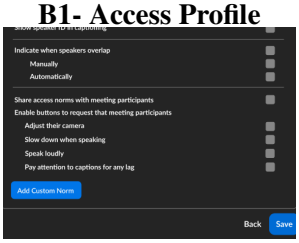
(b) *"Subtly [let] the speaker know if something was unclear"*- Anna

A3- Speaker Identification



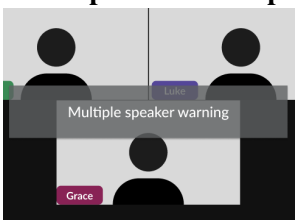
(c) *"[It lets you] intervene and say 'Hey, like, turn-tale here"*- Amelia

B1- Access Profile



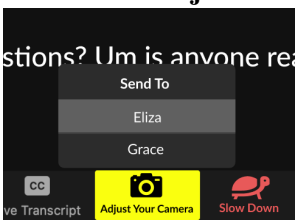
(d) *An accessible group "standard ... provided automatically"*- Brenda

B2 - Speaker Overlap



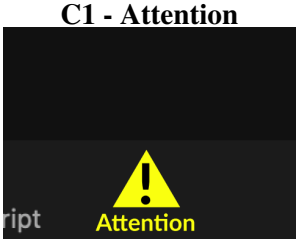
(e) *"In the middle of the screen, in a way that it's not too jarring, but you can't ignore it"*- Blake

B3- Camera Adjustment



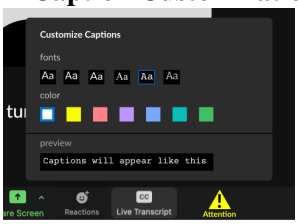
(f) *"If I'm talking ... and I'm muted ... that same kind of approach could be used"* - Bea

C1 - Attention



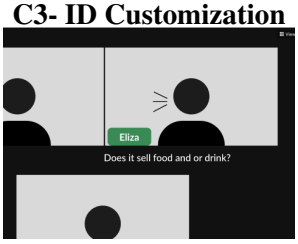
(g) *"'Something is wrong' button [that] mean[s] several things; 'hold on', 'I'm lost'..."*- Cad

C2- Caption Customization



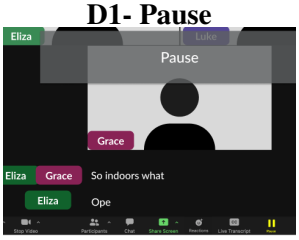
(h) *"Adjust the captions' size, colors, font from within the meeting"*- Colin

C3- ID Customization



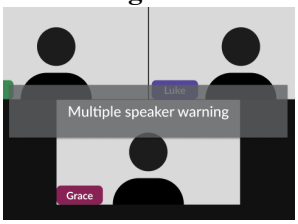
(i) *"If I don't know who's speaking, I don't have the contextual information"*- Camille

D1- Pause



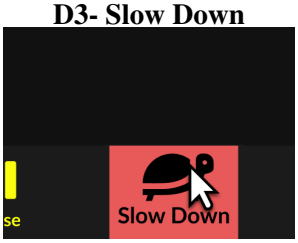
(j) *A way to "realize ... 'I need to wait"'*- Dot

D2 - Background Noise



(k) *Could display background noise like "(barking in the background)"*- Daisy

D3- Slow Down



(l) *An "alert for speaking too fast"* - Daisy

Figure 5.1: 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.

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 (Figure 5.1c). 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 was 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 (Figure 5.1i) garnered mixed reactions. Amelia liked that it would help consolidate information and limit instances where people textit"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). A suggested solution was 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 explained; *"[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 indicating overlap within speaker identification and sending conversation participants a pop-up when it occurs. When reacting to video prototypes of speaker overlap notifications(Figure 5.1e), participants liked the baseline feature but had suggestions for nuances to build into future implementations. Participants stressed the importance of language choice in notifica-

tions, 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).

Support for Behavioral Feedback

The next category of features involved 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, often using a similar pop-up implementation (Figure 5.1b, Figure 5.1f, Figure 5.1g, Figure 5.1j, Figure 5.1l). 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 for other speakers to slow down. While these access barriers have been discussed in prior work (e.g., [36, 50, 136, 139, 157, 196, 254, 255]), we focus on how technical tools could help to mitigate them by driving behavior change. While Group A's design (Figure 5.1b) included the opportunity to either flag or correct captioning errors in real time, participants only saw promise in flagging errors. 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 errors. 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 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 (Figure 5.1f), 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, as for some *"the exhaustion of being on camera all the time"* (Bridget) was 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 (Figure 5.1l). Speech rate alerts were considered in context of participants' current strategies to get speakers to slow down. For example, Brian shared his approach when new vendors at work would *"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 (Figure 5.1j) and attention (Figure 5.1g) 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 identifying 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.

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 (Figure 5.1d) and desired greater customizability over current captioning interfaces (Figure 5.1h). Customizability has been highlighted throughout prior work [50, 72, 105, 196], but the role of platform infrastructure on accessibility for DHH people has so far only been explored in the context of sign language use [247]. Group B's access profile (Figure 5.1d) allowed groups to enable desired features (such as the behavioral feedback tools discussed in section 5.3.2) 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, concerns 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 (Figure 5.1h) 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 well, and participants highlighted the need for greater control and customizability.

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 (Figure 5.1a) and identifying non-speech sounds (Figure 5.1k). Though Google has integrated sound recognition into the Live Transcribe app [109], it is not available within commercial videoconferencing tools. Volume visualization has only been explored in the context of pre-recorded captions [124].

Group A's proposal to visualize speaker volume in captions (Figure 5.1a) 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 this is important: *"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 (Figure 5.1k), 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 considered 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.

Summary

Overall, our participants identified a number of features that they would be excited to use and, 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 exciting new paradigm. Finally, products like Live Transcribe [13] have integrated sound recognition into their ASR apps, and our results indicate that this would be valuable during videoconferences as well. Participants also identified features they did not want. While increasing group awareness of conversation break-

downs 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.

5.4 Discussion

In this chapter we report on our codesign practice with established groups of DHH and hearing people, finding that people actively negotiate and build accessible group practices on top of captioning use and identifying participant-driven priorities for future videoconferencing features to support accessible group communication strategies. 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.

5.4.1 Designing with Established Groups

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 [149, 196] and communication practices of established DHH/hearing pairs [293], whereas we highlight varied ways that existing groups with mixed hearing abilities engage with captions (e.g., work colleagues establish formal rules while cousins rely on shared history). 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 [254, 256, 255], 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. Though recruiting existing groups is nontrivially difficult, it was through codesigning with these groups that 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 learning 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.

5.4.2 Implications for the Design of Future Videoconferencing

While prior work has identified the value of conversation norms (e.g., [99, 149, 196, 293]) and Seita et al. [257] 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. Underlying the specific feature designs participants developed throughout our codesign sessions are broader considerations for the design of videoconferencing tools that can support 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), access is a fundamentally social problem. 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 identi-

fied are not technically complex to implement and leverage the existing functionality of video conferencing systems (e.g., participants' design of a 'slow down' button). However, these tools could be critical in helping shape accessible conversation dynamics. We encourage designers and researchers to explore these avenues that may be less obviously novel but desired and socially impactful.

- **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.

- **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. [50, 72, 105, 196]) 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 [157] 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., [28, 29, 105, 106, 124, 143, 232, 254, 258]), our participants' perspectives agreed that reading captions requires too much cognitive load to make on-caption visualizations useful [29, 258]. 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.

- **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.

- **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. If designing technology to guide behavior change, it must 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.

5.4.3 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 [220]. Disability Justice activists have furthered this thinking and operationalized it in their principle of collective access, and 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 [196, 293], 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 [81], 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 [247].

5.4.4 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. Additionally, our study was conducted using Zoom, biasing designs and reflections. While we believe that findings are applicable beyond this single platform, future work may want to explicitly explore the role 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 world 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.

5.5 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.

5.5.1 Within-Dissertation Context

By codesigning collective communication access technologies with mixed hearing ability groups, I further demonstrate the ways that social, environmental, and relational factors impact groups' use of and preferences for technologies. This chapter also practically demonstrates ways that a collective communication access approach can be built into captioning technologies, providing both specific designs that merit further development and design principles to navigate building tools to be used by all members of mixed hearing ability teams. Though researchers' perspectives and interpretations are never eliminated from the research process, by utilizing codesign methodologies we anchor these designs in groups' ideas and priorities, rather than a top-down authorial view of what collective

communication access technologies ought to be. Throughout this process, groups' primary reactions to proposed technologies focused on how they would shape the kinds of conversations they held and the social impact tools would have on their specific groups and other imagined groups. This further suggests that technologies' realistic context of use is a primary driver of participants' preferences for captioning tools.

Chapter 6

CHARACTERIZING USER-DRIVEN CAPTIONING PRACTICES ON TIKTOK

This section contains research published at CHI 2024 titled *"'Caption It In An Accessible Way That Is Also Enjoyable': Characterizing User-Driven Captioning Practices on TikTok"*. I led this project, coordinating a student research team consisting of Tessa Eagle, Pitch Sinlapanuntakul, and Soo Hyun Moon, and working with faculty advisors Kate Ringland, Jon Froehlich, and Leah Findlater.

While previous chapters explored real-time captioning use in synchronous, small-group conversation, for this chapter I explore a different captioning context: user-generated captioning on TikTok videos. I chose to switch contexts for three key reasons: first, to demonstrate that even during asynchronous, indirect communication choices communicators (i.e., video creators) make significantly impact caption use, second, to explore a different set of social, relational, and environmental factors, and, finally, because user-generated captioning is a context where hearing communicators are already assumed to be responsible for captioning. In synchronous conversation, hearing people do not traditionally interact with captions, but for user-generated content to be captioned, its creator must make choices to enable this. While platforms such as YouTube provide automatically generated captions by default if videos' language is specified [290], to caption TikToks requires active engagement on the part of the video creator (even after TikTok added automatic captioning support in 2021 [121]).

Therefore, in this chapter we explore what happens when platform users (rather than solely algorithms, trained professionals, or television or movie studios) are in charge of captions, identifying the captions that emerge in the specific social and environmental constraint of the TikTok platform, aligned with DRQ1. We end by reflecting on how future user-generated captions could be better designed to match captioning users' access needs and preferences, and how to engage video creators in making that change, engaging with DRQ2.

6.1 Motivation

There has been a significant shift in how people consume entertainment media; moving from traditional media (e.g., TV, movies) to online, user-generated content [296]. In 2020, for example, TikTok was the most downloaded app [216] and, in 2023, was viewed 4.4 billion minutes *per* day by US adult users alone [35]. This paradigm shift has important accessibility implications. Globally, there are well-established, legally enforced captioning standards and requirements for media that appears on television [302, 84, 92] as well as emerging legislation for professionally produced content uploaded to the internet [1]. However, the accessibility of user-generated videos remains unregulated. This leaves caption users with no legally protected access to platforms like YouTube and TikTok [173] — a notable gap.

Unlike other video sharing platforms, TikTok became popular at a time when caption use has become mainstream, including among their young user base [246]. Despite no formal requirement and significant initial obstacles, TikTok creators have developed a culture of captioning content [57, 137, 46]. Though TikTok rolled out automatic captioning and a built-in closed captioning interface in April 2021 [121], many TikTok creators have adopted a highly stylized, open captioning approach that embeds captions into their videos.[1] Prior HCI research on captioning design has identified strong preferences for online captioning styles that align with television standards (e.g., [28, 59, 158, 258]), but studying TikTok offers an opportunity to explore both how creative captioning practices organically emerge on a social media platform as well as how they compare to traditional captioning practices. Despite media coverage and emerging academic interest in TikTok captioning [260], there has not yet been a comprehensive study of captioning practices on the platform. As media consumption continues to shift toward unregulated, user-generated content, studying captioning practices on TikTok provides an opportunity to understand how online videos are captioned and how those approaches serve or fail caption users.

Therefore, we set out to answer the following research questions:

1. How is user-generated captioning implemented on TikTok?

[1]Closed captions can be toggled on and off, using a TV remote or video player settings, whereas open captions are burned into a video and will always be on screen.

2. How do choices made in generating and placing captions impact TikTok users who need captions to meet a Deafness or disability-related access need?

To address these questions, we ran a two-phase study: a large-scale content analysis and a complementary interview study. We identified three major dimensions of user-generated open captions: how videos represent *audio and language* in text, how captions are *styled and placed*, and how well the *content* of captions matches a video's audio. By integrating participant perspectives with our content analysis, we found that 1) the current state of audio and language coverage in captions aligns with participant preferences—speech is nearly completely captioned while music and sound effects are rarely captioned, 2) some captions' color, size, placement, and timings varied from standard expectations, often decreasing readability, and 3) non-verbatim captioning and errors, while present in captions, were often minimally disruptive, and additional content (e.g., emoji) could provide richer paralinguistic information. Notably, we found that, despite the lack of regulation, the current state of user-generated captioning on TikTok does allow caption users to meaningfully engage with the platform. However, participants still identified significant room for growth, highlighting the need for user-generated video-specific standards, along with tools that encourage more creators to not only caption their videos, but to caption them *well*, could further extend access.

In summary, our research contributes 1) a large-scale analysis of TikTok open captioning, contextualized by its impact on caption users, and 2) steps toward future standards for user-generated captioning.

6.2 Related Work

This chapter draws on the prior work discussed in Chapter 2 and we also provide background on the specific landscape of TikTok research in order to contextualize our findings.

6.2.1 TikTok and Research

The social internet has shifted in recent years toward short-form video content. TikTok, developed in 2016, is an algorithmically-driven social media platform primarily focused on video sharing. Since 2023, the platform has reached over 1 billion active users worldwide[2] and was the most downloaded

[2]https://www.demandsage.com/tiktok-user-statistics/

app in 2020 [216]. Given this rise in popularity, other platforms introduced similar features for publishing brief video content (e.g., Instagram Reels – released in August 2020[3] and YouTube Shorts – released in September 2020[4]). Content on TikTok is primarily short-form video-based, with videos initially limited to 60 seconds in length, and more recently extended to ten minutes [280]. TikTok is not only notable in its bias toward short content, which may be easier to caption, but recent work highlights ways that TikTok's platform incentivizes specific kinds of content (e.g., a strong bias toward repetitive trends, a desire to optimize content for the algorithm) [307, 32], which can encourage a culture of open captioning. While initially adopted by younger populations for dance-related challenges, TikTok's user base has since diversified, and the app is now primarily used by 19 to 29-year-olds[5].

TikTok has recently gained considerable attention in HCI research. Much of this work has centered on sensemaking around the proprietary and elusive TikTok algorithm [146, 151, 204, 261, 71] and specific sub-communities that vary widely from grieving individuals [89], to those with experiences of psychiatric hospitalization [252] and eating disorders [119], to users discussing acne and skincare [304, 97]. Marginalized groups have also found community and belonging on TikTok. For example, research has highlighted LQBTQ+ communities [71, 262], neurodivergent-related content [6, 78], inclusive technology for disabled individuals [77], and discussions of shadowbanning in queer, trans, and disabled TikTok communities [240]. Research on the credibility of information disseminated on TikTok has also proliferated, especially with regards to the COVID-19 pandemic [263, 17, 176]. Notably, these studies often analyze the 100 most liked or viewed TikToks within certain topics or hashtags [300, 248, 154, 304].

However, little work has focused on the accessibility of the platform. TikTok did not introduce automatic captioning until April 2021 [121]. Given the primarily video-and-audio-based nature of the platform, captions are an integral part of participation for d/Deaf, hard of hearing, neurodiverse, and disabled communities. Simpson et al. found that much of TikTok's accessibility has stemmed from grassroots community efforts, largely by disabled communities who have developed workarounds to address app inaccessibility [260].

---

[3]https://about.instagram.com/blog/announcements/introducing-instagram-reels-announcement

[4]https://blog.youtube/news-and-events/one-year-of-youtube-shorts-what-weve-learned-so-far/

[5]https://www.demandsage.com/tiktok-user-statistics

6.3 Methods

We employed a two part mixed-methods study. We first collected and analyzed a dataset of TikToks to characterize how user-generated TikTok videos are captioned. Then, we performed complementary interviews with TikTok users who need captions to access the platform to identify the impact of these current captioning approaches.

6.3.1 Content Analysis

We began by collecting a dataset of TikTok videos and developing an initial codebook. We describe our method for collecting and analyzing videos, which led to the overview of TikTok captioning approaches we present in Section 6.4.1.

Data Collection and Analysis

We created a dataset comprised of (1) TikToks likely to be shown to a general audience and (2) TikToks related to Deafness and disability. We took this two-pronged approach[6] to understand how content is captioned both when it reaches a broad audience and when it is made by communities invested in access. We chose to collect both general audience and Deafness and disability-related videos to analyze a breadth and variety of captioning practices on TikTok. Further, our analysis was targeted at understanding current practices on the platform, independent of captioning users' viewing patterns. For this research, we define captioning as a textual representation of audio or language, including spoken language, signed languages, and other sounds.

Captions can be either open or closed [73]. Open captions are burned into video content, whereas closed captions can be toggled on and off. See Figure 6.1 for an illustration of the difference. While TikTok supports both open and closed captioning we observed inconsistency in the availability of closed captions during preliminary analysis, with variation over time (e.g., videos appeared closed captioned one day and not another) and across devices and browsers (e.g., at the same time on the same device, videos appeared with closed captions in a mobile browser and without closed captions in the TikTok app). Due to this inconsistency, we scoped content analysis to open captions.

[6]Other accessibility research has also used multiple data sources to gain a more complete picture of the area of study (e.g., [184])

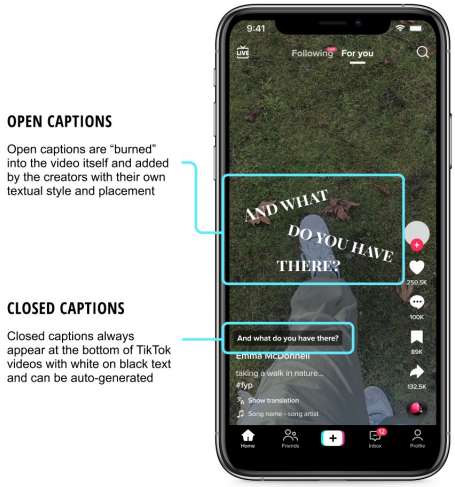


Figure 6.1: Simulated screenshot of a TikTok illustrating the difference between open captions (top text of the video) and closed captions (bottom of video). Closed captions appear at the bottom of a TikTok video as white text on a translucent black background and can be toggled on and off. Open captions can be any color, size, font, and in any location on the screen and are permanently part of a video.

Because TikTok's terms of service prohibit *"us[ing] automated scripts to collect information from or otherwise interact with the Services"* [281], we collected data manually. To collect data, researchers created new accounts and liked or favorited [7] videos that met each data collection phase's inclusion criteria. After liking and favoriting the quota of videos for each data collection round, researchers requested their account's data from TikTok. This resulted in a JSON file containing the links to all videos that a user had liked and favorited. We parsed and combined these files, using the resultant list of video links to form our dataset.

[7]These functions serve equivalent purposes, but we observed occasional malfunctions with the "Like" feature and used "Favorite" as a backup

General Audience Data Collection

To characterize broader trends in TikTok captioning, we sought videos that were likely to be shown to a broad audience. However, TikTok's emphasis on personalized, automatically generated video feeds (a 'For You Page' or FYP), means there is not a core set of videos all TikTok users see. Therefore, we targeted videos that TikTok serves a user it has very little information about, as a proxy for general audience data. To collect this data, four members of the research team generated new TikTok accounts, and, over five days in early February 2023, each researcher liked or favorited 100 captioned videos a day. We scrolled through the research account's FYP, liking a video if it was captioned and scrolling as soon as we determined it was uncaptioned. We excluded ads, live videos, and sponsored posts from consideration but had not yet discovered inconsistencies with closed captions, so we collected both open and closed captioned videos.

We initially collected 2000 general audience videos. Among these, 1654 were unique URLs, signifying unique videos. At the time of submission, 65.3% (1050) of the 1654 unique videos featured open captions, 28.1% (464) had no open captions, and 8.5% (140) had been taken down since initial collection. Having intentionally over-sampled, researchers then randomly selected 150 videos from the set of 1050 open-captioned videos for coding and analysis, informed by sample sizes in prior work (see 6.2.1).

Deafness and Disability Data Collection

We also collected data from communities we hypothesized to be at the cutting edge of video accessibility—Deafness and disability-related content creators. We identified Deafness and disability-related videos via the following five hashtags: #Deaf, #HardOfHearing, #Disability, #Accessibility, and #Neurodiversity. We selected these hashtags to balance gathering videos with a broad focus (e.g., #Disability, #Accessibility) with videos targeting communities likely to use captions to access videos (e.g., #Deaf, #HardOfHearing, #Neurodiversity) [260, 175]. We sought this balance to ensure we had representation from communities that value captioning while not excluding groups we did not consider in advance. To collect this data, we generated five new TikTok accounts, which were used to collect data by the same four researchers who collected the general audience data (the lead author collected data on two accounts during this data collection cycle). Each account was

Table 6.1: Frequency of Deafness and disability data collection hashtags in the final dataset

Hashtag	Count
#Disability	49
#Deaf	44
#Accessibility	30
#Neurodiversity	29
#HardOfHearing	22

assigned a different hashtag to collect data from daily, assigned over a five-day period. We used a Latin Square design to ensure that 1) each research account was used to collect data exactly once from each hashtag and 2) we collected data from every hashtag for each of the five days of data collection.

Over five days in April 2023, researchers searched their designated hashtag on the TikTok 'hashtags' results tab and scrolled until they had liked or favorited 100 open-captioned videos daily. These hashtags often had a high concentration of videos from a small set of creators and sometimes contained irrelevant or offensive content. Therefore, while collecting data researchers strategically avoided liking videos from the same creator to diversify our dataset and excluded content they deemed irrelevant. Researchers were instructed to skip a video if it was 'clearly non-topical', 'ableist mockery', or an ad, and borderline videos were included to be discussed later. The data collection process was designed to collect 2500 videos, 500 per hashtag. However, one researcher's device did not consistently register 'liked' videos and was only able to record 237 out of 500 'liked' videos. We, therefore, collected 2,237 videos, 1,208 of which were duplicates, resulting in a final dataset of 1,029 videos. To match our general audience data, we coded and analyzed a random set of 150 of these videos. The distribution of hashtags in our final dataset is shown in Table 6.1.

Video Content Analysis

We iteratively developed a coding scheme to analyze how videos are captioned. Over three cycles, four researchers drafted a set of codes, applied them to 25-30 videos, and discussed gaps, redundancies, and disagreements before settling on a coding scheme. After the final round of coding, researchers achieved an average Krippendorf's Alpha inter-rater reliability score of .71 on all quan-

titatively analyzed codes. The final coding scheme tracked three key components of captioning: audio and text coverage, style and placement, and caption content.

We then applied our coding scheme to 150 general audience and 150 Deafness and disability TikToks. To diversify our examination of captioning practices, we analyzed only one video per creator. The same four researchers who collected data and generated the coding scheme coded the videos, with two researchers coding each video over the course of two rounds. In the first coding round, each researcher coded 75 open captioned videos (one half General Audience, one half Deafness and Disability Related). Each coder's set of videos was then randomly sorted into thirds and distributed to other members of the coding team. During the second round, each researcher again coded 75 videos. Upon completing both rounds of coding, each pair of researchers discussed and resolved the differences between their coding of the 50 videos they both analyzed. This process produced a single, authoritative coding for each of the 300 videos we analyzed.

We then performed a mixed-methods analysis of our coded data. For quantitative data, we calculated summary statistics and for qualitative data, we open coded responses. To conduct this analysis, we split our coding scheme into thirds and had two researchers analyze each third, with the lead researcher taking part in all analyses.

We also conducted a word error rate (WER) analysis on the videos we identified as containing at least one error. Though we identified errors in 59 videos, only 55 were still posted on TikTok at the time of calculation. For those videos, we transcribed the open captions directly and then manually generated a verbatim transcript of the video. We used the Amberscript implementation of the NIST Scoring Toolkit[8], to calculate the WER for each video and computed the overall average WER. We also calculated the WER for the three videos shown during the interviews.

6.3.2 Interview Study

To complement our TikTok video content analysis, we performed semi-structured interviews with TikTok users who need captions to access the platform, seeking to understand the impact of common TikTok captioning approaches. Following Mack and McDonnell et al. [186], we defined eligibility by captioning use, rather than specific disability, recruiting participants who use captions *"due to*

[8]https://www.amberscript.com/en/wer-tool/

Deafness, disability, neurodiversity, or related condition." We relied on established connections within Deaf and disability communities to recruit participants, reaching out to relevant mailing lists and using snowball sampling.

The semi-structured interviews, conducted over video conference, lasted one hour and had three parts. First, we asked participants to reflect on their current experiences with TikTok and how (in)accessible videos are to them. We then selected three videos, all with more than 500,000 likes[9], from our dataset, which exemplified key aspects of captioning identified in our content analysis. The first video[10] (WER = 6.1%) captioned speech but not background music and used varied caption color, placement, and size. The second video[11] captioned one of the two speakers but did not caption the dog – a salient audio source – or the yelling by the two speakers. Participants were shown the original video (WER = 37.5%) as well as an edited version made by the first author, which captioned all audio (WER = 0.0%). The final video's[12] captions (WER = 0.0%) were formatted standardly but represented a voice-over track not connected to on-screen actions. After participants viewed the videos, we probed for their reaction to certain aspects of the captioning. Finally, the study session concluded with a discussion of what participants would like to see in the future and a comparison of current TikTok norms to other video content. Participants were compensated $40, automatic captions were always enabled, and we arranged ASL interpretation and CART transcription upon request.

We had nine participants in this study. Their average age was 39.1 years old (range 19-73), and five identified as Deaf, three as deaf[13], two as hard of hearing, one as neurodiverse, and two as having some other disability (some participants held multiple identities). With regard to gender, seven participants identified as women and two identified as men. Participants self-reported their race: 66% were white, 11% Black, 22% Asian or Pacific Islander and 11% Native American. We required that participants frequently use TikTok, and 44% reported using the platform multiple times

---

[9] We intentionally selected videos with a large viewership to mitigate privacy concerns

[10] https://www.tiktok.com/@austinandlexi/video/7188243037972106539

[11] https://www.tiktok.com/@bananna_k/video/7198305835943185710

[12] https://www.tiktok.com/@ripleysaquariums/video/7167494942204497157

[13] Capital "D" Deaf often signals identity with Deaf community, whereas lower-case "d" deaf more frequently refers to the audiological experience of deafness [212]

Table 6.2: An outline of the key facets of user-generated open captions that we discuss in Section 6.4.1

Category	Facets to Consider
Audio, Language, and Text	- Types of sounds in videos - Amount of audio covered by captions - Approaches to captioning signed languages
Style and Placement	- Caption timing and animation - Caption placement and alignment - Use of color - Text formatting
Content	- Deliberate non-verbatim transcription - Additional content - Captioning accuracy

daily, 22% reported daily use, 22% reported using TikTok 3-5 times a week, and 11% reported weekly use. We required participants have experience reading captions in English and five also reported communicating using ASL.

We used a mix of top-down coding and reflexive thematic analysis [37, 38] to analyze interview data. Upon completing interviews, researchers reviewed transcripts, flagging data aligned with content analysis findings and taking notes to form a codebook. Data that aligned with content analysis findings was open coded and integrated accordingly. Researchers coded the remaining data in two stages - one researcher completed the initial coding pass and a second reviewed their work. Across this process, the lead researcher reviewed all transcripts. Coded data was then developed into themes using an inductive, semantic, and critical realist approach. Thematic analysis emphasizes the role of authors' positionality. We are a mixed-ability research team, with some members identifying as DHH, neurodivergent, and/or disabled. The lead researcher is a hearing person with conversational ASL skills. Authors identify as white and Asian.

6.4 Findings

We present a content analysis of TikTok captions, highlighting key considerations that go into captioning and integrating participant perspectives on the impact of different captioning choices. We then identify broader themes around the state of TikTok accessibility for captioning users.

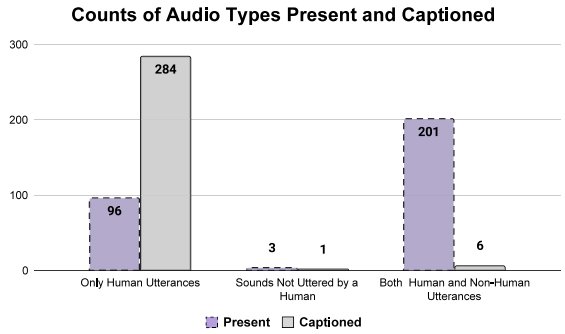


Figure 6.2: Bar chart displaying the number of videos where each audio type was present and videos that captioned each respective audio type. Videos with only human utterances accounted for the majority of captioned videos (94.7%, 284/300)

6.4.1 The Current State of TikTok Captioning

To characterize TikTok captioning, we examine how audio and language are represented in caption text, style and design choices, and caption content. As relevant, we contextualize these video analysis findings with interview participants' perspectives and preferences. Table 6.2 provides an overview of the facets of captioning we analyze in this section.

Audio, Language, and Text

We required videos in our dataset have captions, but creators did not caption all audio equally. Understanding what audio is present in videos and how comprehensively it is captioned reveals what creators prioritize when making a video accessible. Therefore, we focus on *how much* audio is captioned in TikTok videos and how that aligns with participant preferences. We provide an overview of audio types then discuss how human speech, music, non-speech sound, and signed languages were captioned.

We categorized video soundscapes as containing 1) human utterances only (e.g., speech, singing, laughter), 2) sounds not uttered by a human (e.g., dog barking, instrumental music, clapping, ap-

pliance beeping), or 3) a mix of both. Figure 6.2 shows how often each audio type occurred in our dataset and how frequently each type was captioned. Most, but not all, videos (72.7%, 218/300) consistently captioned each audio type. Videos were considered to be captioned consistently in two situations – when *all* of one audio type was captioned in videos that audio type, or if an audio type was present and *never* captioned. For example, if a video contained spoken and sung human utterances and did not caption singing, we considered its captioning inconsistent.

Human Utterances. Human utterances were part of nearly every video in our datasets (99%, 297/300), and were largely captioned (96.7%, 290/300). Most commonly, these captioned human utterances were speech – 85.7% of videos (257/300) contained people talking, and only five did not caption all speakers. This prioritization matched participant preferences: they unequivocally agreed that uncaptioned speech was inaccessible. As P1 put it, *"no captions, and I just scroll past"*.

Music. Though many videos in our dataset contained lyrical or instrumental music, it was often uncaptioned—interestingly, a decision participants supported. Though we could not reliably quantify the presence of music in videos, we observed that it was rare for the presence of music (instrumental or lyrical) to be indicated in captions. When captioned at all, lyrics were often selectively captioned rather than fully transcribed. Largely, participants did not find captioning music to be necessary on TikTok. Many participants agreed that it was *"honestly easier not to know"* (P9) about most music because, on a small screen, *"it just gives more things that I need to read and then it gets frustrating"* (P6). Some participants linked this to their Deaf identity: *"music does nothing for me, I don't understand it—I am profoundly Deaf and have always been Deaf"* (P4). If music is captioned, participants preferred that creators use a music note emoji (P2) or briefly note the tone indicated by the music (*e.g.,* [upbeat music] P8) rather than transcribing lyrics.

Signed Languages. We observed several videos that captioned sign language (13.3%, 40/300), all from our Deafness and disability data collection, though our interview participants reported infrequently coming across videos that captioned signed languages. Signed videos took a wide range of approaches to audio: 22.5% (9/40) captioned no audio and only captioned signing. However, often people signed and spoke content simultaneously[14], interpreted music and TikTok sounds, or used text-to-speech to voice an English interpretation of their signing. However, because our data

[14]Known as simultaneous communication or sim-comming

collection processes sought captioned videos, they likely do not reflect all signed content on TikTok. P6 and P8 both reflected on encountering uncaptioned signed videos as DHH people who know some ASL but do not primarily sign, noting it is *"kind of weird for me, because I'm like, you want me to understand you, but you're going to make me work for it"* (P6). P9 pointed out that captioning signed videos poses a challenge, as there is not a *"standard way to have captions for our language."*[15]

Non-Human Utterances. Only 2.3% of videos (7/300) in our dataset captioned non-human utterances, a stark contrast to how spoken and signed languages were captioned but somewhat aligned with participant preferences. Interview participants were mostly interested in captioning 'important' sounds and sounds that were not obviously visually indicated. Participants stressed the importance of considering the purpose and impact of sound in a video: if *"someone's just making a like, kind of annoying, stupid noise, I don't really need context for the noise they're making"* but captioning relevant sounds *"added flavor to the video"* (P6). Additionally, P9 noted that the TikTok format made environmental audio less relevant than in other media: "if you miss sound on a [TikTok] video, you can still enjoy it, but for movies you are left wondering."

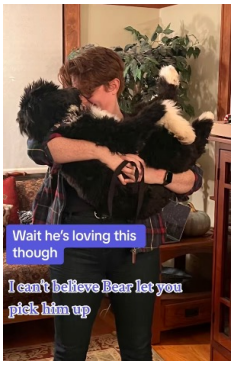
Style and Placement

TikTok captions are notable in their use of a wide range of approaches to style and placement. As P3 put it: *"I think that the captions on TikTok are way way way more creative and people seem to be having more fun with captioning compared to [traditional video platforms]."* We sought to understand how videos in our dataset approached caption timing and animation, placement, color, and formatting. Overall, while there was nonstandard style and placement throughout our dataset, participants preferred captions that prioritized practical access over novel designs.

Timing and animation. When choosing how to time and animate captions, the majority of videos aligned with participants' preferences for captions to be *"static, right there, simple, clean"* (P1). Most videos (83.3%, 250/300) timed their captions similarly to movies and TV: a few lines appear on screen at a time and refresh once all content is spoken. Other timings included captioning speech one or a few words at a time (5.0%, 15/300 videos) and emulating live captions (3.3%, 10/300), with words appearing as they are spoken, building into captioned lines (see Fig-

[15]English language captions can never directly represent signing and only provide a written interpretation [24]

(a) Captions are animated, with words appearing one at a time, initially appearing large and then shrinking to the center of the screen.



(b) Different fonts, color, and caption placement differentiate speakers, stacking captions from a quick exchange on top of each other

(c) Captions appeared a few words at a time, using different sizes, color, and placement for each set of words and contributing to video tone

Figure 6.3: Simulated stills representing aspects of caption style and placement we observed throughout our dataset

ure 6.4). Most participants stressed the need for captions to not disappear *"so quickly that I don't have time to read it"* (P2). P9 provided a differing perspective, noting that while rapid-fire captions are *"not 100% accessible"*, she *"really like[s] it, cause it shows me the way [a TikTok creator is] talking."* Caption rate has long been considered an obstacle to caption readability [139], and user-driven choices around caption timing add another dimension to this discussion. 10.0% of videos (30/300) animated captions in some way, occurring more frequently in general audience videos (15.3%, 23/150) than in Deafness and disability videos (4.7%, 7/150). Common animation styles include fading, bouncing, and erratic motion (e.g., strobing, shaking) (see Figure 6.3). Participants noted that this amount of motion on screen *"can be really jarring"* (P3).

Placement. Despite interview participants' strong preference for captions that stay in one location, over a third of videos (34.3%, 103/300) moved captions around the screen over the course of the video. Variable caption placement was more prevalent in Deafness and disability videos (43.3%, 65/150) than in general audience videos (25.3%, 38/150). Commonly, caption placement was used to differentiate information—for example, separating types of audio (e.g., TTS, laughter, human speech) or contributions from multiple people (see Figure 6.3). Caption movement could also serve

as a meta-structure to organize the video's content (e.g., separating questions and answers, moving from topic introduction to content). However, many videos included seemingly random placement or placement motivated solely by a high-contrast background. Captions were placed in all regions of the screen, with a slight bias toward the top of the video, a departure from established practice [41].

Participants consistently reported problems with poorly placed captions. Often, TikTok's dense UI elements on the bottom and right sides of the screen overlap with captions and make it so *"I can't see those captions"* (P9). In contrast to many captioning standards, P7 suggested that creators should default to placing captions along the top of the video as *"more things are happening on that floor 80% of the screen instead of like the top 20%."* Additionally, participants did not like when the "captions felt far from the action" as their "eyes were doing double work, popping up and down" (P4), a common captioning consideration known as visual dispersion [157]. However, the value of placing captions near relevant visuals comes into conflict with the desire to not move captions around the screen. If captions move throughout a video, P2 noted *"I had to look all around to figure it out . . . If it was all in one place each time, then I know where to look for placement."*

Color. Color choice had a strong impact on caption comprehension, including both the text outline and fill colors. Over 87% (262/300) of videos used black-and-white captions (see Table 6.3), which were preferred by our interview participants. However, 29.3% (88/300) used other color combinations, most commonly white text on a colorful background (15.7%, 47/300 videos) or colorful text outlined in white (7.7%, 23/300). Videos used multiple captioning colors 22.3% (68/300) of the time, which most frequently served to differentiate speakers and sounds or to emphasize specific phrases within the video (see Figure 6.3). Participants' priority for color schemes was that they produce *"simple captions that I can read"* (P5), allowing for more colorful captions to differentiate speakers or sound sources only if readability was the guiding principle. Suggestions included using bright color in the caption background and keeping the text black (P6) or, as P3 suggested, using different combinations of black and white: *"maybe black text with white background for the first speaker and black background and white text for the second speaker, so that way it feels more consistent"*.

Formatting. Videos also leveraged formatting elements such as typeface, size, and capitalization to style their captions, but most deviations from a perceived norm were not well-received by

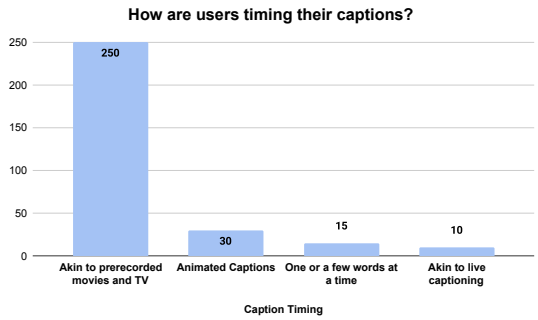
Table 6.3: Frequency of open caption color scheme in our dataset

Caption Style	Example	Frequency
White with Black Outline	Example	54.7% (164/300)
White with No Outline	Example	15.3% (46/300)
Black with White Outline	Example	3.7% (11/300)
Black with No Outline	Example	2.7% (8/300)
White on Gray/Black Background	Example	9.3% (28/300)
Black on White Background	Example	10.3% (31/300)
White with Colorful Background	Example	15.7% (47/300)
Colorful with White Outline	Example	7.7% (23/300)

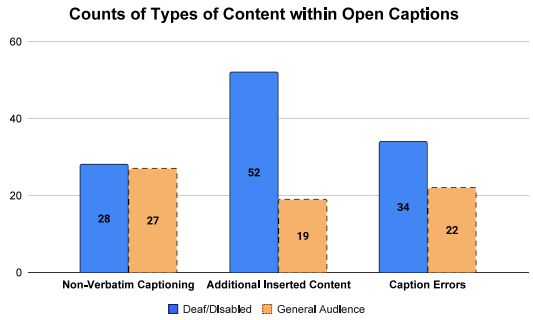
participants. A vast majority (94.7%, 284/300) of videos used the same typeface throughout their video, but variation was used to differentiate video titles from captions, to emphasize the final lines of videos, and to indicate a speaker change (see Figure 6.3). We also observed that videos frequently departed from standard rules of capitalization, using all-caps to emphasize certain words, captioning some videos entirely in all-caps, and sometimes not capitalizing any words in captions. Participants did not find this to be helpful variation: *"I just don't want them changing the style and the font and the letters—that is really hard"*(P5). Though we were unable to consistently quantify trends in font size, we observed a high degree of variation. Size changes could be meaningful, differentiating an important phrase from the rest of captions, or somewhat randomly scaled with regard to the amount of text on screen. Participants suggested that having the font *"large enough to be able to read"* (P2) is critical to readability and noted that if they come across "captions that are like, tiny ... I can't read that ... I'd swipe through [and skip the video]" (P8).

Considerations Around Caption Content

While the above two subsections focused on how audio was translated into captions and how captions were styled and placed, here we consider the content of captions themselves. Going beyond the presence of captions and considering their content, we analyzed instances when language was



(a) Users are primarily (250/300) timing their open captions similarly to precorded movies and TV (a few lines at a time). Displaying captions one word at a time was less common (15/300), followed by 10/300 videos captioning similarly to live captioning, where words are shown as they are spoken and accumulate into a few lines



(b) Breakdown of frequency of occurrence for three types of content in open captions: non-verbatim captioning, inserted content, and captioning errors. Inserting additional content into open captions was most common, especially within the Deaf or Disabled dataset (52/150)

Figure 6.4: Bar charts visualizing caption timing and patterns of notable content changes among general audience and Deafness and disability-related videos.

not captioned verbatim, when additional content was added, and when the captions had errors (see Figure 6.4).

Deliberate Non-Verbatim Captioning. Though the Deaf community has long advocated for verbatim captions [138]—as opposed to summarized or censored captions—our interview participants had nuanced, context-dependent perspectives on non-verbatim, user-generated captions. In our video analysis, we found that 18.3% (55/300) videos deliberately used non-verbatim captions, most often to caption a curse word or other vulgarity (e.g., "shit" captioned as "sh*t") though sometimes to replace content that was not obviously a censorship target (e.g., "autism" captioned as "the 'tism"). Popular strategies for altering words included using asterisks or other punctuation in place of vowels, removing letters from words, replacing the spoken word with an alternative (e.g., "fucking" captioned as "friggin"), and using acronyms, abbreviations, and emoji in the place of fully voiced words. Notably, these instances of non-verbatim transcription were limited in scope, often impacting single words in videos. Participants largely echoed P6's reaction that they *"haven't come*

across censored captions too much", but that *"it's kind of just like something you have to deal with."* Breaking with a long tradition of strong opposition to censored captions [303], many participants shared P9's sentiment that, while *"I don't like it when they [censor content] . . . I understand the creator's reasoning."* Participants still disliked the ways that censorship feels like *"you're treating me as if I'm less than or as if I'm fragile because I can't hear"* (P6). However, they considered that audiences contain *"young people too"* (P8) and that creators may need to protect their content on a platform prone to censoring videos [240]. In fact, P3's initial reaction to much of the non-verbatim content she saw on TikTok was that it *"feels like I'm getting older"* as she noticed patterns of captioning that *"kind of became a language and a culture to get around the censors."* Overall, while non-verbatim captions provide lower-quality access, participants took a nuanced view, understanding them as part of platform culture in the face of censorship and shadowbanning.

Adding Content to Captions. In our dataset, 23.7% (71/300) of videos added content to captions beyond direct transcription, often communicating paralinguistic aspects of speech (e.g., tone). This occurred more frequently in Deafness and disability videos (34.7%, 52/150) than in general audience videos (12.7%, 19/150). Most commonly, videos included emojis, frequently used to indicate the tone of the spoken content (e.g. 😭, 🥰, 😤) or to match the topic of the video (e.g. 🦀, 🐟, 🐙 for a video using the song "Under the Sea"). Participants largely liked emoji additions but emphasized that while sparing emojis can *"help me understand mood and the perspective,"* excessive use is *"a little bit cringey"* (P6). P4 likened emojis that matched the tone of a caption to non-manual markers, a key component of ASL grammar that often serve as a tone modifier. While many videos entirely omitted punctuation, when used, punctuation helped to differentiate types of content (e.g., indicating that *whispers* was a tonal description, not a captioned word), to convey volume or emphasis (e.g., using !!! and !?), and to convey the pace of speech in captions (e.g., a caption that reads "It's just . . . I'm").

Captioning accuracy. In our analysis of TikTok captioning accuracy, we found that captions were largely accurate—which also reflected our participants' experiences. We identified at least one error in 19.7% of videos (59/300), with errors in 24.0% (36/150) of Deafness and disability-related videos and in 15.3% (23/150) of general audience videos. The average word error rate

(WER) among videos with at least one error was 7.9%, ranging from 0.5% to 35.7%[16]. Error types included word substitutions (e.g., "old on" instead of "hold on", "rep saint of" instead of "Representative"), deletions (e.g., captioning "what's great" as "great"), and insertions (captioning "got her dressed" as "got it her dressed"). Overall, participants reported noticing errors in TikTok captions but largely agreed with P2's assessment that *"there's always going to be some words that are missed or incorrect, but you basically get the overall content, and you're able to follow."* Errors did still impact participants' experience, as participants skipped videos with highly inaccurate captions and stressed that when captioning *"doesn't have as many spelling errors and word choice errors, I'll have fewer misunderstandings"*(P5).

6.4.2 Participant Experiences with and Desires for User-Captioned Content

While specific facets of captioning, as we explored in Section 6.4.1, are crucial to video accessibility, participants also reflected broader factors shaping their use of TikTok. Here, we highlight findings on the impact of changing caption norms, perspectives on accessibility on TikTok, and desires for the future. These findings draw entirely from participant interviews.

Changing Captioning Norms

Participants reflected on the impact of two changing norms in captioning: the use of automatic captioning and a shift toward open captioning.

Participants felt that automatic captioning tools produced passable TikTok captions, though they did not match the gold standard of human captioning. Since TikTok rolled out automatic captioning in spring 2021 [219], participants reported significant increases in access: *"I do feel like now with the automatic captions, almost all videos are accessible"* (P9). There was, however, still a perceived drop in quality. P5 expressed his desire for *"not the automated, not the kind of robotic one, but the person, the live person doing the captioning."* Many shared P6's experience that, when viewing automatic captions, *"you don't get the full context of what they're saying, but you kind of have, like, a broad spectrum of what they're saying."* Participants stressed that not all automatic captioning errors have the same impact. Despite usually being able to guess at errors, P8 recounted once

[16]This analysis was conducted over 55 videos, as four were unavailable at the time of calculation

spending *"half the video"* trying to make sense of a single error—the name of the subject of the video—and concluded that *"how disruptive [an error] is is not absolute."*

Closed captioning is usually considered the best practice for captioning a video, but many participants preferred the shift toward open captioning because embedding captions in a video makes it durably accessible. Some participants primarily watched TikToks off the app and, therefore, only had access to open captioned TikToks, as closed captions are an app-specific feature [121]. While some had a hard time adjusting away from closed captions, *"what I've used most of my life"* (P6), others valued that open captions were *"non-turnable-offable"* (P4). P9 highlighted the cultural shift that open captions represented: *"[closed captioning] feels like an assistive tool instead of a complete experience, which the embedded captions do feel like."*

Nuances of Platform Accessibility

The kind of content that participants consumed significantly shaped their access needs for that video. If videos primarily contained speech, participants needed captions: *"if people are just standing there talking to each other or to the camera . . . then I need to know what the words are"*(P5). However, participants still reported enjoying a substantial body of uncaptioned content. P4 explained that *"if it is more action, and show rather than tell"* then uncaptioned videos were still worth watching (e.g., gymnastics (P1), cooking (P3), animals (P2, P4, P7)). Notably, participants had varied interests and watched a diverse range of TikTok content, emphasizing that it is important that all kinds of content are made accessible. As P3 put it,*"if the video is not captioned, I'll just be like 'hmmmm I guess this person doesn't care about us."'*

When participants perused TikTok individually, uncaptioned videos proved to be less of an access barrier than when engaging with TikTok socially. Participants attributed the relatively high amount of captioned content they were shown to TikTok's algorithmically mediated nature, which they understood to *"keep your preferences . . . so you can use that algorithm to watch things you like"* (P9). This curated view led P8 to reflect: *"I think it is quite accessible . . . but I think I could argue that it is inaccessible if there are videos that don't have captions on them, I just don't see many of them."* TikTok's endless scroll design made it so that it was usually easy for participants to just *"skip any videos without a [captioning] option"* (P3), but this became complicated when view-

ing specific videos sent by family and friends. In P4's experience, sharing and discussing videos is *"part of the social engagement nowadays"* and this interaction breaks down if the videos being shared are not captioned. P2 shared that when her siblings send her uncaptioned TikToks she'll respond *"Hello! – I'm Deaf"*, to which they reply *"oh, right, sorry"* and then explain the content of the video.

Participants did not all feel similarly about the volume of inaccessible content on TikTok. Some felt that *"TikTok videos are not like a 'need' thing ... it's free entertainment"* and therefore did not take issue with the fact that not everything is accessible since *"you will find something eventually"* (P7). However, others recounted that their reaction to inaccessible content was *"a sense of resignation"* because *"it's frustrating, honestly. It means that Deaf people are falling further and further behind"* (P5).

Desires for the Future

When considering what they'd like to see on TikTok in the future, participants had one overarching desire: *"I would love it if every video was captioned"* (P8). P1 envisioned this world: *"I would love to wake up in the morning and just go, 'oh, I can tell what's going on'."* Toward that goal, participants considered how creators could better prioritize access, ways to integrate captioning standards and guidelines into the platform, and opportunities for customization.

Participants highlighted the ways that individual creators' choices shape video accessibility and proposed ways to improve norms on the platform. P3 reflected that, while TikTok creators often start with some knowledge of captioning, there is room for improvement: *"I think people are so much better at captioning their videos, but they're still learning to caption it in an accessible way that is also enjoyable."* Many saw creators' investment in captioning as a way to win their viewership: *"my time, time is valuable right? And, basically, I'm going to give the reward to watch something to someone who's invested time to make it accessible"* (P2). When considering how to move toward more captioned content, P5 reflected that, rather than a technical approach, creators should *"maybe just listen to us, I guess"* and prioritize including Deaf and disabled viewers. Recognizing that creators are key stakeholders and that captioning is effortful, participants proposed ways that TikTok's design could encourage and support creators in adding captions. P6 imagined adding

a way to contact creators about their video accessibility, hoping that direct feedback would help creators realize *"this would really help. And then it doesn't get lost."* P7 also envisioned that TikTok could help teach creators how and why to caption videos: *"whenever they are posting something, they can have like a prompt . . . 'do you wanna caption the video' or. . . benefits of captioning."* Multiple participants also noted that TikTok, as a platform, could build in automatic captioning by default, making it so that if higher-quality creator-generated captions were not available, the video would retain a modicum of accessibility.

Across the board, participants noted the lack of guidelines for TikTok captions. P2 contrasted the state of most captioned media to TikTok: *"when it comes to the captioning industry, there are rules, there are standards, and they know what they are. But TikTok, it's wide open, anything goes. It's an open source."* Participants suggested that there should be a way *"to clarify some rules"* (P5) for captioning in a way that would still *"allow for a little creativity"* (P2). For P3, this looked like building guidance into the tools creators use to make videos:*"it would be really fun if everyone had a selection of captioning styles to choose from that they know would be really accessible . . . [and] some technology to tell them 'hey, your captions are overlapping this and that, let's move them to a different place'."*

Participants also imagined that the platform would become more accessible if captions were customizable. P1 reflected that, when watching streaming television, *"you can actually pick your own background and color of the [captions]—that's really awesome."* Others noted their experience with platforms like Zoom, where they can *"drag [captions] and move them anywhere on the screen"* (P3). P6 stressed that TikTok has the opportunity to not have to make captioning a *"one size fits all"* experience, and that customization would lead to a more accessible experience. P7 believed that being able to change the color, resize, move, or turn off captions was also key to improving the user experience of TikTok. This customizability could extend to being able to configure settings that instructed TikTok: *"don't even bother to send me things that are not, you know, captioned or whatever"* (P1).

6.5 Discussion

Our findings highlight relevant factors to consider when assessing how a user-generated video is captioned and point to a need for greater standardization of user-generated captions. We, therefore, discuss steps toward a captioning standard for user-generated social video, consider the future of user-driven captioning, and envision how disability justice concepts can help guide future user-generated captioning efforts.

6.5.1 Steps Toward a Captioning Standard for User-Generated Content

Having analyzed the current user-generated captioning practices in our TikTok dataset, we compare these unregulated, user-generated approaches to formal captioning standards. Participants frequently made sense of TikTok captions in relation to their understanding of standard practices, indicating that participants' preferences for captions were strongly influenced by such standards. We therefore compare our findings on the current state of TikTok captioning with an established standard–the Described and Captioned Media Program (DCMP)–, as an exploration of what a future user-generated captioning standard could consider. Because standards vary internationally and have been shown to shape geographically-specific captioning preferences [238], we compare our findings from English-language videos and interviews with US-based participants to an American standard. While the Federal Communications Commission (FCC) is the US regulatory group that controls captioning, their guidelines are broad, focusing primarily on *"accuracy, synchronicity, program completeness, and placement"* [56]. However, incorporating FCC rulings and a wide body of research, the Described and Captioned Media Program (DCMP), a project by the National Association of the Deaf, has developed a comprehensive set of standards known as the Captioning Key [70]. The DCMP's level of detail allows us to identify specific areas where the user-generated practices we observed align with or diverge from a respected standard. Their guidance is also applicable to both open and closed captions, in line with current practices on TikTok. In the following list, we compare our findings, as relevant,[17] to the DCMP's major sections—text, language mechanics, presentation rate, sound effects and music, speaker identification, and special considerations.

[17]The Captioning Key is targeted toward captioning creators, so not all recommendations are relevant or information we tracked in our analysis

- **Capitalization.** DCMP standards recommend mixed-case capitalization (e.g., "My dog and I played fetch.") except to indicate shouting. While not the norm, we observed videos that used no capitalization, mixed case within words (e.g., WoOoOo), and all-caps regardless of sound volume.

- **Typeface and Color.** The DCMP narrowly recommends captions use the same typeface and use white text over a translucent box. We found limited typeface variation, but a wide variety of combinations of black and white, as well as multicolored captions.

- **Caption Rate.** Per the DCMP, captions should be a minimum duration of 40 frames (slightly over one second) and a maximum of six seconds. Caption rates should also stay between 130 and 160 WPM. While we did not quantify caption duration or speed, participants reported that TikTok captions felt too fast, and we observed captions that updated with each word. Prior work finds that captions are maximally readable at 145 words per minute, but that this varies with a person's experience using captions [139].

- **Caption Placement.** Standards recommend placing captions at the bottom or, as a backup, at the top of the screen, moving captions left to right to identify speakers during dialogue. We observed captions moving across the entire screen with no clear norms for how placement can differentiate information.

- **Punctuation.** The DCMP stresses adhering to formal punctuation rules, but we observed both a lack of punctuation and creative use of punctuation.

- **Censorship.** The DCMP explicitly instructs creators to caption profanity and slang verbatim. We observed some non-verbatim captioning, and findings suggest that, on social video sharing platforms that censor videos, captioning guidelines must account for the fact that creators' choice to generate non-verbatim captions can come not from paternalism toward captioning users, but from material consequences due to platform censorship practices [240].

- **Sound Effects.** Standard captioning of sound effects does not caption all non-speech sounds but prioritizes those necessary to understand or enjoy a video. This aligned with partici-

pant expectations but not with trends in our TikTok data—very few videos captioned non-human utterances. The DCMP provides a format for captioning sound effects or describing the quality of sounds (e.g., [whispering]), but we observed a greater variety of approaches (e.g., *whispering*).

- **Music.** The DCMP requires that instrumental music be described only when it is essential to understanding the video but suggests that music lyrics should always be captioned. We observed that music, instrumental and lyrical alike, was rarely captioned in our datasets, but participants did not identify this as a problem. They did not want the additional cognitive load of descriptions or transcription of music that was not vital to their understanding of a video. If music was captioned, participants prioritized mood descriptions over transcription of lyrics.

- **Speaker Identification.** The DCMP suggests captions identify speakers by being placed underneath the current speaker and to identifying each speaker by name, but, in our datasets, captions were more likely to use different colors than names to differentiate speakers and varied placement both vertically and horizontally.

Ultimately, we find that any future guidelines for user-generated captioning should build from traditional standards, with key points of departure. While current standards stress formal mechanics of grammar, language, and punctuation, the looser standard we observed seemed to be appropriate for the tone of videos and was not a notable accessibility barrier. Regarding font and color, the current state of color use on TikTok often resulted in less readable captions, but participants agreed that a greater range of caption color than the DCMP's recommendation could be useful, if readability is prioritized. Although current standards recommend captioning all music, smaller screens and a different artistic role of music in user-generated social videos suggest that music should be captioned sparingly to lessen cognitive load. Our small set of participants' initial reaction highlighted that captions on TikTok are displayed too rapidly, and future work should explore both an optimal captioning rate and presentation style, taking into account the impact of varied literacy, hearing status, and experience using captions. Finally, the algorithmic censorship of videos on platforms like TikTok raises questions about verbatim captioning, and future standards may consider what kinds of non-verbatim captioning methods preserve information access without risking content removal and

shadow-banning [240, 152]. Recent work by Klug et al. [152] found that TikTok creators largely use non-verbatim 'algospeak' to evade algorithmic consequences, suggesting that future user-generated captioning standards must account for the content moderation behaviors of video-hosting platforms. Our findings demonstrate a need for guidance to ensure that user-generated captions successfully extend video accessibility, and we present this comparison as a first step toward shaping future standards.

6.5.2 Toward The Future of User-Generated Captioning

The videos in our dataset are representative of a new era of considerations for captioning: they are open-captioned by users engaged in internet culture. This poses new considerations for captioning design and research, namely how to engage video creators, who to study as captioning users, and how to systematically study open captions.

Traditional captioning tool design either assumes captions will be generated by professional CART captioners (e.g., [149, 157, 135]), or by automatic speech recognition-based tools (e.g., [198, 45]), and therefore does not consider the needs of non-expert captioners. Video creators are fundamental to the existence of user-generated captioning, and our findings reveal many avenues for future change that require significant effort from video creators. Future platform design should consider ways to both incentivize and enforce high-quality captioning, and future work needs to engage video creators in the design of those systems.

Captions have traditionally been studied as a tool used by DHH people (e.g., [157, 59, 31]), but recent research has emphasized that other disability communities, particularly neurodivergent people, also use captions to access audio/video content [260]. Correlating assistive technology use with a single disability group thus misses the perspectives of these other potential users [186]. Further, even within Deaf and hard of hearing captioning users, preferences and experiences can differ [158, 276]. To account for these varied users and experiences, we explicitly recruited "captioning users" broadly rather than focusing on a specific group such as DHH participants. However, all but one of our participants identified as DHH, which means that we were unable to explore tensions among the needs of different groups of captioning users—an important direction for future work. Analogously, past work has found that users of alt text, another user-created digital access tool, have

a variety of preferences [182, 268]. Researchers and designers have begun to propose approaches to alt text provision that meet varied needs, namely customization. In a similar vein, understanding and including the needs of all people who use captions to meet an access need is crucial to ensuring an inclusive future of captioning design.

Finally, the shift toward open captioning produces a new set of considerations when assessing captioning quality. Recently, HCI captioning researchers have emphasized the importance of metrics to understand and improve caption accuracy [144] and placement [10]. The set of features we analyzed (see Table 6.2) could serve as a step toward a structured analytical tool for understanding the quality of open captions. Such a tool could support future researchers in assessing factors beyond accuracy, holistically encompassing elements of audio coverage, design, and captioning content, which are necessary to consider when engaging with open-captioned videos. These features could also be useful in creating future tools to guide non-expert caption creators in making considered decisions when generating new captions or understanding the state of their past content.

6.5.3 Disability Justice and TikTok Access

Accessibility legislation and research overwhelmingly focus on access to critical or educational information, often to the exclusion of entertainment or content deemed less important. Within HCI captioning literature, research overwhelmingly focuses on access to education (e.g., [157, 149]), work (e.g., [188, 29]), or informative media (e.g., news [59], education [31]). However, while a few participants used TikTok for informative purposes, most recounted enjoying watching silly pet, cooking, and trending dance videos. In fact, some participants wrestled with the idea that something that is "*not like a 'need' thing*" (P7) ought to be accessible. We argue that ensuring accessibility to content, even when it does not fulfill a specific need, is essential—all people deserve access to idle entertainment and the ability to participate in the *"social engagement"* (P4) of sharing and discussing silly videos.

For user-generated content to become accessible content, we argue that creators must embrace principles of disability justice, particularly collective access and interdependence. The disability justice principles of collective access - accessibility is a group, not individual responsibility - and interdependence - that we all rely upon each other to navigate the world - articulate a world where

everyone is responsible for considering how to extend access to others [131]. Prior work has often been motivated by the idea that not all online videos will be well-captioned (e.g., [258, 31]). We argue that by adopting a lens of collective access and interdependence, we can imagine a world where high-quality captioning is seen as inherent to user-generated video content and focus future efforts on building tools that help realize that world.

6.6 Limitations

Our study has a few key limitations. First, although we reached saturation while analyzing 300 videos, this represents a fraction of the videos uploaded to TikTok every minute. The type of mixed methods analysis we conducted does not scale indefinitely, and future quantitative analyses of TikTok captioning at scale could complement this work. Second, we scoped our dataset to English language videos because of our research team's fluency in English. All participants were also based in the US. Future work should examine how captions and caption users' perspectives vary in non-English language and international contexts. Next, we focused on interviewing individuals who already use TikTok and therefore cannot address whether viewers who need captions to access videos find TikTok to be accessible overall. Our participants considered TikTok to be accessible enough to be enjoyable, but we cannot speculate whether this perspective holds universally. Further, we intentionally defined our recruitment criteria based on use of captions to meet an access need, rather than a specific d/Deaf or disability identity. However, all but one participant identified as DHH. Future research should seek to have greater participant diversity. Finally, nine participants is a small sample. Future research, particularly work exploring a standard for user-generated captioning, should seek to validate our findings with a much larger participant pool.

6.7 Conclusion

As the world's most downloaded app in 2020, [216] and a massive repository of user-generated video content, TikTok provides an exciting opportunity to understand current trends in user-generated captioning and explore how those captions impact the many caption users viewing TikTok daily. We conducted a content analysis of 300 TikTok videos, evenly distributed between general audience and Deafness and disability datasets, and interviewed nine frequent TikTok viewers who rely on

captions. Our findings reveal that current TikTok captioning practices facilitate access but could be improved, perhaps with the aid of a user-generated content-specific captioning standard. This work contributes the first empirical understanding of the state of captioning on TikTok, and provides approaches to advancing toward a world with universal captioning for user-generated content.

6.7.1 Within-Dissertation Contextualization

In this chapter I identified trends within captions on TikTok, representing norms specific to the environment of the TikTok platform and social practices of TikTok creators. By studying a different context of caption use, I demonstrate that it is not only during real-time conversation that communicators' choices impact how intelligible they are to caption users. Arguably, TikTok is already a platform that sees collective access approaches to captioning, though future work is needed to identify creators' motivations and investigate how they can be guided to adapt more accessible standards, such as the steps we outline in our discussion of a future user-generated captioning standard. Envisioning TikTok creators as conversation partners of their viewers who rely on captions to access content is also an exciting step for me in broadening our understanding of collective communication access to include many digital accessibility efforts, not only synchronous, spoken conversation.

Chapter 7

THEORETICAL LITERATURE REVIEW

In this chapter I present the other half of the work submitted as the paper *"Envisioning Collective Communication Access: A Theoretically-Grounded Literature Review of Captioning Literature from 2013-2023"*, coauthored with Leah Findlater and forthcoming at ASSETS '24. This chapter contains a review of the last decade of captioning literature through the lens of my collective communication access framework (see Chapter 3). Note that this review includes analysis of two of the previous chapters (Chapter 4 and Chapter 5), as they fit the parameters of our literature review. While the work presented in Chapter 6 meets all other criteria, it was not officially published until 2024, making it out of scope of this review.

In this chapter we analyze published captioning research with regards to how it, among other things, engages communication context and takes a collective access approach. By doing so, we identify that papers engaged with communication context and using a collective communication approach do indeed better match captioning's realistic context of use and are more engaged in disability and Deaf community politics. This chapter contributes to answering DRQ1 and DRQ2 by exploring socioenvironmental factors and collective access approaches in other published research.

7.1 Motivation

Rapidly changing communication technologies and digitally mediated communication bring both new accessibility barriers and new opportunities to make communication accessible (e.g., [53, 183]). Human-computer interaction (HCI) researchers have done significant work in this area, with 16% of accessibility research papers between 2010 and 2019 focusing on communication access [184]. Yet, as accessibility research matures as a field and begins to integrate more interdisciplinary critique (e.g., [189, 25, 297]) we identify an opportunity to reflect on how we study communication and what perspectives are or are not included in current practice.

In this chapter, we draw upon a framework of *collective communication access* (see Chapter 3),

derived from thinking in disability studies, Deaf studies, disability justice, and communication studies. We identify how, if at all, a collective communication access approach manifests in one specific body of HCI communication access work: captioning research. Captioning, or representing language or audio in written text, is a technology that has undergone significant change in recent years. Automatic speech recognition (ASR) is newly viable [290], videoconferencing platforms now make automatic captioning omnipresent [273, 274, 275], and user-generated videos are increasingly captioned by their creators [57, 46, 260]. Furthermore, captioning is designed for d/Deaf and hard of hearing (DHH) people but, as a tool primed to support spoken communication, is less tightly connected to Deaf culture than signed languages such as American Sign Language (ASL) [164]. Given the rapid evolution of captioning technologies and less-determined connection to political understandings of accessibility, we chose to analyze captioning literature published between 2013 and 2023 to identify how it aligned with or differed from our collective communication access framework.

In this chapter, we answer the following question:

1. What is the state of last decade of human-centered, accessibility focused captioning research and how, if at all, does it align with that framework?

We applied our framework for collective communication access to the 73 human-centered accessibility-focused captioning papers published at ACM-, SIGCHI-, and SIGACCESS- sponsored venues between 2013 and 2023. We find that most captioning research published in this timeframe does not take a collective communication access approach, though the field has quietly enacted some activist demands. Most published papers demonstrate a focus on improving access by improving technologies, aiming to change neither DHH nor hearing people's current practices, and implicitly situate DHH people as experts on captioning technologies. Yet, the small body of research aligned with collective communication access demonstrates that engaging conversation partners in developing more accessible norms better matches captioning's realistic context of use, enacts a more political understanding of accessibility, and identifies novel directions for technology development.

In this chapter we contribute 1) a framework and rubric for collective communication access, 2) critical reflections on the past decade of captioning research, and 3) directions for future collective communication access research.

7.2 Related Work

Much of the related work that guides this chapter has been discussed in Chapter 2 and Chapter 3, but we highlight one body of work uniquely relevant to this chapter: work focused on literature reivews in HCI accessibility.

7.2.1 Literature Reviews in HCI Accessibility

The key contributions of this chapter is a structured literature review of the last decade of captioning research, a method that is increasingly popular within HCI accessibility research.

Structured reviews of HCI accessibility literature can either take a broad focus on the field or track a specific disability group or kind of technology. Mack et al. [184] published the first comprehensive review of accessibility research, identifying that the field is growing rapidly and unevenly. For instance, 43.5% of papers from 2010 to 2019 focused on blind and low vision people and all other groups were significantly less studied. Other reviews have aimed to understand how specific populations have been studied within accessibility research, including children with 'special needs', [21], blind and low vision people [43], autistic people [179, 229, 265], older adults [286], and people with ADHD [267]. Notably, prior systematic reviews have not focused specifically on technologies designed primarily for DHH people, the focus of this chapter.

Structured reviews tend have two goals: to summarize and/or critically reflect upon a trend within the field. For example, Brule et al. [43] focus on the kinds of technologies studied and research methods employed with people with visual impairments, aiming to identify best practices for quantitative, empirical evaluations with this group. Vines et al. [286], on the other hand, perform a discourse analysis on literature around ageing, identifying the often-stereotypical discourses around older adults and age that are embedded in HCI research and drawing on critical gerontology to set directives for future work. We opt to use this method in a manner similar to Vines et al. [286], reading the last decade of captioning research through a lens of critical perspectives on disability and communication.

7.3 Methods

Through our theoretical synthesis we developed five tenets of a collective communication access approach. To assess how these tenets appear in HCI communication accessibility research, we developed a rubric with to apply to research papers. We used this rubric to analyze one specific body of communication accessibility technology research: captioning research. We identified a dataset of all captioning research published between 2013 and 2023 that took a human-centered, accessibility lens on the captioning of audio into text. We then used our theoretical rubric to code all 73 papers in this dataset. We do not report quantitatively on scores generated by our rubric, but rather used it to structure a qualitative analysis of captioning literature.

7.3.1 Identifying Relevant HCI Captioning Papers

To systematically assess the current HCI accessibility approach to communication and how it aligns or diverges from a collective communication access approach, we defined a dataset of relevant recent research. We scoped our analysis to a specific communication technology frequently studied by HCI researchers: captioning. We chose to focus on captioning because it is a common form of communication access that is necessarily technologically mediated and has been studied extensively within HCI literature and in myriad contexts, including synchronous communication. We reviewed recent research, scoping to the most recent decade of published research, a practice that has been used in other reviews of accessibility literature to study recent trends in depth [184].

While captioning has many forms, we opted for a permissive definition that still maintained focus on using captions as an assistive tool. There are several terms, often used interchangeably, to refer to the practice of representing audio in text, notably: “captions”, “open captions”, “closed captions”, and “subtitles”. By studying captions generally, we include both open (i.e., captions burned into a video) and closed (i.e., captions displayed in an overlay that can be toggled on and off) captions [70]. While subtitles are often seen as synonymous to captions, they are relevantly different. Within a North American context, captions are intended to serve DHH audiences who cannot access underlying audio, and therefore contain speech transcription and audio information (e.g., ‘music’, ‘wind blowing’, [softly]). Subtitles, in contrast, are intended to provide linguistic access to those who are not fluent in English and provide only speech transcription [291]. We therefore did not

include subtitles within the scope of our study. Captions can also serve different roles: access to synchronous conversation or to rerecorded media. While these contexts of use are different, there are many shared considerations. For instance, research on how to best style and place captions for real-time videoconferencing can draw heavily from research into television captions, and vice versa. Further, while it may be less obvious how collective communication access would apply to research into prerecorded media, prior work finds that choices made by people generating captions for prerecorded media can impact caption utility [181, 260, 175]. Therefore, we included captioning research for both synchronous conversation and prerecorded media in our literature review, noting differences and similarities between the two contexts when relevant.

We first searched the Association of Computing Machinery's Digital Library[1] for all entries in the ACM Full-Text Collection that contain the stem "caption*"[2] in the title, abstract, or keywords published between 2013 and 2023. Note that, while this spans 11 years, we refer to it as the past decade of captioning research. Our initial query returned 765 results, many of which focused on developing a text summary of video content (known in computer vision as a caption), or on image captioning techniques to facilitate access to images for blind and low vision people. Therefore, the first author manually reviewed all initial results, deeming them relevant, irrelevant, or questionable. *Relevant papers* were defined as original research that 1) study captioning in the context of representing audio in text, 2) study captions as an accessibility tool, and take a 3) human-centered approach. We defined a human-centered approach to captioning research as one that studied caption use or user-facing design, and excluded papers that solely focused on generating captions (including both algorithmic and crowd-computation methods). We also scoped our literature review to publications written in English, as it is the only language all authors read fluently. Publications that did not present original, peer-reviewed research were also excluded, including workshop abstracts, doctoral consortia, newsletter or magazine articles, and student design competition entries. We included conference papers, posters, extended abstracts, journal papers, and other reviewed formats.

After a first pass, we discussed and further narrowed down the dataset. Throughout this process, we applied a high threshold to deeming a paper irrelevant and a relatively low threshold to deeming

[1] https://dl.acm.org

[2] this ensured that the search would return results that contain 'caption', 'closed caption', 'captioning', 'captioned', 'captioner', etc.

papers relevant, as they would be re-reviewed during coding. Initially, the first author identified 122 papers as relevant, 31 papers questionable, 26 of an excluded format and 586 papers irrelevant. We lightly discussed the initial relevant and questionable set, updating inclusion and exclusion criteria to be more specific about what constituted 'human-centered' and filtering for papers whose full text was written in a language other than English. After the second pass over relevant and questionable papers, the first author identified 97 relevant, 14 questionable, and 42 irrelevant papers. We then met to discuss the remaining 14 questionable papers, settling on a dataset of 101 relevant captioning papers. We then reviewed the venues these papers came from, including only papers that were published at ACM[3], SIGCHI[4], or SIGACCESS[5] sponsored conferences. This resulted in a candidate dataset of 93 publications, which we further narrowed in the process of applying the theoretical framework to a final dataset of 73 papers (see subsection 7.3.2 for a full description of this process).

7.3.2 Applying the Theoretical Framework

To analyze relevant captioning papers, we developed our framework for collective communication access (see section 3.2) into a rubric. Inspired by Williams et al.'s [297] rubrics for counterventional criteria, we translated our theoretical framework into a tool for our own analysis and to enable others to more easily take up a collective communication access approach in the future. In this rubric we identified what it would mean to strongly meet, strongly fail, or land in the middle for each tenet of collective communication access. When appropriate, we also identified what would make a criteria non-applicable (e.g., criteria A is not applicable to research that does not study interactive communication). Taken together, these constitute a rubric for collective communication access, shown in Table 7.1.

The first author coded 15 papers using a rubric draft and then the second author reviewed a random subset of five of those papers. Authors met and discussed differences to further refine and finalize the rubric, at which point the first author analyzed the rest of the dataset by coding each paper using the final rubric. While this coding step may resemble more quantitative coding meth-

[3]https://dl.acm.org/conferences, https://dl.acm.org/jouranls

[4]https://sigchi.org/conferences/

[5]https://www.sigaccess.org/assets/related-conferences/

Criteria	Fails to Meet	Neither strongly meets nor fails to meet	Strongly Meets	Not applicable
A. Approach to Collective Access	Research treats the Deaf/disabled person as the only party that needs to consider communication access	Research recognizes the impact of nondisabled communication partners on communication access, but does not engage or design for them	Research engages all communicators, Deaf/disabledand nondisabled alike, as having an impact on communication access	Research does not study interactive communication (e.g., perspectives on algorithmic confidence measures)
B. Target of Change	Intervention places the burden of change on Deaf/disabled people to change towards nondisabled norms	Intervention tries to support access into unchanged social worlds by putting the burden of change on technologies	Intervention puts the burden of change on nondisabled people to adapt to Deaf/disabled people's current practices	The paper does not describe an intervention (e.g. foundational empirical research)
C. Historical Harms	Research is actively engaged in perpetuating historical harms (e.g., has explicitly oralist, eugenicist aims)	Research does not perpetuate historical harms, but does not actively challenge or acknowledge them	Research sets out to challenge historical harms	
D. Whose Knowledge and Expertise	Proxies or researcher expertise stand in for the perspective of the impacted community	Paper includes perspectives from impacted communities, but does not frame this as a way to center marginalized expertise	Paper uses participatory/member researcher methods to center and empower the impacted community	Not user research
E. Considering Context	Research does not account for the role context plays on communication in either study design or results	Research considers the role of context, but does not account for it in either study design or results	Research centers the role context plays in communication access, in both study design and results	Research does not study interactive communication (e.g., perspectives on algorithmic confidence measures)

Table 7.1: Our rubric, derived from the theoretical framework presented in Chapter 3, used to analyze captioning papers in our dataset

ods common to literature reviews (e.g., [184, 267]), we conceptualized it as similar to a deductive codebook used in qualitative analyses of interview data. Acknowledging the subjectivity and nuance inherent in our rubric, we do not use this coding process to produce data we can summarize quantitatively (i.e., reporting the distribution of scores for each rubric criteria). Instead, we used this coding pass as a data familiarization and sorting process, to enable our qualitative analysis of how the last decade of captioning literature does or does not engage criteria of collective communication access. After coding, the first author then went through the dataset, integrating each paper into the findings guided by its scores on our rubric (e.g., a paper that strongly met the "considering context" criteria was discussed in that section).

During this analysis, the first author identified more papers that did not meet the inclusion criteria for our dataset. Of the 93 papers included in this coding step, the first author identified 20 more *irrelevant* papers, which the second author reviewed before the papers were excluded from final analysis. Ultimately, 73 papers constitute the final dataset and informed the paper's findings.

7.3.3 Coding for Trends in the Dataset

We also conducted a secondary analysis to systematically identify high-level themes in research scoping and design decisions. The first author coded for an initial set of trends, and then the second author coded a random subset of ten papers using that codebook. At that point, authors had 80% raw agreement and discussed and arrived at consensus over the 20% of codes they applied differently. This drove the refactoring of some codes, which the first author then reapplied to the entire dataset. The second author then coded another random subset of 10 papers. After the second pass, authors achieved an average Krippendorf's Alpha inter-rater reliability reliability score of 0.88 (range .45-1), indicating satisfactory agreement [191]. Our lowest agreement came in identifying participant groups engaged in research.

Final codes for trends fell into two categories: binary and categorical. For some trends, we developed binary codes to capture whether or not papers contained a user study, if they had an explicitly educational focus, and if they studied head-mounted displays (HMDs), videoconferencing, or television captioning. Other data was better-captured by categorical codes, selecting all options that applied for participant type, caption source and communication style. For participant type, we

tracked if papers' participants were DHH, hearing experts, hearing generally, or other. We coded for the following caption sources: CART, C-Print, automatic speech recognition (ASR), crowd sourced captions, pre-generated captions, unspecified source, and other. Finally, we tracked if papers studied the following communication styles: live interactive conversations, live lectures, prerecorded media, unspecified style, and other. We calculated basic summary statistics for this data, reported in subsection 7.4.1.

7.3.4 Positionality

This work is profoundly shaped by authors' positionality and research experience. As a hearing disabled person, the first author has been working in the space of collective communication access for over five years, motivated to explore how technology can support an empowered party in an interaction to see access as something they too are responsible for. In working with d/Deaf and hard of hearing communities as a hearing person, they have been learning ASL, advocating for communication access across personal and professional contexts, and anchoring their work in Deaf critique. They have received graduate education in both HCI and disability studies. The second author, also a hearing person, has conducted and advised research into communication accessibility technologies for approximately a decade.

7.4 Findings

In this section we analyze the body of captioning research defined in section 7.3 through the lens of our theoretical rubric. We begin by summarizing the dataset, providing a high-level overview of the last decade of captioning research. Then in subsection 7.4.2 we combined findings around papers' **collective access approach** and **targets of change** to highlight how captioning researchers conceptualize and aim to increase communication access. Next, in subsection 7.4.3 we identify the extent to which captioning research situates its goals and findings in the context of historical harms and exclusions, combining analysis around how papers engage **historical harms** and **whose knowledge and expertise** is centered. Finally, in subsection 7.4.4 we identify how papers do or do not engage **broader relational, social and environmental context** in captioning research.

7.4.1 Summary Of Dataset

The decade of captioning research we analyzed (2013-2023) spanned varied venues, research foci, and types of captioning studied.

Overall, there was a consistent increase in number of papers published on captioning over time, ranging from three papers published in 2013 to 12 in 2023 (see Figure 7.1), aligned with the overall growth of accessibility publishing over this time period [184, 227, 51]. Papers in our dataset came from 16 different ACM, SIGCHI, and SIGACCESS sponsored conferences. Three venues dominated the dataset, with 34.2% (25/73) of papers published at ASSETS[6], 23.3% (17/73) of papers published at CHI[7], and 15.1% (11/73) of papers published at W4A[8]. The remaining 27.4% (20/74) of papers were published across 13 other venues.

Most research in our dataset engaged participants, with 84.9% (62/37) of papers containing a user study. Of the papers that conducted user studies, 85.5% (53/62) involved DHH participants, 24.1% (15/62) engaged hearing people generally, 4.8% (3/62) recruited hearing experts, and 4.8% (3/62) recruited frequent subtitle users, independent of hearing status. Outliers included studying caption use by people with ADHD [260], recruiting people who speak English as a second language [127], and not reporting on specific demographics of the group engaged in the user study [123, 128, 14, 41, 226, 105]. Note that only one paper explicitly identified captioning as an access tool for people who are not DHH [260]. Throughout our findings we discuss how captioning research regards DHH people, because that is the focus of the last decade of research, but it is important to note that DHH people are not the only people who use captioning as an access tool.

Research studied a range of captioning styles and contexts. 35.6% (26/73) of papers studied caption users' experiences with captioned videos, often not specifying how those videos came to be captioned. 54.8% (40/73) of papers studied ways of providing real-time captioning: 41.6% (30/73) used automatic speech recognition (ASR), 13.7% (10/73) used CART[9], and 6.8% (5/73) used crowdsourced captions. Some papers did not specify caption source or employed multiple caption sources.

---

[6] https://dl.acm.org/conference/assets

[7] https://dl.acm.org/conference/chi

[8] https://dl.acm.org/conference/w4a

[9] CART, or Communication Access Real-time Transcription refers captions provided by expert human transcriptionists, using specialty equipment and software to transcribe in near real-time [2]

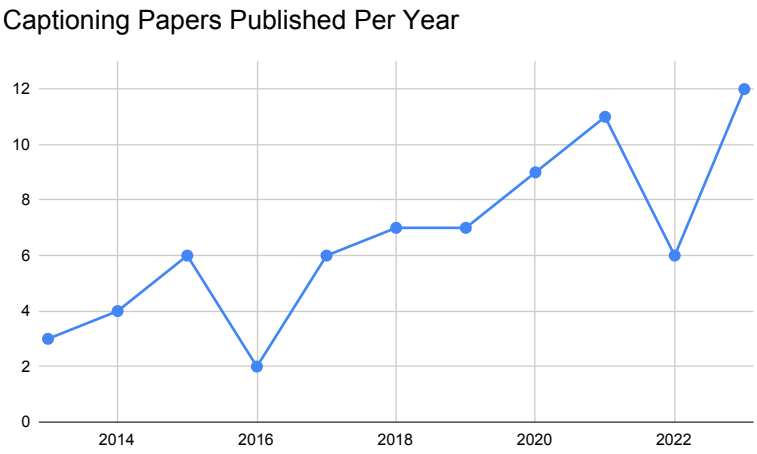


Figure 7.1: The number of captioning papers published per year steadily increased, going from 3 papers in 2013 to 12 in 2023

Researchers also studied a range of communication contexts, researching caption use during live, interactive conversation (42.5%, 31/73), live lectures (10.9%, 8/73), while viewing prerecorded media (42.5%, 31/73), and while watching live television (5.5%, 4/73). There was, sometimes, a difference between the intended type of communication studied and research instruments – 17.8% (13/73) of papers aimed to study live forms of communication but showed participants prerecorded videos. There were also specific communication environments that received notable attention by researchers who tailored their captioning interventions to these contexts: educational captioning (19.1%, 14/73), head mounted displays for captioning (15.1%, 11/73), television captioning (13.7%, 10/73), and use of captions while videoconferencing (8.2%, 6/73).

7.4.2 How Captioning Research Aims to Make Communication Accessible

We identify two key approaches for improving communication access used throughout the last decade of captioning research. Most commonly, researchers aim to improve communication access by iterating on and improving the design of captions and caption form factors, without disrupting the social status quo. Yet, there is a smaller body of research that aims to alter the behavior of nondisabled communication partners, finding that a captioning tool's efficacy depends on the social context it is used within. We draw on our framework's sections on a paper's **approach to collective**

access (Criteria A) and research's **target of change** (Criteria B) to analyze how our dataset envisions communication can be made accessible.

Improving Technology To Improve Access

The majority of captioning research we analyzed focused solely on improving captioning tools themselves to attain the larger goal of extending access. It is, of course, necessary and important to iterate on captioning technologies, but we identify limitations to approaches that only try to change technologies, not inaccessible communication norms.

One of the key targets for improvement is captioning interfaces—researchers aim to make them more readable or include more information. A significant body of research is dedicated to exploring what makes captions more readable and intelligible (e.g., [292, 58, 106, 28, 145, 301, 4, 105, 225, 143, 160, 170, 128, 98, 226, 41, 42]). These interventions aim to lower the cognitive load needed to use captions through approaches such as highlighting [145], formatting [106], and caption placement [41]. For example, Kushalnagar et al. developed a tool to allow Deaf, low vision students to better access lecture recordings while splitting their attention between magnified visuals and captions [162]. Other techniques include visualizing non-speech information dynamically alongside transcription (e.g., [193, 4, 150, 67, 116]), formatting captions to better identify speakers (e.g., [12, 41]), and making captioning tools customizable (e.g., [128, 198]).

Research on the most popular novel interface for captioning over the last decade, head-mounted displays (HMDs), often aimed to sidestep social factors that impact communication accessibility by introducing a new form factor. Research on HMDs aims to make captioning-mediated social interactions more naturalistic, citing two primary motivations: the difficulty of gaze management on the go (e.g., [192, 134, 223]) and a desire for more socially acceptable, discreet solutions (e.g., [136, 108, 134, 40, 94, 223]). HMD form factors tend to place all of the burden of change on DHH people, the only ones who don a new technology to communicate, one that is frequently reported to be burdensome (e.g., [192, 134, 94]). HMD captioning ultimately makes it easier for DHH people to meet hearing norms for spoken conversation rather than adapting communication settings to be more ideal for caption use. As new technologies become viable and promise a better form factor for viewing real-time captions, researchers should also consider who must do more of the work of

access in this envisioned future.

The tendency of captioning research to focus on technology over social context can also be seen in the tendency to study and develop new tools only with and for caption users. Many tools designed for interactive communication are neither studied nor evaluated in interactive contexts. For example, Berke et al. [29, 28] aimed to study how different algorithmic confidence visualizations could impact DHH people's experiences of one-on-one automatically captioned conversations. Rather than observing an interactive conversation between a DHH participant and a hearing person, they asked participants to view prerecorded videos and imagine they are part of a live conversation. This is a relatively common approach to studying new captioning technologies but abstracts away conversation partners' agency and capacity to mutually determine an accessible communication style.

Changing Social Norms to Improve Access

Though less common, we did observe another approach to increasing communication accessibility: engaging caption providers, third-party supports, and direct interlocutors. These approaches more closely align with our collective communication access framework.

When studying caption-generation ecosystems, research often translates insights from DHH caption users into guidance for professional or amateur caption creators. For example, Alonzo et al. [4] proposed novel sound visualizations and, after determining DHH captioning users' needs and preferences, tested their system with its intended users: hearing video creators. Other work provides guidance to professional caption providers, such as how to place captions during live television broadcasts (e.g., [9, 10, 11, 12]) and how networks could design their caption systems for online sharing of broadcast content (e.g., [128, 59, 105, 58, 41]). Several papers studying user-generated video content provided guidance for both platforms and creators [181, 175, 260, 31]. For instance, Mack et al. [181] found that social platforms could *"gracefully teach"* Deaf norms to hearing users, relieving some of the burden of explaining communication etiquette. Other work begins with the assumption that online videos will not be well-captioned, and seeks to identify what should be prioritized to be captioned by outside providers [31, 258]. While there is truth to the notion that not all videos will be well-captioned, whether or not researchers see content creators as key participants

in making communication accessible shapes the future possibilities they imagine.

Some research considered third-party stakeholders such as captioners, instructors, and university disability services in the design of captioning technologies. Particularly when researching captions in the classroom, lecturers and university access service providers are often implicitly situated as stakeholders and users of captioning technologies, though their perspectives are rarely represented in findings (e.g., [45, 145, 156]. However, when Kawas et al. [149] studied DHH students' experiences of captioning, they also engaged disability services professionals, captioners, and professors, finding that instructors need more support to best teach students using captions. There is an interesting tension at play – ultimately these captioning tools only succeed if they provide students access. Yet, by not engaging other parties, researchers miss the opportunity to explore how more accessible communication styles (e.g., lecturers' teaching styles) could coordinate with and be supported by technical systems to improve access.

While the last decade of captioning research infrequently engaged with direct conversation partners, the research that did so finds they have significant impact on conversation access. Seita et al. focused on understanding hearing people's behavior changes while captioned [254] and DHH people's priorities for ideally accessible behaviors [256, 255]. They motivate this work in a vision of collective access: *"technology could more equitably distribute [conversational feedback] responsibility among all parties in conversation"* [255]. Other work has identified how DHH people's experiences of captioned conversations generally, [196, 178, 299], speechreading while using captions [69], and using HMDs [135] can be improved or made more difficult by their hearing conversation partners. While codesigning for mixed-hearing-ability groups, both McDonnell et al. [198] and Seita et al. [257] found that, even when communication access services are available, communication accessibility depends on all parties, DHH and hearing alike, adapting their norms and collaboratively developing a bespoke, situated communication style.

7.4.3 Reckoning with Historical Harms and Centering Community Perspectives

Disability studies, Deaf studies and disability justice all stress that the current state of discrimination that Deaf and disabled people face is a direct result of political, social, and cultural choices. Therefore, working to make the world more accessible ought to mean taking a political stance to counter

ableist, audist, and eugenic forces. However, captioning research infrequently names or aims to counter the historical harms and biases that make communication access an ongoing concern. Yet, the current body of captioning research largely avoids significant harm. While not frequently linking to broader discourses on accessibility or disability activism, research does increase communication access. Furthermore, for decades, engaging disabled people in decision making about disability issues has been one of the key demands of disability activism [131, 230]. Human-centered accessibility focused captioning research overwhelmingly centers DHH people as experts on captioning technology, but does not link this practice to its activist history. The small body of politically-engaged captioning research, however, more powerfully names and addresses specific harms and does epistemic work to empower DHH technology users and researchers. In this section we synthesize our coding around **historical harms** (Criteria C) and **whose knowledge and expertise** (Criteria D) is considered in captioning research.

Ungrounded Captioning Research

Often, captioning research does not explicitly reckon with the historical discrimination and disenfranchisement that makes communication access a pressing issue for DHH people. Traditionally, activists and theorists emphasize the importance of developing a critical consciousness that can precisely address the root causes of historical harms [88]. While often not articulating a critical consciousness, the current body of captioning research does not reproduce the extreme ableist and audist harms discussed in Chapter 3, suggesting that there is an unnamed politics undergirding the current practice of human-centered accessibility research. Here, we highlight that how researchers frame their broader mandate and engage with DHH communities is impacted by a lack of consideration of historical harm and bias.

Often, captioning research does not discuss the harms of inaccessibility and role of technology in increasing access. Papers instead introduce captioning as a tool DHH people use and immediately dive into the problems with current caption provision that their work addresses (e.g., [115, 292, 144, 150]). The frequent lack of grounding in an accessibility or justice-based framework can leave research on captioning devoid of a connection to access as a political project.

While many papers implicitly identify DHH people as experts by testing captioning technologies

with exclusively DHH users, most do not explicitly describe this as a way to place the direction of technical development in the hands of the DHH community. The majority of papers contained user studies, though some short form publications did not include them, often presenting preliminary technical work or summarizing an ongoing research project (e.g., [115, 162, 253, 225, 305, 42]). Yet, overwhelmingly, while most papers did engage caption users to shape the design of captioning, they did not frame this as an intentional or value-laden choice. Bringing disabled people into decision-making that impacts disabled people's lives, however, is the most famous demand of the disability rights movement [230]. It is notable that the current state of human-centered accessibility focused captioning research has naturalized this activist demand into a practice that does not require justification. In fact, there were two short-form publications [284, 283] within our dataset that had hearing people simulate being DHH. Yet, Tu et al. [284, 283] engaged with critiques of disability simulation (see [213, 222, 279]), emphasizing that not engaging DHH communities limits the applicability of their research.

Connecting Captioning Technologies to Historical Harms

There is also research that connects technology design to a broader historical and political context, sometimes driven by specific research areas and sometimes as part of a broader research ethos.

There are some areas of captioning research that are more likely to engage with historical legacies. When research is motivated to support DHH people with lower literacy, it sometimes situates lower literacy rates in the context of oralism and educational discrimination (e.g., [30, 181]), but other times simply states low literacy as a fact without context (e.g., [106]). Discussion of social acceptability of technology is also historically situated. Some research notes that the desire for technology to be socially acceptable is rooted in discrimination and builds to protect people from that discrimination, rather than challenging it. For instance, Olwal et al. [223] justify their focus on HMDs as follows: *"many people do not wish to call attention to their disability for fear of exploitation or discrimination ... for example, eyeglasses may be more desirable than hearing aids due to the perception that hearing aids are for older adults."* Others actively engage this tension, such as Findlater et al.'s [95] discussion of how to balance well-documented desires for socially acceptable assistive technologies with the Deaf community's commitment to distinctly Deaf visual

communication styles. Educational captioning tools were more often studied in a context of historical discrimination [145, 305, 156], exemplified by Kushalnagar's [156] motivation: *"nearly fifty years after the first educational mandate for accessible services in the early 1970s, only 16% of DHH individuals complete a bachelor's degree or higher, far less than the graduation rate of 30% for hearing individuals."*

There is a final, smaller body of work that engages with Deaf and disabled history, activism, and culture as an intrinsic motivation. Seita et al [254, 256] motivate their work by noting that *"communication barriers may lead to isolation, miscommunication, or reduced productivity or professional outcomes"*. Citations to Deaf community knowledge include citing Christine Sun Kim[10] as an authority for creatively conveying non-speech information in captioning [193] and citing the "#NoMoreCraptions" campaign [76] to motivate research into automatic captioning approaches (e.g., [193, 196, 175]). Desai et al. [69] grapple with the history of oralism in their research on speechreading and Simpson et al. [260] focus on captioning as an access tool for people with ADHD, noting a historical lack of research that engages adults with ADHD outside of a corrective focus [267]. McDonnell et al.'s work [198, 196] is framed by disability justice and Deaf studies, explicitly exploring collective access captioning approaches. This body of work provides its audience the opportunity to understand captioning technologies as connected to a broader politic and decenters technologists as the sole authority on captioning technologies.

Nothing About Captioning Without Captioning Users

From the disability rights rallying cry of *"nothing about us without us"*, to Deaf studies' emphasis on self determination, to the disability justice principle of leadership of the most impacted, disabled and Deaf scholars and activists are clear: work that impacts Deaf and disabled people should not just include but cede leadership to impacted communities. A small body of work aims to do so, either by engaging participatory methods or by having Deaf-led research teams.

Though far from the dominant methods in the field, there is a body of research that puts the direction of captioning technology development in the hands of the DHH community using participatory methods, such as codesign [149, 228, 257, 198]. For example, Kawas et al. [149] first

[10]Kim is a famous Deaf visual artist: https://christinesunkim.com/

used a range of methods to understand DHH students' current experiences of using captioning and then conducted codesign workshops to set directions for future improvement. Peng et al. [228], on the other hand, began with codesign workshops to guide the implementation and testing of their augmented reality (AR) captioning tool. More so than traditional user research, codesign methods move toward collective communication access by transferring power toward impacted communities.

Another form of community-engaged research is work done by DHH scholars who articulate the ways their lived experience shapes their technical expertise. When research is conducted by DHH researchers, even work that is not explicitly autoethnographic is often described as informed by the DHH author's lived experiences (e.g., [136, 108, 301, 299, 69, 116]). For example, Jain et al. [136] describe determining specifications for a HMD from the hard of hearing first author's experiences using and fine-tuning the system, using his preferences as defaults that could be customized for later participants. Jain et al [134] have also published explicitly autoethnographic work describing the experience of using HMD captioning. ASSETS experience reports[11] provide a unique format for publishing the experiences of disabled technology users. In one, Loizides et al. [178] highlight the variety of use cases that both hearing and DHH authors had for on-demand access to automatic captioning via Google's LiveTranscribe app. Additionally, Mathew et al. [192] describe the experiences of two DHH authors using AR-captioning and AR-interpreting. Having member-researchers conduct captioning research provides opportunities to enrich findings and better align technology design with the lived experience of long-term captioning use.

7.4.4 Considering Communication Context

Finally, we explore whether research considers the role of relational, social, or environmental context in shaping access practices. We highlight that in captioning research, there are two major dimensions in which to consider context: in the design of the study method and in the design of implemented or proposed interventions. We observed both research that abstracted away context in their experimental design and research that deeply considered context when designing for and understanding caption use. This section draws solely from the **context** criteria of our framework (Criteria E).

---

[11] https://assets23.sigaccess.org/experience-reports.html

Researching Without Communication Context

While research on captioning is often framed around matching the specific nature of different kinds of communication, some research abstracted away key contextual features in their study design.

While controlled experiments provide a very valuable form of knowledge around the efficacy of interventions, their necessary control of contextual factors limits the method's viability for collective communication access research. We observed researchers frequently using prerecorded stimuli (e.g., [29, 30, 67, 161, 100, 40]) while aiming to study interactive communication, losing out on relational context or ability of conversation partners to mutually build accessible communication practices. Controlled experiments run during interactive conversations were often highly manufactured. For instance, Seita et al. [256, 255] had an actor repeat the same response multiple times in a row, with varied levels of a communication behavior, to measure hearing people's impact on DHH caption users. While this allows for specific forms of measurement, it does not reflect how people would experience those behaviors in context, and could miss critical nuance (e.g., familiarity with a conversation partner). We did observe some controlled experiments in our dataset that more-closely replicate the conditions in which those captions would be used, such as Al Amin et al.'s research into optimal placement of TV captioning ([9, 10, 11]). The types of knowledge controlled experiments produce can be invaluable in advancing certain arguments – for instance, Berke et al. [30] demonstrated a statistically significant difference in DHH people's ability to identify caption errors depending on literacy, providing impetus for future research to better design for DHH captioning users with less access to literacy. Yet, often the nature of controlled experiments can limit the strength of claims they can make about real-world communication contexts.

We also observed instances where, to try to isolate any varied understanding to an intervention in a controlled experiment, researchers either played videos with no audio (e.g., [67]) or would have their participants listen to white noise to fully equalize their hearing ability (e.g., [294]). While this may yield a more rigorous controlled experiment, it does not reflect how many DHH people communicate, combining their hearing abilities with assistive tools. Further, this impulse to remove participants' residual sensory ability has been criticized as a form of disability simulation [279, 222, 213] in the context of work with people with low vision [83, 237], while this choice often goes unremarked upon in research with people who are hard of hearing.

The Benefits of Considering Communication Context

When researchers investigate captioning technologies in ways that deeply consider the social, relational, and environmental contexts they are used within, it is clear that this context is integral to shaping how effective captioning tools can be.

Some research does not explicitly set out to understand the role context plays on caption use, but demonstrates the importance of context by designing for hyper-specific use cases (e.g., [82, 105, 305]). Some research demonstrated the need to pay close attention to the type of media being captioned. For instance, Al Amin and Hassan et al.'s [10, 11, 9, 12] work on automatically placing captions during live television broadcasts found that, for instance, there are different on screen information demands for a sports game than a daily news broadcast. Other work was driven to match the cultural and linguistic context captioning tools are used within [49, 277]. For example, Takagi et al. [277] explored a combination of ASR and crowd captioning to caption Japanese, as complexities of written Japanese make fully human-generated captions very difficult to produce. Designing captioning tools for narrowly scoped use cases allowed researchers to engage with and leverage the context tools will be used within.

Some work has found that the places captioning tools are used within greatly shape the experience of using those tools. Research on HMDs has revealed that simultaneously reading captions and maintaining situational awareness is highly cognitively demanding [136] and that granular environmental considerations must be taken into account in AR interface design (e.g., how captions should appear if the speaker is not in the user's field of view [228]). The specific environment of the classroom was often central to educational captioning research. For instance, Kushalnagar highlights that classrooms must be well-configured to allow captions to facilitate access [156]. Kawas et al. [149] surfaced the extent to which students' experiences of captioning are impacted by seemingly minor constraints, such as the availability of outlets in a classroom.

Broader sociocultural factors also impact captioning tools. Findlater et al. [95] emphasize that DHH people's sound awareness preferences vary across social contexts, finding statistically significant differences in, for instance, people's concerns around social acceptability of technologies used with close others versus with strangers. People's communication backgrounds play a significant role in what kind of technology is the appropriate solution—Mathew et al. [192] emphasize that

real-time captioning is a much more usable tool for a DHH person who prefers to voice for themselves. Captions only provide access for receptive, not expressive communication, and therefore do not provide access to all DHH people equally [192, 198, 196]. Loizides et al. [178] highlight how communication access is also shaped by broader global context, finding that masking and social distancing in the wake of the COVID-19 pandemic made automatic captioning a more important part of many DHH people's access practices.

Taking a more specific focus on context as a key determinant of communication accessibility, McDonnell et al. [196] propose that caption use must be understood in the context of interrelated social, environmental, and technical factors. This framework has been taken up in their follow-up work [198] and by other researchers. Seita et al. [257] discuss their findings in the context of the social and environmental factors that shape both the communication that happened during the study and the designs participants developed and Desai et al. [69] emphasize social, environmental and technical impacts on people's decisions to either speechread or use captions.

7.5 Discussion

In this chapter we assess the last decade of human-centered accessibility research on captioning technologies through a framework of collective communication access (see Chapter 3 and Table 7.1). We find that HCI captioning research does not perpetuate many of the harms technologists have historically caused, which prior literature reviews indicate is not true across all accessibility subdomains [267, 265]. Most captioning research engages DHH communities as the relevant authority on caption design and targets change toward technical infrastructure. However, the small body of work that is strongly aligned with our collective communication access framework shows that technology alone cannot create fully accessible communication. Work that treats access as a political project that hearing interlocutors should be a part of demonstrates the rich potential for future work in this burgeoning space. We now discuss two key considerations for future collective access captioning technologies: conducting future collective communication access research and balancing social acceptability and social change.

7.5.1 Conducting Future Collective Communication Access Research

We envision exploring additional domains of communication access through the lens of collective communication access and other applications of our framework and rubric.

Captioning research provides an interesting domain within which to study collective communication access since it is necessarily technical and used to facilitate oral communication, potentially making it less likely that research will be engaged in Deaf culture and community than work that studies the use of signed languages. However, our framework is not specific to captioning technologies. A comparative analysis of the last decade of ASL research within ACM, SIGCHI, and SIGACCESS sponsored conferences could be illuminating (e.g., exploring whether ASL research is more grounded in historical context and harms) and is a promising direction for future work. Augmented alternative communication (AAC) devices frequently facilitate access for people with intellectual, developmental, or speech-related disabilities. Work in communication studies [5] and HCI [285] has identified that AAC technologies can be designed and deployed in ways that limit their users' agency and are primarily designed to support AAC users to engage in normative, spoken conversation. Bringing a collective communication access lens to AAC research could reveal ways to redesign AAC-mediated communication as a negotiation between communication partners. Furthermore, digital accessibility is, at its core, a question of communication access – can users communicate their inputs to a technical system and can they access the outputs of that technology. Considering research areas such as making websites screenreader-navigable, providing high quality alt-text on social media, or ensuring documents are formatted accessibly through a lens of collective communication access could reveal new angles and approaches to address these long-intractable access barriers.

Our analysis of captioning research through this framework identifies practical considerations for future work. Currently captioning is overwhelmingly studied with DHH people, but there are other groups who use captioning for accessibility [260] and narrowly associating an access technology with a user group risks marginalizing many people who could benefit from that technology [186]. Further, research often does not engage with the fact that captioning technologies serve only a particular subset of DHH people well during interactive conversation – those who are comfortable voicing their contributions. Future work should explore how captions could better support

DHH people who prefer not to voice. We also recommend that future research more often study all people communicating using captions, rather than only studying DHH captioning users. Finally, we identify that experimental methods are often conducted in a way that does not match real-world caption use. We recommend that future work complement lab-based findings with situated, qualitative analyses of novel tools in the contexts they will actually be used within.

While we used our theoretical rubric[12] to assess published work, we envision it as a tool that could be used during research definition and study design to support researchers in reflecting on their approach and guiding them towards collective communication access framings. This could be particularly useful to researchers who are new to fields such as disability studies, Deaf studies, disability justice, or communication studies. We emphasize that interdisciplinarity is crucial to our analysis – the lead author of this work is both a disability studies scholar and HCI researcher. We hope that work such as this could provide a foothold into the long journey of learning and substantively engaging with other fields that hold crucial knowledge on disability, Deafness, and accessibility.

7.5.2 Balancing Social Acceptability and Social Accessibility

We identify collective communication access as an approach that both aligns research with Deaf and disability scholarship and produces cutting-edge technology designs. However, we must not ignore a main justification of research that designs technologies that preserve communication's status quo—socially acceptable technologies within this status quo can be necessary for and desired by DHH people navigating an inaccessible world.

This tension between the promise of collective access and the complexities of social acceptability is built into disability justice activism's conception of this principle. Sins Invalid stresses that collective access approaches should be arranged *"depending upon an individual's needs, desires, and the capacity of the group"* [131]. We do not create a more just future if access is **only** provisioned in spaces where caption users feel safe enough and have the time and capacity to guide their communication partners toward more accessible communication styles. Caption users may be more likely to communicate orally and or not be at a place in their lives to want to change how

[12]See Supplementary Materials

they communicate with those around them. Yet, practices such as DeafSpace [81] show us that, by default, current design does not support DHH modes of communication that diverge from a hearing norm. The strong skew towards designing technologies that do not challenge inaccessible social norms fails to support many people's current or desired access practices. Furthermore, prioritizing designs that align with the social status quo limits our capacity as a field to change the ableism and audism inherent to current communication environments.

Captioning research and design could make collective access approaches to accessibility more socially acceptable. The technologies we use to communicate shape what we think normal communication is. Recent years have seen a substantial change in how hearing people interact with captioning, with the majority of young people preferring to watch captioned media [46], the rise of captioning practices on user-generated video content platforms [260], and a normalization of real-time automatic captioning on videoconferencing platforms [275, 273, 274]. Captioning researchers should not stop supporting people who would prefer technologies to be discreet and socially acceptable. However, there is an opportunity to commit to designing communication tools and platforms that make not only the availability of captioning socially acceptable, but make collective access approaches unremarkable and expected.

7.5.3 Limitations

Our work does have limitations, namely a scoped context and subjective analysis. While our analysis of a decade of research is aligned with methodological norms, we only queried the ACM Digital Library, and there are many other repositories of research that may hold human-centered, accessibility-focused captioning research. Furthermore, we take an American perspective on disability studies and Deaf studies and analyze HCI research published in English. Global disability studies scholars emphasize that local context matters, and our framework should not be assumed to be global. Our North American focus also translates into our language choices, as the difference we highlight between captions and subtitles is grounded in our cultural context. Finally, our research presents a taste of several vast disciplines, filtered through our perspectives as hearing technologists. Other researchers may have selected different bodies of work or takeaways and interpreted them differently.

7.6 Conclusion

In this research we articulate a framework for collective communication access, drawing from disability studies, Deaf studies, disability justice, and communication studies. We then translate that framework into a rubric to assess how prior captioning research has considered communication as a collective process, who or what is targeted for change, how research grapples with historical harms, whose knowledge and expertise is centered, and how communication context is integrated into research and technology design. From our analysis of captioning research from 2013 to 2023, we find that the current state of captioning research is not dire – research largely aims to improve technology to improve access and treats DHH people as experts on captioning. Yet, work that attends to the social context of communication and envisions access as a collective responsibility—that is, work that better aligns with a collective communication access approach—identifies access barriers and opportunities for technological development not seen in other research. We end by envisioning a future where technology makes collective access more socially acceptable and where collective communication access is extended to other technical contexts.

7.6.1 Within-Dissertation Contextualization

In this chapter, I track the collective communication access framework introduced in Chapter 3 through the last decade of human-centered accessibility-focused captioning literature. A key takeaway from this chapter is that designing for collective communication access is a novel innovation in the field – it is primarily research presented in my previous chapters (see Chapter 4 and Chapter 5) and work led by Dr. Matthew Seita that has engaged hearing people as key determinants of communication accessibility, and my research uniquely engages communication context (rather than controlling for communication context to enable experimental findings). Work that engages with aspects of collective communication access identifies considerations that do not emerge in research less-grounded in context, community politics, or collective access.

Chapter 8

DISCUSSION

Over the course of my dissertation, I have demonstrated the thesis:

"Centering social, relational, and environmental factors as key determinants of communication accessibility while designing captioning technologies better matches captioning's realistic context of use and can advance disability and Deaf community values."

I began by outlining a framework for collective communication access in Chapter 3, and then empirically demonstrated the role that intertwined social, environmental and technical factors play in shaping the use and usefulness of real-time captioning tools in Chapter 4. In Chapter 5 I detailed the codesign process I conducted to design collective communication access tools to support mixed hearing ability groups, demonstrating how this framework can produce novel technical approaches and identifying the role that relational context plays on captioning experiences. In Chapter 6 I explored the role of context and collective access in a different type of captioning—user-generated captions on TikTok—identifying the social and environmental factors that shape captioning on platform and considering how collective access approaches could apply in a context where hearing creators are already participating in captioning content. I ended with Chapter 7, tracking my collective communication access framework through the last decade of captioning research and identifying how considering context and designing for collective access does better match captioning's realistic context of use.

I now end my dissertation with this chapter by reflecting on and discussing the major provocations of my research and conclude with limitations and future directions for my dissertation research.

8.1 Considering An Activist Accessibility Research Practice

A tension I have engaged throughout my work is how to balance competing calls within HCI accessibility to do disability studies or activist-informed research and user-centered research. Within diverse communities like the DHH community, not all participants' perspectives align with activist

communities. Further, as my work targets change toward a party not often engaged in accessibility work, in some ways my research intentionally counters hearing people's preferences and default desires. Here I reflect on the competing asks of these two philosophies of responsible accessibility research and discuss a path forward.

As we push toward an activist accessibility practice, it is important to remember those who may not be well-served by this approach to technology design. Since Mankoff et al.'s 2010 *Disability Studies as a Source of Critical Inquiry for the Field of Assistive Technology* [189], some within HCI accessibility have aimed to align our work with disability studies and activist perspectives. This is a core commitment of my work and one that I argue for strongly – turning to scholars of disability and disability activists provides a theoretical source that can guide more responsible accessible technology research. However, disabled activists are not representative of all disabled people. The technology someone who proudly identifies as a disabled activist might prefer could be a terrible option for someone with acquired hearing loss who does not want to change how they communicate with their family and friends. Disability studies scholars and activists identify that person with acquired hearing loss as possibly struggling with internalized ableism [141]. While those scholars and activists may illuminate a path toward disability pride for that person in years to come, the question remains of what technologists ought to do for them now. Deaf cultural perspectives that place moral value on communicating only in ASL and eschewing connection to the hearing world have been critiqued [242] for being essentialist and dismissive of those without privileged access to a signing community. Yet, it is through communities that affirm disability as an *"ingenious way to live"* [90] and develop disabled and Deaf pride that we can *"begin to anticipate presents and to imagine futures that include all of us"* [142]. Balancing the imaginative power of disability studies and activism with the myriad realities that might make challenging an ableist status quo undesirable to many disabled people is a real challenge for activist accessibility research.

Yet, conducting user-centered research without an activist perspective risks designing technologies that reify the status quo and do not enable a less-ableist future. The last few decades of HCI research have emphasized that research ought to be user-centered. Coming from more traditional design legacies [218] and newer practices of design thinking [48] is the idea that design recommendations and assessment must come from the current or imagined users of that design. Within HCI accessibility, critique of disability simulation [279, 213, 222] has largely succeeded in ensuring

that accessibility technologies are designed with the input of disabled people (see Chapter 7 for an discussion of this with regard to captioning technologies). Yet, perhaps particularly in spaces where technologies can advance or further specific forms of marginalization, treating all participants as equally qualified to advise on the design of accessibility technologies is epistemically dangerous. We see from the history of disability activist movements that not being intentional about leadership of the *most impacted* results in leadership of and legislation and design for the *most privileged* [207]. This is not to say that accessibility researchers ought to only consult activists or any other specific subset of disability communities, but that we must contextualize our participants' politics and experiences. The research I present in this dissertation was done with people who held ableist beliefs and with people who are politically committed to Deaf-centric communication styles. The range of experiences participants shared with me all had understandable causes – DHH people routinely encounter ableism and audism and some internalize those beliefs while others develop a proud Deaf identity. To the best of my ability and evolving perspective, I hope that I did justice to the range of perspectives participants shared with me while still maintaining an anchor in disability justice politics.

I argue that neither activist-grounded nor user-centered research practices are a panacea and that we should seek a middle path forward. Specifically, I call for a design practice that serves communities with diverse preferences right now while also working to achieve a future where mainstream communication technologies embed activist values around interdependence and collective access. In the United States, most DHH people are older adults who spent a lifetime communicating orally – only 16% of people 20-69 use hearing aids, while 30% of those over 70 do [118]. I argue that accessibility researchers do not fulfill our mandate to DHH communities without providing technologies that support someone who would prefer to communicate orally and with minimal change to their prior communication strategies. However, I also firmly believe that without an activist commitment, accessibility research cannot advance us along a path to a future that is not only more accessible but also less ableist. We do not have to assume that, just because a collective access approach to communication that decenters hearing norms is intimidating to many DHH and hearing people today, that this will always be true. As I demonstrated in Chapter 1, the last five years of captioning technologies alone have seen significant changes in what people consider to be normal uses of captioning. We can take inspiration from the largely-bottom up changes of the last five years

to engineer collective access technologies for the next five years and beyond.

8.2 Tensions in Designing for Collective Communication Access

Throughout my dissertation, I demonstrate the promise and potential of designing captioning technologies for groups, rather than individuals. Specifically, I propose a theory of collective communication access that is tightly linked to the disability justice concept of collective access. To reiterate, Sins Invalid describes collective access: *"Access needs can be articulated and met privately, through a collective, or in community, depending upon an individual's needs, desires, and the capacity of the group"* [131]. This redirection towards treating access needs as a group responsibility animates my research. While I strongly believe in this potential, it is not without complications. Here I identify tensions around designing for unequal use across groups, designing for group use of biased tools, and designing collective access with multiply disabled groups.

Designing for Different Uses of Group Tools

While my dissertation work envisions access as something that people do together and proposes captioning technologies for group use, we must grapple with the different ways different users would interact with such systems and the power dynamics involved. As I demonstrate in chapters 4, 5, and 6, the choices communication partners make and the attitudes they hold make a material difference in the extent to which captioning users can access communication. Nondidsabled people do not face the same consequence for engaging in inaccessible communication behaviors. Therefore, designing captioning technologies for group use must **not** mean designing for equivalent use. Literature on the design of groupware explores how to build systems that have different user roles [107, 68, 247] (e.g., videoconference meeting host, participant), a crucial body of work to engage in future research.

Adding to the different roles people may have while using captioning tools (e.g., DHH users notifying when things become accessible, hearing users learning accessible behavioral norms), we must also consider power dynamics. Unlike the dynamics of meeting host and meeting participant, where the person with greater control over the technology has implicitly been identified as the person with more control over that interaction, collective communication access tools may often place control in the hands of less-empowered members of a group. As discussed in Chapter 5,

these tools must be designed carefully to enable captioning users to ask for access anonymously in some situations and aim to keep hearing people from becoming embarrassed or hostile. Because captioning tools must afford more control to caption users than their interlocutors, these tools must also be carefully designed to allow a less empowered member of a group (e.g., a student in class, a junior colleague) to get the access they need without facing social consequences.

Yet, there are limits to this differential consideration. Often caption users face workplace [87, 86], educational [156, 305], and social [254, 255] discrimination in mixed ability settings, but this is not a durable assumption. Consider a scenario where a white Deaf supervisor does a poor job of considering the input of junior, Black, hearing women on his team. If group captioning tools are designed to prioritize captioning users' access needs, in ways that make it harder for the junior, hearing members of this team to interrupt, they may inadvertently further workplace discrimination in the name of accessibility. To design for groups requires not just an analysis of ableism, but an intersectional analysis [60, 55].

Furthermore, a central assumption of collective access is that groups will be willing to work together to communicate more accessibly, but this is not without its challenges. Asking all conversation partners to consider access means that people must work at changing their default communication styles, which is no small feat. While adopting accessible norms does have tangible benefits for all communication partners—as I contend in Chapter 3, communication is not effective if it is not understood by its audience—nondisabled people face far fewer obvious consequences for failing to communicate accessibly. At its core, my dissertation research is anchored in activist thinking and is focused on how technologies can support people in enacting a world that is more in line with values of disability justice. Yet, it would be naive and counterproductive to ignore the fact that not only are people resistant to change, but that, even when strongly motivated, changing behaviors is fundamentally difficult. Designs for collective communication access must acknowledge and aim to address this difficulty and resistance. As highlighted in Chapter 4 and Chapter 5, these tools must also respect the fact that DHH people do not want or always feel safe to renegotiate communication norms with all conversation partners. Yet, the pool of loved ones, friends, classmates, and colleagues of caption users is non-trivial and designing for use with close others may be a promising first use case for collective communication access tools.

Collective Use of Biased Access Technologies

Next, collective communication access approaches risk reifying bias and discriminatory group dynamics given the bias inherent to both human and machine-centered captions. Here I want to highlight some of the work that I intentionally did not do over the course of my dissertation research. I had initially imagined systems that encouraged conversation partners to monitor captions for errors and to take responsibility for their being captioned accurately. While this is a practice some participants described in their conversations (see Chapter 5), I pulled back from designing for caption monitoring. First, when floated as a design probe (see Chapter 4) or raised by participants (see Chapter 5), reflection revealed that correcting errors in real time would be infeasibly cognitively demanding and did not match the way caption readers made sense of errors (see also [145]). But, more importantly, literature on bias in ASR (e.g., [153, 101]) and online discussion of the failure of white interpreters and captioners to adequately provide access to non-white speakers and signers [233] demonstrates that who is captioned easily is not a neutral matter.

When both the humans and algorithms who may generate captions demonstrate bias, a system that makes speakers responsible for the errors made while captioning their voice encodes racism, audism, and other forms of discrimination. Conversations with DHH colleagues about caption error correction also highlighted that, given ASR's poor performance on Deaf accents [101], encouraging hearing conversation partners to correct captions could further situate DHH people's speech as unintelligible, an assumption already made of Deaf accents [101, 33]. It is a significant harm that algorithmic and human captions do not equally provide access for all, and an area that deserves significant future work centering the perspectives of Deaf people of color. While the uneven performance of captions often makes collective access solutions even more necessary, we must take care to not build tools that amplify the impact of algorithmic discrimination.

Cross-Disability Collective Access

A further consideration for collective communication access tools is how to design for scenarios where multiple disabled people interact. Historically, HCI has studied disability as a set of siloed identities that rarely interact, not looking at cross-disability spaces or the experiences of multiply disabled people [184]. While I studied groups with multiple DHH people (see Chapter 5) and

intentionally recruited on basis of caption use, rather than identity, for the work recounted in Chapter 6, my dissertation work largely does not challenge this trend. Yet, disability community activists [231, 205, 131] and a small body of HCI research [183, 185, 122] demonstrate that when disabled people collaborate, access needs emerge that are often not anticipated by even existing accessible technologies.

Consider the video prototypes I put forward in Chapter 5. While participants stressed that it was important to them that colors be color blindness friendly and one participant grappled with the access needs that might make being on camera full-time difficult, many of these proposed ideas rely on visual cues and do not assume that hearing conversation partners may have their own access needs. Groups with multiple DHH members did discuss strategies (e.g., switching to text) when captions did not support mutual understanding, a small taste of how disabled people work to meet each others' access needs. While accurate to the groups codesigning ideas for their own use in this particular study, the small range of access needs considered is a limitation of this work and a provocation for the future.

As we see the power of people coming together to create access, making it an active conversation rather than shamefully private, it is exciting to think about how to both bring in nondisabled people and people with different disabilities. The provocation to make access together, to me, has a different valence when speaking to nondisdabled people – it is asking nondisabled people to cede some of the power they hold and to see themselves as able to change and engage in making their interactions more accessible. Engaging nondisabled people, then, is often an educational process and a call toward justice. This is what made this area of research so compelling to me when I began it, identifying at the time as nondisabled. Now, finishing this work, identifying as a disabled person myself, I am also excited by the potential of technology to support disabled people in making spaces we can be in together. Disability communities already use technology as a key tactic for addressing conflicting access needs [231, 167], appropriating technologies that are not purpose-built for collective access. I look forward to future work (be it mine or others [183]) that can build technology to support disabled people engaging in collective access together, highlighting the capacity that disabled people have to lend support to each other [25, 186, 122].

8.3 A Future For Interdisciplinary Disability Research

My dissertation research, and my success in publishing my dissertation research, is beholden to the work of those before me who have pushed the field to understand disability studies as a needed source of theory and analysis within HCI accessibility, many of whom I am lucky to count as friends and mentors. As I also join a cohort of junior scholars increasingly taking up disability studies and activist perspectives, I want to reflect on the future for interdisciplinary disability research.

Over the last 5-15 years[1], HCI practitioners have developed practices of engaging disability studies scholarship and activism, but we have done less to engage disability studies scholars and activists. Some of us identify as disability studies scholars in our own right (with varied degrees of formal education) and more as disability activists, but HCI venues see few publications by authors whose primary affiliation is in disability studies or activism (for some notable exceptions, see [186, 234, 235, 236, 190, 122, 132]). We also see publications on topics such as interdependence that reference primarily HCI sources on the topic, rather than the rich activist sources used by Bennett et al. [25] to introduce the concept to HCI. Further, we do not have sufficient infrastructure to train interested HCI students in academic disability studies. I had the privilege to receive a formal graduate education in disability studies, an option that did not even exist when I began at The University of Washington. UW is one of relatively few universities with an official disability studies program [270] and there are only two official PhD programs in disability studies in the United States[2]. The analytical capacity I now have as someone engaged in disability studies as a field in its own right is significantly stronger than what I possessed when hunting for disability studies articles or perspectives to back up specific arguments. Building infrastructure to support the many HCI PhD students interested in learning disability studies is a crucial piece of building a future for thriving interdisciplinary disability research.

A connected problem to the lack of engagement with disability studies scholars and activists is that HCI accessibility perspectives do not get engaged with in these areas. I do not intend to presume that disability studies and disability activism need HCI accessibility. However, academic

[1]While Mankoff et al.'s [190] first invocation to engage disability studies was published in 2010, there was little research in this vein between then and Bennett et al.'s work on interdependence [25], published late 2018

[2]At the University of Illinois Chicago and Syracuse University

disability studies particularly is preoccupied with the ways that *medical* technologies construct and shape disability. I have spent the last five years searching for an analysis of how many 21st century mainstream disability technologies have become the domain of not medical professionals but computer scientists with limited background in disability and an unstated (but present) idea of what disability and accessibility are and what ought to be done about it. The closest I have come is Elizabeth Petrick's 2015 book [230] *Making Computers Accessible: Disability Rights and Digital Technologies*, which ends its analysis in the late 90s. There is rich scholarship to be done, of a kind that does not get published at ACM venues but in *Disability Studies Quarterly*, to interrogate current HCI accessibility practice and identify opportunities to build a more robust understanding and orientation toward disability in future practice. I think there must be HCI scholars involved in that work, but that it would be strengthened by collaboration with disability studies scholars. How to bridge these fields remains an open question, but one that I argue would be well-worth both fields' effort.

8.4 Limitations and Future Work

While I have achieved both breadth and depth in my dissertation work, many limitations and opportunities for future work remain.

A primary limitation of my work is that it does not implement, deploy, and study the long term use of group captioning tools. I believe this is critical future work, however, it was beyond my capacity and educational priorities during my PhD. When it comes to what tools to build, I argue the ideas my work and others (e.g., [247, 257, 210]) propose for videoconferencing are the most logical place to begin, as the technically mediated nature of online communication easily enables interventions for all interlocutors. I think it will be very important to build and qualitatively study captioning tools for group use, as many open questions around how technology would disrupt group dynamics and when it would be acceptable to disrupt group dynamics remain. I remain curious about the learning effect these types of tools would have, and envision a longitudinal field deployment with mixed hearing ability groups, with one research question focusing on use over time. Such research could determine whether a tool can help teach accessible communication norms that groups then internalize and use to self-regulate or whether ongoing technical support is needed.

I have approached my dissertation as a case study of what emerges when studying one specific communication accessibility technology–captioning–through the lens of collective access. I have not even begun to exhaust the wide range of captioning uses, let alone explored broader communication access contexts. While I studied small-group live captioning and user-generated pre-recorded captioning, one could study collective access in the context of one on one interactions, classroom environments, broadcast television, and many more. Additionally, captioning is far from the only communication access technology. ASL technologies, augmented and alternative communication devices, and many nonvisual access approaches mediate communication access and would be fascinating to study with the lens of collective access.

Given that the start of my dissertation research happened to coincide with the beginning of widely-available, viable automatic speech recognition systems, there is data throughout my findings about how communities adapt to algorithmically generated access tools. As I finish my PhD at a moment when large language models (LLMs) and generative AI (GAI) are proposed as the solutions to many access barriers, I argue that future work specifically studying the role of ASR in DHH communities could illuminate a broader path forward for AI-mediated access. In my work participants have articulated benefits (e.g., on-demand access) and drawbacks (e.g., unreliability) of automatic captions and over the last five years ASR has become a mainstay of captioning workflows. Designers of LLM and GAI-based access tools could turn toward automatic captioning users to better understand how providing an error-laden and bias-prone but free and on-demand access tool impacts disabled communities and what approaches could mitigate potential harms.

While I broadened the population I engaged with my research in my TikTok study, my dissertation work overwhelmingly focuses on the experiences of d/Deaf and hard of hearing people. This is an important population to study, but it is not the only group that benefits from captioning (e.g., neurodiverse people [260, 65]). Following the findings of work that I co-led [186], I reimagined my recruiting criteria to include anyone who needs captions to meet an access need, and I would like to do more research that recruits accordingly. I have studied captioning technology with some hearing people, exploring their perceptions of accessibility when communicating with DHH people. However, as my dissertation argues, hearing people are systematically under-considered as stakeholders in conversational accessibility and more work is needed. Furthermore, the majority of participants I interviewed were white and all were living in the United States at the time we met to conduct

research. This is a limited perspective. Disability studies emphasizes that disability must be considered within its national context [203] – there is not one global opinion on disability issues, and my work is informed by Western-focused literatures and conducted within an American context.

Additionally, despite HCI's prominent focus on DHH people when we study communication accessibility [184], many other communities experience access barriers when communicating. I would like to study many of the arguments I have explored in my dissertation with people with intellectual and developmental disabilities that impact communication, as that is a group whose agency and autonomy is often not prioritized in accessibility research (e.g., [297, 267, 264]). Additionally, many digital access needs are fundamentally communication access needs - information is not communicated accessibly (e.g., screenreader access) or users can't communicate with digital systems accessibly (e.g., lack of keyboard navigability). I look forward to exploring how to approach many classic digital accessibility problems with a collective communication access lens.

Chapter 9

CONCLUSION

I conclude by revisiting my dissertation-level contributions:

- Empirical evidence that social, relational, and environmental context are critical determinants of how captioning technologies mediate access
- Design guidance identifying 1) practical considerations for captioning technology that shapes behavioral norms, 2) guidance for matching caption design to specific contexts, and 3) a design space of low technical complexity, high social impact captioning interventions
- A theoretical framework of collective communication access

9.1 Empirical Contribution

Throughout my dissertation, I argue that to design for collective communication access we must identify and understand the role of contextual features. While considering the ways that social, relational, environmental, and technical factors interact may seem self-evident, I demonstrate in Chapter 7 that, often, captioning research is done without active attention to this context. Throughout my dissertation work I demonstrate a practice of attending to contextual factors and their interplay (particularly in Chapters 4, 5, and 6), resulting in a deeper understanding of how captioning technologies are used in the world and novel directions for captioning design. Synthesizing my dissertation work, rather than reading each chapter in isolation, makes clear the different demands for captioning technologies in different contexts (e.g., during online videoconferencing vs. on TikTok). I envision a future captioning research practice that can build on my work to 1) design studies in a way that makes context visible, 2) tailor interventions to specific contexts, and 3) build greater customization capacity into captioning tools, motivated by tools' varied contexts of use. The work I present in Chapter 4 has already been used to motivate other researchers' study design [257] and as a lens

for analysis [69], which is gratifying and makes me excited for a future of contextual captioning research.

9.2 Design Guidance

My dissertation also contributes guidance that has implications for how we design technologies to encourage accessible behavior change, how to match the styling of captions themselves to broader context, and what kind of technical infrastructure is needed to make communication more accessible.

In Chapters 4 and 5 I provide specific design guidance around features that could be built to guide more accessible communication, assessed by DHH individuals and mixed hearing ability groups. I have been excited to see others take up this work (alongside other research I advised [247]) to guide the design of research prototypes [210], and look forward to future research assessing how these tools are (or are not) taken up by groups over time (either conducted by myself or others).

From Chapter 6, I identify another way contextual, collective research can impact captioning design – by identifying captioning guidelines for user-generated captioning. Though so far developed with a small number of participants, I am excited by to further develop these standards and experiment with disseminating them as another approach to designing collective communication access.

Finally, many of the technology design recommendations I make across my dissertation work do not require a new generation of technical development to be feasibly implemented. While this requires challenging predominant assumptions of what makes for technical innovation, I am excited by a future of accessible technology design that can make tangible improvements to the quality of communication access by identifying the social norms could easily be supported once known.

9.3 Theoretical Framework

Finally, the collective communication access framework I present in Chapter 3 combines five domains infrequently put in conversation with each other (disability studies, Deaf studies, disability justice, communication studies, and HCI), into an approachable and practical guide for future research. As I have demonstrated in Chapter 7, this framework is useful for analyzing and identifying

the often under-stated and unattended to politics embodied in prior work. I also am excited by this framework as a guide for future research. The work I present in Chapters 4, 5, and 6 was informally guided by this framework, as I developed it over the course of years of study and experience conducting this empirical research. I look forward to having a framework and rubric to reach for when planning my future studies, teaching future students, and, hopefully, one that will have a life outside of my research.

9.4 Final Remarks

Throughout my dissertation, I have blended critical disability and Deaf perspectives with accessibility research practice, illuminating but-one area where beginning with a theoretical and activist sense of disability reveals novel opportunities for technology design. My dissertation is just the beginning of collective communication access research, and I am also excited for all the other forms of accessibility research practice that can emerge from this approach (such as my work on chronic illness and accessibility [184]). I have learned so much from disability communities, personally and professionally, over the last five years and am motivated to continue to drive innovations in accessible technologies in a field that is open to critical disability epistemologies.

BIBLIOGRAPHY

[1] 3PlayMedia. Accessibility Laws, Aug 2022. Available at https://www.3playmedia.com/learn/accessibility/accessibility-laws/.

[2] Captioning Activism and Community. What is CART and FAQ's – Find a CART Captioning Provider, September 2012. Available at http://ccacaptioning.org/faqs-cart/.

[3] Rachel Adams, Benjamin Reiss, and David Serlin. *Keywords for disability studies*, volume 7. NYU Press, 2015.

[4] Oliver Alonzo, Hijung Valentina Shin, and Dingzeyu Li. Beyond Subtitles: Captioning and Visualizing Non-speech Sounds to Improve Accessibility of User-Generated Videos. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, pages 1–12, New York, NY, USA, October 2022. Association for Computing Machinery.

[5] Meryl Alper. *Giving voice: Mobile communication, disability, and inequality*. MIT Press, 2017.

[6] Meryl Alper, Jessica Sage Rauchberg, Ellen Simpson, Josh Guberman, and Sarah Feinberg. TikTok as algorithmically mediated biographical illumination: Autism, self-discovery, and platformed diagnosis on# autisktok. *New Media & Society*, page 14614448231193091, 2023.

[7] Hearing Loss Association Of America. Position Papers - Hearing Loss Association of America, 2020. Available at https://www.hearingloss.org/about-hlaa/position-papers/.

[8] Akhter Al Amin. Audio-Visual Caption Evaluation Metric for People who are Deaf and Hard of Hearing. *Articles*, July 2020.

[9] Akhter Al Amin, Saad Hassan, and Matt Huenerfauth. Caption-occlusion severity judgments across live-television genres from deaf and hard-of-hearing viewers. In *Proceedings of the 18th International Web for All Conference*, W4A '21, pages 1–12, New York, NY, USA, May 2021. Association for Computing Machinery.

[10] Akhter Al Amin, Saad Hassan, Sooyeon Lee, and Matt Huenerfauth. 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, pages 1–14, New York, NY, USA, April 2022. Association for Computing Machinery.

[11] Akhter Al Amin, Saad Hassan, Sooyeon Lee, and Matt Huenerfauth. Understanding How Deaf and Hard of Hearing Viewers Visually Explore Captioned Live TV News. In *20th International Web for All Conference*, W4A '23. ACM, April 2023.

[12] Akhter Al Amin, Joseph Mendis, Raja Kushalnagar, Christian Vogler, and Matt Huenerfauth. Who is Speaking: Unpacking In-Text Speaker Identification Preference of Viewers Who Are Deaf and Hard of Hearing While Watching Live Captioned Television Program. In *Proceedings of the 20th International Web for All Conference*, W4A '23, page 44–53, New York, NY, USA, 2023. Association for Computing Machinery.

[13] Android. Live Transcribe | Speech to Text App, 2024. Available at `https://www.android.com/accessibility/live-transcribe/`.

[14] Lawrence Angrave, Zhilin Zhang, Genevieve Henricks, and Chirantan Mahipal. Who Benefits?: Positive Learner Outcomes from Behavioral Analytics of Online Lecture Video Viewing Using ClassTranscribe. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, SIGCSE '20. ACM, February 2020.

[15] National Court Reporters Association. Captioning around the world, June 2014. Available at `https://www.thejcr.com/2014/06/30/captioning-around-the-world/`.

[16] Dean C Barnlund. A Transcational Model of Communication. In *Language Behavior*, pages 43–61. DE GRUYTER, Berlin, Boston, December 1970.

[17] Corey H Basch, Zoe Meleo-Erwin, Joseph Fera, Christie Jaime, and Charles E Basch. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Human vaccines & immunotherapeutics*, 17(8):2373–2377, 2021.

[18] H.-Dirksen L. Bauman. Audism: Exploring the Metaphysics of Oppression. *The Journal of Deaf Studies and Deaf Education*, 9(2):239–246, April 2004. Publisher: Oxford Academic.

[19] H.-Dirksen L. Bauman and Joseph J. Murray. *Deaf Gain: Raising the Stakes for Human Diversity*. U of Minnesota Press, October 2014. Google-Books-ID: ey90DwAAQBAJ.

[20] Macy Bayern. Zoom grew by 574% in less than two months, but Skype for Business reigns supreme, April 2020. Available at `https://www.techrepublic.com/article/zoom-grew-by-574-in-less-than-two-months-but-skype-for-business-reigns-supreme/`.

[21] Gökçe Elif Baykal, Maarten Van Mechelen, and Eva Eriksson. Collaborative Technologies for Children with Special Needs: A Systematic Literature Review. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, New York, NY, USA, April 2020. Association for Computing Machinery.

[22] BBC. Subtitle guidelines, Jul 2022. Available at https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/.

[23] Ursula Bellugi and Susan Fischer. A comparison of sign language and spoken language. *Cognition*, 1(2):173–200, January 1972.

[24] Collin Matthew Belt. American Sign Language is Not English on the Hands, Jul 2013. Available at https://www.lifeprint.com/asl101/topics/history8.htm.

[25] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 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*, pages 161–173, Galway, Ireland, 2018. ACM Press.

[26] R.J. Berger. *Introducing Disability Studies*. Lynne Rienner Publishers, 2013.

[27] Larwan Berke. Displaying confidence from imperfect automatic speech recognition for captioning. *ACM SIGACCESS Accessibility and Computing*, (117):14–18, February 2017.

[28] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers. 2019. ISBN: 9781450359719.

[29] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 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*, pages 155–164, Baltimore Maryland USA, October 2017. ACM.

[30] Larwan Berke, Sushant Kafle, and Matt Huenerfauth. Methods for Evaluation of Imperfect Captioning Tools by Deaf or Hard-of-Hearing Users at Different Reading Literacy Levels. 2018. ISBN: 9781450356206.

[31] Larwan Berke, Matthew Seita, and Matt Huenerfauth. Deaf and hard-of-hearing users' prioritization of genres of online video content requiring accurate captions. In *Proceedings of the 17th International Web for All Conference*, W4A '20, pages 1–12, New York, NY, USA, April 2020. Association for Computing Machinery.

[32] Aparajita Bhandari and Sara Bimo. Why's everyone on TikTok now? the algorithmized self and the future of self-making on social media. *Social Media + Society*, 8(1), Mar 2022.

[33] Jeffrey P Bigham, Raja Kushalnagar, Ting-Hao Kenneth Huang, Juan Pablo Flores, and Saiph Savage. On how deaf people might use speech to control devices. In *Proceedings of the 19th international ACM SIGACCESS conference on computers and accessibility*, pages 383–384, 2017.

[34] Hannah Boland. Google launches free version of meet in bid to topple Zoom, Apr 2020. Available at https://www.telegraph.co.uk/technology/2020/04/29/google-launches-free-version-meet-video-calling-app-bid-topple/.

[35] Sydney Bradley and Dan Whateley. TikTok has surged past Instagram in time spent on the platform, but lags in ad dollars. , Aug 2023. Available at https://www.businessinsider.com/tiktok-surpasses-instagram-time-spent-us-adults-emarketer-8-2023.

[36] Alessandra Brandão, Hugo Nicolau, Shreya Tadas, and Vicki L. Hanson. 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, pages 25–32, New York, NY, USA, October 2016. Association for Computing Machinery.

[37] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[38] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, August 2019. Publisher: Routledge _eprint: https://doi.org/10.1080/2159676X.2019.1628806.

[39] Virginia Braun and Victoria Clarke. *Thematic Analysis: A Practical Guide*. SAGE, October 2021. Google-Books-ID: eMArEAAAQBAJ.

[40] Gabriel Wade Britain, David Martin, Tyler Kwok, Adam Sumilong, and Thad Starner. Preferences for Captioning on Emulated Head Worn Displays While in Group Conversation. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '22. ACM, September 2022.

[41] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '15, pages 103–112, New York, NY, USA, June 2015. Association for Computing Machinery.

[42] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. Subtitles in 360-degree Video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '17. ACM, June 2017.

[43] Emeline Brulé, Brianna J. Tomlinson, Oussama Metatla, Christophe Jouffrais, and Marcos Serrano. Review of Quantitative Empirical Evaluations of Technology for People with Visual

Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, April 2020. Association for Computing Machinery.

[44] Janine Butler. The Visual Experience of Accessing Captioned Television and Digital Videos. *Television & New Media*, 21(7):679–696, November 2020. Publisher: SAGE Publications.

[45] Janine Butler, Brian Trager, and Byron Behm. 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, pages 32–42, Pittsburgh, PA, USA, October 2019. Association for Computing Machinery.

[46] Sebastian Cahill. From TikTok to Netflix, Gen Z is embracing closed captions. We asked them why., Aug 2023. Available at https://www.insider.com/generation-z-subtitles-closed-captioning-millenials-no-hearing-loss-2023-8.

[47] Kelly Caine. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 981–992, New York, NY, USA, May 2016. Association for Computing Machinery.

[48] Maria Camacho. David Kelley: From design to design thinking at Stanford and IDEO. *She Ji: The Journal of Design, Economics, and Innovation*, 2(1):88–101, 2016.

[49] Beiyan Cao, Changyang He, Muzhi Zhou, and Mingming Fan. Sparkling Silence: Practices and Challenges of Livestreaming Among Deaf or Hard of Hearing Streamers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023.

[50] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 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*, page 67, Pittsburgh, Pennsylvania, USA, 2009. ACM Press.

[51] GVC Center. Weaving CHI - Top Keyword Topics, 2019. Available at https://public.tableau.com/views/WeavingCHI-TopKeywordTopics/TopKeywordTopics?:embed=y&:display_count=yes&publish=yes:showVizHome=no.

[52] National Deaf Center. Communicating With Deaf Individuals, September 2017. Available at https://www.nationaldeafcenter.org/resource/communicating-deaf-individuals.

[53] Anfan Chen, Hichang Cho, Richard Evans, and Runxi Zeng. Editorial: Reimagining communication in a post-pandemic world: The intersection of information, media technology, and psychology. *Front. Psychol.*, 14:1154044, February 2023.

[54] J.B. Christiansen and I. Leigh. *Cochlear Implants in Children: Ethics and Choices*. Gallaudet University Press, 2002.

[55] Combahee River Collective, Demita Frazier, Barbara Smith, and Beverly Smith. *The Combahee River Collective Statement*. Gato Negro Ediciones, 2017.

[56] Consumer and Governmental Affairs. *Closed Captioning of Video Programming; Telecommunications for the Deaf and Hard of Hearing, Inc*. Number FCC 14-12. Feb 2014.

[57] Brian Contreras. Why captions are everywhere on TikTok: "glasses for your ears", Sep 2021. Available at `https://www.latimes.com/business/technology/story/2021-09-23/subtitles-once-a-barrier-to-foreign-film-have-found-a-home-on-tiktok`.

[58] Michael Crabb, Rhianne Jones, and Mike Armstrong. The Development of a Framework for Understanding the UX of Subtitles. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15*, ASSETS '15. ACM Press, 2015.

[59] Michael Crabb, Rhianne Jones, Mike Armstrong, and Chris J. Hughes. Online News Videos: The UX of Subtitle Position. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '15, pages 215–222, New York, NY, USA, October 2015. Association for Computing Machinery.

[60] Kimberlé Williams Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. In *The public nature of private violence*, pages 93–118. Routledge, 2013.

[61] CSDHH. *Communication Services for the Deaf and Hard of Hearing*, Jul 2020.

[62] Elaine Czech, Paul Marshall, and Oussama Metatla. "Just like meeting in person" - Examination of interdependencies in dementia-friendly virtual activities. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22, New York, NY, USA, 2022. Association for Computing Machinery.

[63] Maitraye Das, Darren Gergle, and Anne Marie Piper. "It doesn't win you friends': Understanding Accessibility in Collaborative Writing for People with Vision Impairments. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):191:1–191:26, November 2019.

[64] Maitraye Das, Thomas Barlow McHugh, Anne Marie Piper, and Darren Gergle. Co11ab: Augmenting Accessibility in Synchronous Collaborative Writing for People with Vision Impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–18, New York, NY, USA, April 2022. Association for Computing Machinery.

[65] Maitraye Das, John Tang, Kathryn E Ringland, and Anne Marie Piper. Towards accessible remote work: Understanding work-from-home practices of neurodivergent professionals. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–30, 2021.

[66] Hannah Davies. Lights, camera, caption! why subtitles are no longer just for the hard of hearing, Jul 2019. Available at `https://www.theguardian.com/tv-and-radio/2019/jul/21/subtitles-tv-hearing-no-context-twitter-captions`.

[67] Caluã de Lacerda Pataca, Matthew Watkins, Roshan Peiris, Sooyeon Lee, and Matt Huenerfauth. Visualization of Speech Prosody and Emotion in Captions: Accessibility for Deaf and Hard-of-Hearing Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023.

[68] Dominique Decouchant, Sonia Mendoza, Gabriela Sánchez, and José Rodríguez. Adapting groupware systems to changes in the collaborator's context of use. *Expert Systems with Applications*, 40(11):4446–4462, 2013.

[69] Aashaka Desai, Jennifer Mankoff, and Richard E Ladner. Understanding and Enhancing Speechreading Experiences for Accessible Communication in an Online Context. *In submission to CHI 2023: The ACM CHI Conference on Human Factors in Computing Systems*, page 25, April 2023.

[70] Described and Captioned Media Project. Captioning Key, 2023. Available at `https://dcmp.org/learn/captioningkey`.

[71] Michael Ann DeVito. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):380:1–380:31, November 2022.

[72] Becca Dingman, Garreth W. Tigwell, and Kristen Shinohara. Designing a Podcast Platform for Deaf and Hard of Hearing Users. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21, pages 1–4, New York, NY, USA, October 2021. Association for Computing Machinery.

[73] DO-IT. What is the difference between open and closed captioning?, Apr 2021. Available at `https://www.washington.edu/doit/what-difference-between-open-and-closed-captioning`.

[74] DREDF. NAD v. Netflix, Mar 2014. Available at https://dredf.org/legal-advocacy/nad-v-netflix/.

[75] Fabio Duarte. Google workspace user stats (2024), Dec 2023. Available at https://explodingtopics.com/blog/google-workspace-stats.

[76] Katie Dupere. Deaf youtubers call out crappy captions with Hashtag-NoMoreCraptions movement, Oct 2021. Available at https://mashable.com/2016/11/11/youtube-closed-captioning-nomorecraptions/.

[77] Jared Duval, Ferran Altarriba Bertran, Siying Chen, Melissa Chu, Divya Subramonian, Austin Wang, Geoffrey Xiang, Sri Kurniawan, and Katherine Isbister. Chasing Play on TikTok from Populations with Disabilities to Inspire Playful and Inclusive Technology Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–15, New York, NY, USA, May 2021. Association for Computing Machinery.

[78] Tessa Eagle and Kathryn E. Ringland. "You Can't Possibly Have ADHD": Exploring Validation and Tensions around Diagnosis within Un-bounded ADHD Social Media Communities. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23, New York, NY, USA, October 2023. Association for Computing Machinery.

[79] Gil Eastman and Mary Lou Novitsky. Deaf Mosaic Ep 402: Deaf President Now, Mar 1988. Available at https://media.gallaudet.edu/media/Deaf+Mosaic+402A+Deaf+President+Now/1_09xzy4hx.

[80] Shari Eberts. Opinion — people with hearing loss shouldn't have to pay Zoom for captions, Dec 2020. Available at https://www.washingtonpost.com/opinions/2020/12/17/zoom-captions-hearing-loss-free/.

[81] Claire Edwards and Gill Harold. DeafSpace and the principles of universal design. *Disability and Rehabilitation*, 36(16):1350–1359, August 2014.

[82] Ryohei Egusa, Shuya Kawaguchi, Tsugunosuke Sakai, Hiroshi Mizoguchi, Fusako Kusunoki, Miki Namatame, and Shigenori Inagaki. Implementation and Evaluation of Accessible Caption System in Universal Puppetry: A Case Study on Hearing Impaired Children at the Elementary School of Deaf. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '17. ACM, October 2017.

[83] Frank Elavsky. Okay, I want to be as gentle as possible with a review of this now-award-winning work, April 2023. Available at https://twitter.com/FrankElavsky/status/1643627602882486276.

[84] Elizabeth Ellcessor. Captions On, Off, on TV, Online: Accessibility and Search Engine Optimization in Online Closed Captioning. *Television & New Media*, 13(4):329–352, July 2012. Publisher: SAGE Publications.

[85] Lisa Elliot, Michael Stinson, Donna Easton, and Jennifer Bourgeois. College students' learning with C-Print's educational software and automatic speech recognition. In *American Educational Research Association Annual Meeting*, New York, NY, USA, March 2008.

[86] Lisa Elliot, Michael Stinson, James Mallory, Donna Easton, and Matt Huenerfauth. Deaf and Hard of Hearing Individuals' Perceptions of Communication with Hearing Colleagues in Small Groups. 2016. ISBN: 9781450341240.

[87] Lisa B. Elliot, Michael Stinson, Syed Ahmed, and Donna Easton. 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*, pages 405–406, Baltimore Maryland USA, October 2017. ACM.

[88] Sheena Erete, Yolanda Rankin, and Jakita O. Thomas. Defining Damage-Centered Research in HCI: A Black Feminist Perspective. *Interactions*, 31(3):41–45, may 2024.

[89] Moa Eriksson Krutrök. Algorithmic Closeness in Mourning: Vernaculars of the Hashtag #grief on TikTok. *Social Media + Society*, 7(3):20563051211042396, Jul 2021.

[90] Esther Erlich and Niel Marcus. Neil Marcus: Performance Artist, 2004.

[91] FCC. Closed Captioning Quality Report and Order, Declaratory Ruling, FNPRM, December 2015. Available at `https://www.fcc.gov/document/closed-captioning-quality-report-and-order-declaratory-ruling-fnprm`.

[92] FCC. Closed Captioning of Internet Video Programming, jan 2021. Available at `https://www.fcc.gov/consumers/guides/captioning-internet-video-programming`.

[93] Maria Federico, Servizio Accoglienza, Studenti Disabili, and Marco Furini. *Enhancing Learning Accessibility through Fully Automatic Captioning*. 2012.

[94] Peter Feng, David Martin, and Thad Starner. ToozKit: System for Experimenting with Captions on a Head-worn Display. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, UbiComp/ISWC '23. ACM, October 2023.

[95] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, New York, NY, USA, May 2019. Association for Computing Machinery.

[96] Jon Froehlich, Leah Findlater, Marilyn Ostergren, Solai Ramanathan, Josh Peterson, Inness Wragg, Eric Larson, Fabia Fu, Mazhengmin Bai, Shwetak Patel, and James A. Landay. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2367–2376, New York, NY, USA, May 2012. Association for Computing Machinery.

[97] Jayden Galamgam and Justin L Jia. "Accutane Check": insights into youth sentiment toward isotretinoin from a TikTok trend. *Pediatric Dermatology*, 38(4):980–981, 2021.

[98] Abraham Glasser, Joseline Garcia, Chang Hwang, Christian Vogler, and Raja Kushalnagar. Effect of caption width on the TV user experience by deaf and hard of hearing viewers. In *Proceedings of the 18th International Web for All Conference*, W4A '21. ACM, April 2021.

[99] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 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*, pages 427–432, Baltimore Maryland USA, October 2017. ACM.

[100] Abraham Glasser, Edward Mason Riley, Kaitlyn Weeks, and Raja Kushalnagar. Mixed Reality Speaker Identification as an Accessibility Tool for Deaf and Hard of Hearing Users. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. ACM, November 2019.

[101] Abraham T. Glasser, Kesavan R. Kushalnagar, and Raja S. Kushalnagar. 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*, pages 373–374, Baltimore Maryland USA, October 2017. ACM.

[102] Sara Goering. Rethinking disability: the social model of disability and chronic disease. *Current Reviews in Musculoskeletal Medicine*, 8(2):134–138, June 2015.

[103] Steven Goodman, Susanne Kirchner, Rose Guttman, Dhruv Jain, Jon Froehlich, and Leah Findlater. Evaluating Smartwatch-based Sound Feedback for Deaf and Hard-of-hearing Users Across Contexts. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, New York, NY, USA, April 2020. Association for Computing Machinery.

[104] Google. How google is improving technology for deaf people, Dec 2021. Available at `https://about.google/intl/ALL_us/stories/making-conversation-more-accessible-with-live-transcribe/`.

[105] Benjamin Gorman, Michael Crabb, and Michael Armstrong. *Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption*. May 2021.

[106] Michael Gower, Brent Shiver, Charu Pandhi, and Shari Trewin. Leveraging Pauses to Improve Video Captions. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, pages 414–416, New York, NY, USA, October 2018. Association for Computing Machinery.

[107] Saul Greenberg. Personalizable groupware: Accommodating individual roles and group differences. In Liam Bannon, Mike Robinson, and Kjeld Schmidt, editors, *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91*, pages 17–31. Springer Netherlands, Dordrecht, 1991.

[108] Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-mounted Display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, pages 1–4, New York, NY, USA, October 2020. Association for Computing Machinery.

[109] Victoria Hallett. Google live transcribe reaches over one billion downloads, Dec 2023. Available at `https://gallaudet.edu/information-technology/b-s-in-information-technology/google-live-transcribe-reaches-over-one-billion-downloads/`.

[110] Aimi Hamraie. Designing Collective Access: A Feminist Disability Theory of Universal Design. *Disability Studies Quarterly*, 33(4), September 2013.

[111] Aimi Hamraie. *Crip Making*, pages 303–317. New York University Press, New York, USA, 2023.

[112] Aimi Hamraie and Kelly Fritsch. Crip Technoscience Manifesto. *Catalyst: Feminism, Theory, Technoscience*, 5(1):1–33, April 2019. Number: 1.

[113] Christina Harrington and Tawanna R Dillahunt. 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, pages 1–15, New York, NY, USA, May 2021. Association for Computing Machinery.

[114] Christina N. Harrington, Aashaka Desai, Aaleyah Lewis, Sanika Moharana, Anne Spencer Ross, and Jennifer Mankoff. Working at the Intersection of Race, Disability and Accessibility. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23, New York, NY, USA, 2023. Association for Computing Machinery.

[115] Rebecca Perkins Harrington and Gregg C. Vanderheiden. Crowd caption correction (CCC). In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, pages 1–2, Bellevue, Washington, October 2013. Association for Computing Machinery.

[116] Saad Hassan, Yao Ding, Agneya Abhimanyu Kerure, Christi Miller, John Burnett, Emily Biondo, and Brenden Gilbert. Exploring the Design Space of Automatically Generated Emotive Captions for Deaf or Hard of Hearing Users. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023.

[117] Jon Henner and Octavian Robinson. Unsettling Languages, Unruly Bodyminds: A Crip Linguistics Manifesto. *Journal of Critical Study of Communication and Disability*, 1(1):7–37, May 2023.

[118] Nicole Hernandez. Hearing loss statistics 2024: More common than you might think, Nov 2023. Available at `https://www.ncoa.org/adviser/hearing-aids/hearing-loss-statistics/`.

[119] Shannon S. C. Herrick, Laura Hallward, and Lindsay R. Duncan. "This is just how I cope": An inductive thematic analysis of eating disorder recovery content created and shared on TikTok using EDrecovery. *The International Journal of Eating Disorders*, 54(4):516–526, Apr 2021.

[120] Louise Hickman. *32. Willful Dictionaries and Crip Authorship in CART*, pages 332–336. New York University Press, New York, USA, 2023.

[121] Stephanie Hind. Introducing auto captions, Apr 2021. Available at `https://newsroom.tiktok.com/en-us/introducing-auto-captions`.

[122] Megan Hofmann, Devva Kasnitz, Jennifer Mankoff, and Cynthia L Bennett. Living Disability Theory: Reflections on Access, Research, and Design. page 13, 2020.

[123] Scott Hollier and Shadi Abou-Zahra. Internet of Things (IoT) as Assistive Technology: Potential Applications in Tertiary Education. In *Proceedings of the 15th International Web for All Conference*, W4A '18. ACM, April 2018.

[124] Richang Hong, Meng Wang, Mengdi Xu, Shuicheng Yan, and Tat-Seng Chua. Dynamic captioning: video accessibility enhancement for hearing impairment. In *Proceedings of the 18th ACM international conference on Multimedia*, MM '10, pages 421–430, Firenze, Italy, October 2010. Association for Computing Machinery.

[125] Stacy Hsueh, Beatrice Vincenzi, Akshata Murdeshwar, and Marianela Ciolfi Felice. Cripping Data Visualizations: Crip Technoscience as a Critical Lens for Designing Digital Access. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–16, 2023.

[126] Yongtao Hu, Jan Kautz, Yizhou Yu, and Wenping Wang. Speaker-Following Video Subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(2):1–17, January 2015.

[127] Yun Huang, Yifeng Huang, Na Xue, and Jeffrey P. Bigham. Leveraging Complementary Contributions of Different Workers for Efficient Crowdsourcing of Video Captions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17. ACM, May 2017.

[128] Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. Responsive design for personalised subtitles. In *Proceedings of the 12th International Web for All Conference*, W4A '15. ACM, May 2015.

[129] Hilary Hutchinson, Ben Bederson, Allison Druin, Catherine Plaisant, Wendy Mackay, Helen Evans, Heiko Hansen, Stéphane Conversy, Michel Beaudouin-Lafon, Nicolas Roussel, Loïc Lacomme, Björn Eiderbäck, Sinna Lindquist, Yngve Sundblad, and Bo Westerlund. Technology Probes: Inspiring Design for and with Families. *Conference on Human Factors in Computing Systems - Proceedings*, February 2003.

[130] Sins Invalid. 10 Principles of Disability Justice, Sep 2015. Available at `https://www.sinsinvalid.org/blog/10-principles-of-disability-justice`.

[131] Sins Invalid. *Skin Tooth and Bone: The Basis of Movement is Our People, a Disability Justice Primer*. Sins Invalid, 2nd edition, November 2019.

[132] Liz Jackson, Alex Haagaard, and Rua Williams. Disability dongle. *Platypus: The CASTAC Blog*, 2022.

[133] Vanessa Jackson. In Our Own Voice: African-American Stories of Oppression, Survival and Recovery. *Off Our Backs*, 33(7/8):19–21, 2003.

[134] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, DIS '18. ACM, May 2018.

[135] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 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, pages 241–250, Seoul, Republic of Korea, April 2015. Association for Computing Machinery.

[136] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 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*, ASSETS '18. ACM, October 2018.

[137] Cordilia James. Why do all these 20-somethings have closed captions turned on?, Sep 2022. Available at =https://www.wsj.com/articles/why-do-all-these-20-somethings-have-closed-captions-turned-on-11663386473.

[138] C. Jensema, R. McCann, and S. Ramsey. Closed-captioned television presentation speed and vocabulary. *American Annals of the Deaf*, 141(4):284–292, October 1996.

[139] Carl Jensema. Viewer Reaction to Different Television Captioning Speeds. *American Annals of the Deaf*, 143(4):318–324, 1998.

[140] Richard G. Jones. *Communication in the real world: An introduction to communication studies*. Flat World Knowledge, 2013.

[141] Heesoon Jun. *Disablism/Ableism*, pages 271–302. Springer Nature Switzerland, Cham, 2024.

[142] Alison Kafer. *Feminist, Queer, Crip*. Indiana University Press, May 2013. Google-Books-ID: F4X6yaiCNOcC.

[143] Sushant Kafle, Becca Dingman, and Matt Huenerfauth. 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, October 2021.

[144] Sushant Kafle and Matt Huenerfauth. Evaluating the Usability of Automatically Generated Captions for People who are Deaf or Hard of Hearing. 2017. ISBN: 9781450349260.

[145] Sushant Kafle, Peter Yeung, and Matt Huenerfauth. Evaluating the Benefit of Highlighting Key Words in Captions for People who are Deaf or Hard of Hearing. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 43–55, Pittsburgh PA USA, October 2019. ACM.

[146] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):305:1–305:44, October 2021.

[147] Devva Kasnitz. Discussion on Disability Studies and Technology, November 2020.

[148] Devva Kasnitz. The Politics of Disability Performativity: An Autoethnography. *Current Anthropology*, 61(S21):S16–S25, February 2020.

[149] Saba Kawas, George Karalis, Tzu Wen, and Richard E. Ladner. 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*, pages 15–23, Reno Nevada USA, October 2016. ACM.

[150] JooYeong Kim, SooYeon Ahn, and Jin-Hyuk Hong. Visible Nuances: A Caption System to Visualize Paralinguistic Speech Cues for Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023.

[151] Daniel Klug, Yiluo Qin, Morgan Evans, and Geoff Kaufman. Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. In *13th ACM Web Science Conference 2021*, WebSci '21, pages 84–92, New York, NY, USA, June 2021. Association for Computing Machinery.

[152] Daniel Klug, Ella Steen, and Kathryn Yurechko. How Algorithm Awareness Impacts Algospeak Use on TikTok. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 234–237, New York, NY, USA, 2023. Association for Computing Machinery.

[153] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689, April 2020. Publisher: National Academy of Sciences Section: Social Sciences.

[154] Wenwen Kong, Shijie Song, Yuxiang Chris Zhao, Qinghua Zhu, and Ling Sha. TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos. *Journal of Medical Internet Research*, 23(9):e30409, Sep 2021.

[155] Barbara Krasnoff. Microsoft Teams and google meet both have closed captions, here's how they compare, Jun 2020. Available at `https://www.theverge.com/21291452/microsoft-teams-google-meet-closed-captions-cc-zoom-comparison`.

[156] Raja Kushalnagar. A Classroom Accessibility Analysis App for Deaf Students. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19. ACM, October 2019.

[157] Raja S. Kushalnagar, Gary W. Behm, Aaron W. Kelstone, and Shareef Ali. 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 & Accessibility*, ASSETS '15, pages 223–230, Lisbon, Portugal, October 2015. Association for Computing Machinery.

[158] Raja S. Kushalnagar, Gary W. Behm, Kevin T. Wolfe, Peter Yeung, Becca Dingman, Shareef Sayel Ali, Abraham Glasser, and Claire Elizabeth Ryan. RTTD-ID: Tracked Captions with Multiple Speakers for Deaf Students. June 2018.

[159] Raja S Kushalnagar, Walter S Lasecki, and Jeffrey P Bigham. *A Readability Evaluation of Real-Time Crowd Captions in the Classroom*. 2012.

[160] Raja S Kushalnagar, Walter S Lasecki, and Jeffrey P Bigham. *Captions Versus Transcripts for Online Video Content*. 2013.

[161] Raja S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. Accessibility Evaluation of Classroom Captions. *ACM Transactions on Accessible Computing*, 5(3):1–24, January 2014.

[162] Raja S. Kushalnagar, John J. Rivera, Warrance Yu, and Daniel S. Steed. Avd-lv: an accessible player for captioned stem videos. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility - ASSETS '14*, ASSETS '14. ACM Press, 2014.

[163] Raja S. Kushalnagar and Christian Vogler. Teleconference Accessibility and Guidelines for Deaf and Hard of Hearing Users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, pages 1–6, New York, NY, USA, October 2020. Association for Computing Machinery.

[164] Annelies Kusters, Maartje De Meulder, Dai O'Brien, et al. Innovations in deaf studies: Critically mapping the field. *Innovations in deaf studies: The role of deaf scholars*, 12:1–53, 2017.

[165] Veton Këpuska. Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx). *International Journal of Engineering Research and Applications*, 07(03):20–24, March 2017.

[166] Paddy Ladd. *Understanding Deaf culture: In search of Deafhood*. Multilingual Matters, 2003.

[167] Leah Lakshmi Piepzna Samarasinha. *The Future is Disabled: Prophecies, Love Notes, and Mourning Songs*. Arsenal Pulp Press, 1st edition, 2022.

[168] Harlan Lane. Ethnicity, Ethics, and the Deaf-World. *The Journal of Deaf Studies and Deaf Education*, 10(3):291–310, July 2005. Publisher: Oxford Academic.

[169] Walter S. Lasecki and Jeffrey P. Bigham. Online quality control for real-time crowd captioning. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*, page 143, Boulder, Colorado, USA, 2012. ACM Press.

[170] Walter S. Lasecki, Raja Kushalnagar, and Jeffrey P. Bigham. Helping students keep up with real-time captions by pausing and highlighting. In *Proceedings of the 11th Web for All Conference on - W4A '14*, pages 1–8, Seoul, Korea, 2014. ACM Press.

[171] Jaclyn Lazzari. Are automatic captions helping or hindering students?, Jun 2019. Available at `https://www.3playmedia.com/blog/automatic-captions-helping-hindering-students/`.

[172] Daniel G. Lee, Deborah I. Fels, and John Patrick Udo. Emotive captioning. *Computers in Entertainment*, 5(2):11, April 2007.

[173] Rachel Lerman. Social media has upped its accessibility game. But deaf creators say it has a long way to go. *Washington Post*, April 2021.

[174] Talila TL Lewis. Working Definition of Ableism - January 2022 Update, Jan 2022. Available at `https://www.talilalewis.com/blog/working-definition-of-ableism-january-2022-update`.

[175] Franklin Mingzhe Li, Cheng Lu, Zhicong Lu, Patrick Carrington, and Khai N. Truong. An Exploration of Captioning Practices and Challenges of Individual Content Creators on YouTube for People with Hearing Impairments. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):75:1–75:26, April 2022.

[176] Chen Ling, Krishna P Gummadi, and Savvas Zannettou. "Learn the Facts About COVID-19": Analyzing the Use of Warning Labels on TikTok Videos. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 554–565, 2023.

[177] Xingyu "Bruce" Liu, Ruolin Wang, Dingzeyu Li, Xiang Anthony Chen, and Amy Pavel. CrossA11y: Identifying Video Accessibility Issues via Cross-modal Grounding. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, pages 1–14, New York, NY, USA, October 2022. Association for Computing Machinery.

[178] Fernando Loizides, Sara Basson, Dimitri Kanevsky, Olga Prilepova, Sagar Savla, and Susanna Zaraysky. Breaking Boundaries with Live Transcribe: Expanding Use Cases Beyond Standard Captioning Scenarios. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20. ACM, October 2020.

[179] Elizabeth R Lorah, Ashley Parnell, Peggy Schaefer Whitby, and Donald Hantula. A systematic review of tablet computers and portable media players as speech generating devices for individuals with autism spectrum disorder. *J. Autism Dev. Disord.*, 45(12):3792–3804, December 2015.

[180] Kim Lyons. Zoom now has auto-generated captions available for free accounts, October 2021. Available at `https://www.theverge.com/2021/10/25/22744704/zoom-auto-generated-captions-available-free-accounts-accessibility`.

[181] Kelly Avery Mack, Danielle Bragg, Meredith Ringel Morris, Maarten W. Bos, Isabelle Albi, and Andrés Monroy-Hernández. Social app accessibility for deaf signers. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–31, October 2020.

[182] Kelly Avery Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. Designing Tools for High-Quality Alt Text Authoring. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21, New York, NY, USA, 2021. Association for Computing Machinery.

[183] Kelly Avery Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, John Tang, Andrew Begel, Erin Beneteau, Josh Urban Davis, Abraham Glasser, Joon Sung Park, and Venkatesh Potluri. 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, pages 1–13, New York, NY, USA, October 2021. Association for Computing Machinery.

[184] Kelly Avery Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. What Do We Mean by "Accessibility Research"?: A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 2021. Association for Computing Machinery.

[185] Kelly Avery Mack, Emma McDonnell, Venkatesh Potluri, Maggie Xu, Jailyn Zabala, Jeffrey Bigham, Jennifer Mankoff, and Cynthia Bennett. Anticipate and Adjust: Cultivating Access in Human-Centered Methods. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–18, New York, NY, USA, April 2022. Association for Computing Machinery.

[186] Kelly Avery Mack, Emma J. McDonnell, Leah Findlater, and Heather D. Evans. Chronically Under-Addressed: Considerations for HCI Accessibility Practice with Chronically Ill People. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–15, Athens Greece, October 2022. ACM.

[187] Ability Magazine. Google launches live captions on Google Chrome!, Mar 2021. Available at `https://abilitymagazine.com/google-launches-live-captions-on-google-chrome/`.

[188] James R. Mallory, Michael Stinson, Lisa Elliot, and Donna Easton. 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*, pages 419–421, Baltimore Maryland USA, October 2017. ACM.

[189] Jennifer Mankoff, Gillian R. Hayes, and Devva Kasnitz. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 3–10, Orlando, Florida, USA, October 2010. Association for Computing Machinery.

[190] Jennifer Mankoff, Kateryna Kuksenok, Sara Kiesler, Jennifer A. Rode, and Kelly Waldman. Competing online viewpoints and models of chronic illness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 589–598, New York, NY, USA, May 2011. Association for Computing Machinery.

[191] Giacomo Marzi, Marco Balzano, and Davide Marchiori. K-Alpha Calculator–Krippendorff's alpha calculator: A user-friendly tool for computing krippendorff's alpha inter-rater reliability coefficient. *MethodsX*, 12(102545):102545, June 2024.

[192] Roshan Mathew, Brian Mak, and Wendy Dannels. Access on Demand: Real-time, Multimodal Accessibility for the Deaf and Hard-of-Hearing based on Augmented Reality. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '22. ACM, October 2022.

[193] Lloyd May, So Yeon Park, and Jonathan Berger. Enhancing Non-Speech Information Communicated in Closed Captioning Through Critical Design. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23. ACM, October 2023.

[194] Emma J McDonnell, Tessa Eagle, Pitch Sinlapanuntakul, Soo Hyun Moon, Kathryn E. Ringland, Jon E. Froehlich, and Leah Findlater. "Caption It in an Accessible Way That Is Also Enjoyable": Characterizing User-Driven Captioning Practices on TikTok. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[195] Emma J McDonnell and Leah Findlater. Envisioning Collective Communication Access: A Theoretically Grounded Review of Captioning Research from 2013-2023. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*, Oct 2024.

[196] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 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, October 2021.

[197] Emma J. McDonnell, Kelly Avery Mack, Kathrin Gerling, Katta Spiel, Cynthia L. Bennett, Robin N. Brewer, Rua Mae Williams, and Garreth W. Tigwell. Tackling the Lack of a Practical Guide in Disability-Centered Research. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23, New York, NY, USA, 2023. Association for Computing Machinery.

[198] Emma J. McDonnell, Soo Hyun Moon, Lucy Jiang, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. "Easier or Harder, Depending on Who the Hearing Person Is": Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status. In *The ACM CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, April 2023. ACM.

[199] Microsoft. Live captions and subtitles are coming to skype and powerpoint, Dec 2018. Available at https://news.microsoft.com/en-gb/2018/12/04/live-captions-and-subtitles-are-coming-to-skype-and-powerpoint/.

[200] Ashley Miller, Joan Malasig, Brenda Castro, Vicki L. Hanson, Hugo Nicolau, and Alessandra Brandão. The Use of Smart Glasses for Lecture Comprehension by Deaf and Hard of Hearing Students. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI '17. ACM, May 2017.

[201] Dorian Miller, Karl Gyllstrom, David Stotts, and James Culp. Semi-transparent video interfaces to assist deaf persons in meetings. In *Proceedings of the 45th annual southeast regional conference*, ACM-SE 45, pages 501–506, Winston-Salem, North Carolina, March 2007. Association for Computing Machinery.

[202] Mara Mills. Deaf Jam: From Inscription to Reproduction to Information. *Social Text*, 28(1 (102)):35–58, March 2010.

[203] Mara Mills and Rebecca Sanchez. *Crip Authorship: Disability as Method*. NYU Press, 2023.

[204] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. "I See Me Here": Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, Hamburg Germany, April 2023. ACM.

[205] Mia Mingus. Reflections on an opening: Disability Justice and creating collective access in Detroit, Aug 2010. Available at https://leavingevidence.wordpress.com/2010/08/23/reflections-on-an-opening-disability-justice-and-creating-collective-access-in-detroit.

[206] Mia Mingus. Access Intimacy: The Missing Link, May 2011. Available at https://leavingevidence.wordpress.com/2011/05/05/access-intimacy-the-missing-link/.

[207] Mia Mingus. Changing the Framework: Disability Justice, February 2011. Available at https://leavingevidence.wordpress.com/2011/02/12/changing-the-framework-disability-justice/.

[208] Mia Mingus. Access Intimacy, Interdependence and Disability Justice, April 2017. Available at https://leavingevidence.wordpress.com/2017/04/12/access-intimacy-interdependence-and-disability-justice/.

[209] Ross E Mitchell, Travas A Young, Bellamie Bachleda, and Michael A Karchmer. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies*, 6(3):306–335, 2006.

[210] Anant Mittal, Meghna Gupta, Roshni Poddar, Tarini Naik, Seethalakshmi Kuppuraj, James Fogarty, Pratyush Kumar, and Mohit Jain. Jod: Examining Design and Implementation of a Videoconferencing Platform for Mixed Hearing Groups. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–18, 2023.

[211] Donald F. Moores. The History of Language and Communication Issues in Deaf Education. In *The Oxford Handbook of Deaf Studies, Language, and Education, Vol. 2*. Oxford University Press, 06 2010.

[212] NAD. Community and culture – frequently asked questions, 2023. Available at `https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/`.

[213] Michelle R Nario-Redmond, Dobromir Gospodinov, and Angela Cobb. Crip for a day: The unintended negative consequences of disability simulations. *Rehabilitation psychology*, 62(3):324, 2017.

[214] NCRA. NCRA statistics, 2023. Available at `https://www.ncra.org/home/about-ncra/NCRA-Statistics`.

[215] NCRA. Certified Realtime Captioner (CRC), 2024. Available at `https://www.ncra.org/certification/NCRA-Certifications/certified-realtime-captioner`.

[216] BBC News. TikTok named as the most downloaded app of 2020. *BBC News*, Aug 2021.

[217] OIT News. Google Adds Live Captionign to Slides and Hangouts Meet, Jun 2019. Available at `https://oit.ncsu.edu/2019/06/13/google-adds-live-captioning-to-slides-and-hangouts-meet/`.

[218] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[219] Marije Nouwen and Mathilde Hermine Christine Marie Ghislaine Duflos. TikTok as a data gathering space: the case of grandchildren and grandparents during the COVID-19 pandemic. In *Interaction Design and Children*, IDC '21, pages 498–502, New York, NY, USA, June 2021. Association for Computing Machinery.

[220] Michael Oliver. *Social Work with Disabled People*. Macmillan, 1983. Google-Books-ID: pzSmQgAACAAJ.

[221] Michael Oliver. Disability and Dependency. A Creation of Industrial Societies? In L Barton, editor, *Disability and Dependency*, pages 7–22, Sussex, 1986. The Falmer Press.

[222] Toby Olson. How Disability Simulations Promote Damaging Stereotypes, Jan 2014. Available at https://nfb.org/sites/default/files/images/nfb/publications/bm/bm14/bm1401/bm140107.htm.

[223] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-day Captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 1108–1120. Association for Computing Machinery, New York, NY, USA, October 2020.

[224] Kotaro Oomori, Akihisa Shitara, Tatsuya Minagawa, Sayan Sarcar, and Yoichi Ochiai. A Preliminary Study on Understanding Voice-only Online Meetings Using Emoji-based Captioning for Deaf or Hard of Hearing Users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, pages 1–4, New York, NY, USA, October 2020. Association for Computing Machinery.

[225] Pilar Orero, Marta Brescia-Zapata, and Chris Hughes. Evaluating subtitle readability in media immersive environments. In *Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, DSAI 2020. ACM, December 2020.

[226] Andrew D. Ouzts, Nicole E. Snell, Prabudh Maini, and Andrew T. Duchowski. Determining optimal caption placement using eye tracking. In *Proceedings of the 31st ACM international conference on Design of communication*, SIGDOC '13, pages 189–190, New York, NY, USA, September 2013. Association for Computing Machinery.

[227] CHI PC. CHI 2024 – papers track, post-PC outcomes report, Jan 2024. Available at https://chi2024.acm.org/2024/01/29/chi-2024-papers-track-post-pc-outcomes-report/.

[228] 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. 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, pages 1–10, Montreal QC, Canada, April 2018. Association for Computing Machinery.

[229] Paola Pennisi, Alessandro Tonacci, Gennaro Tartarisco, Lucia Billeci, Liliana Ruta, Sebastiano Gangemi, and Giovanni Pioggia. Autism and social robotics: A systematic review. *Autism Research*, 9(2):165–183, 2016.

[230] E.R. Petrick. *Making Computers Accessible: Disability Rights and Digital Technology*. Book collections on Project MUSE. Johns Hopkins University Press, 2015.

[231] Leah Lakshmi Piepzna-Samarasinha. *Care Work: Dreaming Disability Justice*. 2018. OCLC: 1082979407.

[232] Agnès Piquard-Kipffer, Odile Mella, Jérémy Miranda, Denis Jouvet, and Luiza Orosanu. 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*, pages 36–41, Dresden, Germany, September 2015. Association for Computational Linguistics.

[233] David Player. The power of white gaze: Erasure of black signers, Jan 2024. Available at `https://dplayer84.medium.com/the-power-of-white-gaze-erasure-of-black-signers-77de766e71d8`.

[234] Lauren Race, Chancey Fleet, Joshua A Miele, Tom Igoe, and Amy Hurst. Designing tactile schematics: Improving electronic circuit accessibility. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 581–583, 2019.

[235] Lauren Race, Claire Kearney-Volpe, Chancey Fleet, Joshua A Miele, Tom Igoe, and Amy Hurst. Designing educational materials for a blind arduino workshop. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2020.

[236] Lauren Race, Joshua A Miele, Chancey Fleet, Tom Igoe, and Amy Hurst. Putting tools in hands: Designing curriculum for a nonvisual soldering workshop. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–4, 2020.

[237] Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. Take My Hand: Automated Hand-Based Spatial Guidance for the Visually Impaired, 2023. Available at `https://adildsw.github.io/#/projects/fingerrover/`.

[238] Raisa Rashid, Quoc Vy, Richard Hunt, and Deborah Fels. Dancing with Words: Using Animated Text for Captioning. *Int. J. Hum. Comput. Interaction*, 24:505–519, June 2008.

[239] Kevin Rathbun, Larwan Berke, Christopher Caulfield, Michael Stinson, and Matt Huenerfauth. Eye Movements of Deaf and Hard of Hearing Viewers of Automatic Captions. *California State University, Northridge.*, 2017. Accepted: 2017-04-26T18:56:38Z Publisher: California State University, Northridge.

[240] Jessica Sage Rauchberg. #Shadowbanned: Queer, Trans, and Disabled creator responses to algorithmic oppression on TikTok. In *LGBTQ Digital Cultures*, pages 196–209. Routledge, 2022.

[241] John Roach. AI technology helps students who are deaf learn, April 2018. Available at `https://blogs.microsoft.com/ai/ai-powered-captioning/`.

[242] Octavian Robinson. "We Are of a Different Class: Ableist Rhetoric in Deaf America, 1880-1920". June 2010.

[243] Octavian E. Robinson and Jonathan Henner. The personal is political in *The Deaf Mute Howls* : deaf epistemology seeks disability justice. *Disability & Society*, 32(9):1416–1436, October 2017.

[244] Anwesha Roy. How Reliable is Speech-to-Text in 2021?, March 2021. Available at `https://www.cxtoday.com/speech-analytics/how-reliable-is-speech-to-text-in-2021/`.

[245] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James Landay. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–23, January 2018. arXiv: 1608.07323.

[246] April Rubin. Why young people can't get enough of subtitles, Aug 2023. Available at `https://www.axios.com/2023/08/20/gen-z-millennials-tv-movies-subtitles`.

[247] Jazz Rui Xia Ang, Ping Liu, Emma McDonnell, and Sarah Coppola. "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, pages 1–16, New York, NY, USA, April 2022. Association for Computing Machinery.

[248] Alex M. Russell, Robert E. Davis, Juanybeth M. Ortega, Jason B. Colditz, Brian Primack, and Adam E. Barry. Alcohol: Portrayals of Alcohol in Top Videos on TikTok. *Journal of Studies on Alcohol and Drugs*, 82(5):615–622, Sep 2021.

[249] Ellen Samuels. Six Ways of Looking at Crip Time. *Disability Studies Quarterly*, 37(3), August 2017. Number: 3.

[250] Elizabeth B.-N. Sanders and Pieter Jan Stappers. Co-creation and the new landscapes of design. *CoDesign*, 4(1):5–18, March 2008.

[251] Marsha Saxton. Disability Rights and Selective Abortion. In Lennard J. Davis, editor, *The Disability Studies Reader*, pages 105–116. Psychology Press, 2006.

[252] Anastasia Schaadhardt, Yue Fu, Cory Gennari Pratt, and Wanda Pratt. "Laughing so I don't cry": How TikTok users employ humor and compassion to connect around psychiatric hospitalization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2023.

[253] Chris Schipper and Bo Brinkman. Caption Placement on an Augmented Reality Head Worn Device. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17. ACM, October 2017.

[254] Matthew Seita, Khaled Albusays, Sushant Kafle, Michael Stinson, and Matt Huenerfauth. 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, pages 68–80, Galway, Ireland, October 2018. Association for Computing Machinery.

[255] Matthew Seita, Sarah Andrew, and Matt Huenerfauth. 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*, pages 1–12, Ljubljana Slovenia, April 2021. ACM.

[256] Matthew Seita and Matt Huenerfauth. 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, pages 1–8, New York, NY, USA, April 2020. Association for Computing Machinery.

[257] Matthew Seita, Sooyeon Lee, Sarah Andrew, Kristen Shinohara, and Matt Huenerfauth. 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, pages 1–13, New York, NY, USA, April 2022. Association for Computing Machinery.

[258] Brent N. Shiver and Rosalee J. Wolfe. 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*, pages 231–238, Lisbon, Portugal, 2015. ACM Press.

[259] S. Silberman and O. Sacks. *NeuroTribes: The Legacy of Autism and the Future of Neurodiversity*. Penguin Publishing Group, 2015.

[260] Ellen Simpson, Samantha Dalal, and Bryan Semaan. "Hey, Can You Add Captions?": The Critical Infrastructuring Practices of Neurodiverse People on TikTok. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–27, April 2023.

[261] Ellen Simpson, Andrew Hamann, and Bryan Semaan. How to Tame "Your" Algorithm: LGBTQ+ Users' Domestication of TikTok. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):22:1–22:27, January 2022.

[262] Ellen Simpson and Bryan Semaan. For You, or For "You"'? Everyday LGBTQ+ Encounters with TikTok. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):252:1–252:34, Jan 2021.

[263] Lauren Southwick, Sharath C Guntuku, Elissa V Klinger, Emily Seltzer, Haley J McCalpin, and Raina M Merchant. Characterizing COVID-19 content posted to TikTok: public sentiment and response during the first phase of the COVID-19 pandemic. *Journal of Adolescent Health*, 69(2):234–241, 2021.

[264] Katta Spiel, Emeline Brulé, Christopher Frauenberger, Gilles Bailly, and Geraldine Fitzpatrick. Micro-Ethics for Participatory Design with Marginalised Children. In *Participatory Design Conference*, volume 1, Hasselt & Genk, Belgium, 2018.

[265] Katta Spiel, Christopher Frauenberger, Os Keyes, and Geraldine Fitzpatrick. Agency of autistic children in technology research—a critical literature review. *ACM Trans. Comput. Hum. Interact.*, 26(6):1–40, December 2019.

[266] Katta Spiel, Kathrin Gerling, Cynthia L. Bennett, Emeline Brulé, Rua M. Williams, Jennifer Rode, and Jennifer Mankoff. Nothing About Us Without Us: Investigating the Role of Critical Disability Studies in HCI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pages 1–8, New York, NY, USA, April 2020. Association for Computing Machinery.

[267] Katta Spiel, Eva Hornecker, Rua Mae Williams, and Judith Good. ADHD and Technology Research – Investigated by Neurodivergent Readers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–21, New York, NY, USA, April 2022. Association for Computing Machinery.

[268] Abigale Stangl, Nitin Verma, Kenneth R. Fleischmann, Meredith Ringel Morris, and Danna Gurari. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who Are Blind or Have Low Vision. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21, New York, NY, USA, 2021. Association for Computing Machinery.

[269] Jr. Stokoe, William C. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *The Journal of Deaf Studies and Deaf Education*, 10(1):3–37, January 2005. _eprint: https://academic.oup.com/jdsde/article-pdf/10/1/3/1034248/eni001.pdf.

[270] Allegra Stout and Ariel Schwartz. "It'll Grow Organically and Naturally": The Reciprocal Relationship between Student Groups and Disability Studies on College Campuses. *Disability Studies Quarterly*, 34(2), March 2014.

[271] Morpheyes Studio. Deaf identity — episode one of "deaf and...", Feb 2024. Available at `https://www.youtube.com/watch?v=nythXwJY61Y&ab_channel=MorpheyesStudioatRIT%7CNTID`.

[272] Cella M Sum, Rahaf Alharbi, Franchesca Spektor, Cynthia L Bennett, Christina N Harrington, Katta Spiel, and Rua Mae Williams. Dreaming Disability Justice in HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[273] Google Support. Use captions in a meeting - Computer - Google Meet Help, 2022. Available at https://support.google.com/meet/answer/9300310?hl=en&co=GENIE.Platform%3DDesktop.

[274] Microsoft Support. Use live captions in a Teams meeting, 2022. Available at https://support.microsoft.com/en-us/office/use-live-captions-in-a-teams-meeting-4be2d304-f675-4b57-8347-cbd000a21260.

[275] Zoom Support. Enabling Automated Captions, Apr 2024. Available at https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0058810.

[276] Agnieszka Szarkowska, Izabela Krejtz, Zuzanna Klyszejko, and Anna Wieczorek. Verbatim, Standard, or Edited?: Reading Patterns of Different Captioning Styles Among Deaf, Hard of Hearing, and Hearing Viewers. *American Annals of the Deaf*, 156(4):363–378, November 2011. Publisher: Gallaudet University Press.

[277] Hironobu Takagi, Takashi Itoh, and Kaoru Shinkawa. Evaluation of real-time captioning by machine recognition with human support. In *Proceedings of the 12th International Web for All Conference*, W4A '15. ACM, May 2015.

[278] John Tang. Understanding the Telework Experience of People with Disabilities. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):30:1–30:27, April 2021.

[279] Garreth W. Tigwell. Nuanced Perspectives Toward Disability Simulations from Digital Designers, Blind, Low Vision, and Color Blind People. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[280] TikTok. Creating Longer Videos, Jul 2021. Available at https://www.tiktok.com/creators/creator-portal/en-us/tiktok-content-strategy/creating-videos-longer-than-one-minute/.

[281] TikTok. Terms of service, Jul 2023. Available at https://www.tiktok.com/legal/page/us/terms-of-service/en.

[282] Alexandros Tsilfidis, Iosif Mporas, John Mourjopoulos, and Nikos Fakotakis. Automatic speech recognition performance in different room acoustic environments with and without dereverberation preprocessing. *Computer Speech & Language*, 27(1):380–395, January 2013.

[283] Jason Tu, Georgianna Lin, and Thad Starner. Conversational greeting detection using captioning on head worn displays versus smartphones. In *Proceedings of the 2020 International Symposium on Wearable Computers*, UbiComp/ISWC '20. ACM, September 2020.

[284] Jason Tu, Georgianna E Lin, and Thad Starner. Towards an Understanding of Real-time Captioning on Head-worn Displays. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20. ACM, October 2020.

[285] Stephanie Valencia, Amy Pavel, Jared Santa Maria, Seunga (Gloria) Yu, Jeffrey P. Bigham, and Henny Admoni. Conversational Agency in Augmentative and Alternative Communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–12, New York, NY, USA, April 2020. Association for Computing Machinery.

[286] John Vines, Gary Pritchard, Peter Wright, Patrick Olivier, and Katie Brittain. An age-old problem. *ACM Trans. Comput. Hum. Interact.*, 22(1):1–27, March 2015.

[287] J. Virdi. *Hearing Happiness: Deafness Cures in History*. Chicago Visions and Revisions. University of Chicago Press, 2020.

[288] Christian Vogler, Paula Tucker, and Norman Williams. 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, pages 1–5, New York, NY, USA, October 2013. Association for Computing Machinery.

[289] Quoc V. Vy and Deborah I. Fels. Using Placement and Name for Speaker Identification in Captioning. In Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, Lecture Notes in Computer Science, pages 247–254, Berlin, Heidelberg, 2010. Springer.

[290] Kaveh Waddell. Lost in Transcription: Auto-Captions Often Fall Short on Zoom, Facebook, Google Meet, and YouTube, 2022. Available at `https://www.consumerreports.org/disability-rights/auto-captions-often-fall-short-on-zoom-facebook-and-others-a9742392879/`.

[291] Jena Wallace. Captions vs. Subtitles: Breaking Down the Differences, Dec 2023. Available at `https://www.3playmedia.com/blog/closed-captioning-vs-subtitles/`.

[292] James M. Waller and Raja S. Kushalnagar. Evaluation of Automatic Caption Segmentation. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 331–332, New York, NY, USA, October 2016. Association for Computing Machinery.

[293] Emily Q. Wang and Anne Marie Piper. 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, November 2018.

[294] Yiwen Wang, Ziming Li, Pratheep Kumar Chelladurai, Wendy Dannels, Tae Oh, and Roshan L Peiris. Haptic-Captioning: Using Audio-Haptic Interfaces to Enhance Speaker Indication in Real-Time Captions for Deaf and Hard-of-Hearing Viewers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. ACM, April 2023.

[295] Susan Wendell. Unhealthy Disabled: Treating Chronic Illnesses as Disabilities. *Hypatia*, 16(4):17–33, 2001.

[296] Kevin Westcott, Jana Arbanas, Chris Arkenberg, Brooke Auxier, Jeff Loucks, and Kevin Downs. 2023 digital media trends Immersed and connected, May 2023. Available at `https://www2.deloitte.com/xe/en/insights/industry/technology/media-industry-trends-2023.html`.

[297] Rua Mae Williams, Louanne Boyd, and Juan E. Gilbert. Counterventions: a reparative reflection on interventionist HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[298] John S. Wrench, Narissra M. Punyanunt-Carter, and Katherine S. Thweatt. *Interpersonal Communication: A mindful approach to relationships*. Open SUNY, 2020.

[299] Kenta Yamamoto, Ippei Suzuki, Akihisa Shitara, and Yoichi Ochiai. 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, pages 1–4, New York, NY, USA, October 2021. Association for Computing Machinery.

[300] Anthony Yeung, Enoch Ng, and Elia Abi-Jaoude. TikTok and Attention-Deficit/Hyperactivity Disorder: A Cross-Sectional Study of Social Media Content Quality. *The Canadian Journal of Psychiatry*, page 07067437221082854, Feb 2022.

[301] Carmen Yip, Jie Mi Chong, sin yee kwek, Yong Wang, and Kotaro Hara. Visionary Caption: Improving the Accessibility of Presentation Slides Through Highlighting Visualization. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21. ACM, October 2021.

[302] Sean Zdenek. *Reading Sounds: Closed-Captioned Media and Popular Culture*. University of Chicago Press, December 2015. Google-Books-ID: lY6wCgAAQBAJ.

[303] Sean Zdenek. Designing Captions: Disruptive Experiments with Typography, Color, Icons, and Effects. *23.1*, August 2018. Publisher: Kairos: A Journal of Rhetoric, Technology, and Pedagogy.

[304] David X. Zheng, Anne Y. Ning, Melissa A. Levoska, Laura Xiang, Christina Wong, and Jeffrey F. Scott. Acne and social media: A cross-sectional study of content quality on TikTok. *Pediatric Dermatology*, 38(1):336–338, Jan 2021.

[305] Leyla Zhuhadar, Bryan Carson, Jerry Daday, and Olfa Nasraoui. A Universal Design Infrastructure for Multimodal Presentation of Materials in STEM Programs: Universal Design. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15. ACM, May 2015.

[306] Zoom. Update on live transcription for free accounts, Feb 2021. Available at `https://www.zoom.com/en/blog/update-on-live-transcription-for-free-accounts/`.

[307] Diana Zulli and David James Zulli. Extending the internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. *New Media & Society*, 24(8):1872–1890, 2020.

[308] Matthijs Zwinderman, Rinze Leenheer, Azadeh Shirzad, Nikolay Chupriyanov, Glenn Veugen, Biyong Zhang, and Panos Markopoulos. Using Video Prototypes for Evaluating Design Concepts with Users: A Comparison to Usability Testing. In Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, Lecture Notes in Computer Science, pages 774–781, Berlin, Heidelberg, 2013. Springer.