Bridging Language and Disability Lenses in the Design of Deaf and Hard-of-Hearing Communication Access Technologies

Aashaka Desai

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

Doctor of Philosophy

University of Washington

2026

Reading Committee:
Jennifer Mankoff, Chair
Richard E. Ladner, Chair
Jon E. Froehlich

Program Authorized to Offer Degree:
Computer Science and Engineering

University of Washington

Abstract

Bridging Language and Disability Lenses in the Design of Deaf and Hard-of-Hearing Communication Access Technologies

Aashaka Desai

Co-Chairs of the Supervisory Committee:
Professor Jennifer Mankoff
Paul G. Allen School of Computer Science and Engineering

Professor Emeritus Richard E. Ladner
Paul G. Allen School of Computer Science and Engineering

Deaf and hard-of-hearing (DHH) individuals communicate in many different ways – using sign, speech, writing, touch and more. In designing technologies for DHH individuals, most research takes a disability lens, exploring how technologies can support this wide range of communication access needs. However, engagement with language and language-related factors is limited. My dissertation argues that a language lens can be generative in designing for DHH communication accessibility, allowing us to better understand multilingual and multimodal realities and build technologies for both language access and communication access. To demonstrate this, I present three studies exploring DHH communication online: exploring speechreading in video-calls, captioning in multilingual media and conversations, and tech-mediated sign language interpreting experiences. Across this variety of communication contexts, I highlight how language and language-related factors influence use and efficacy of accessibility technologies. I additionally explore how we might design technologies to meet this diversity of language practices and access needs. To demonstrate how this approach can be technically generative, I turn to the development of sign language dictionaries, investigating methods for video-based search and vocabulary expansion – thereby guiding design and development in both access and language needs. Through empirical, design, and technical

contributions, my dissertation demonstrates the generative power of a language lens in DHH communication accessibility, and contrasts dual goals of language access and communication access guiding technology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Getting to the acknowledgements section of each paper and talk has always been my favorite part – finally, a chance to express the immense gratitude I feel. Research is far from a solo endeavor and becoming a researcher even more so – and there are so many people who have been critical to this journey.

Firstly, my advisors: Jennifer Mankoff and Richard Ladner. Thank you for showing me how to center access not just in my PhD but in every aspect of life. Your continued support and advocacy for disabled students and community has fundamentally shaped how I think research and impact. Your patience and support when all the hard life things came knocking taught me how to be compassionate with myself and others, and to remember we are all humans first and researchers second. As much as I learned from each of you alone, I learned from your interactions with each other: how to disagree, how to ask for advice, how to offer support. It is between those interactions I learned how to say what I wanted to say – thank you for helping me find my voice.

It takes a committee – Jon Froehlich, Alex Lu, Danielle Bragg, Amy Ko: Thank you for believing in me and offering a wealth of perspective through these years. Your unbridled enthusiasm for my work kept me going through the years, and I knew I could always turn to you when I couldn't see the forest through the trees. Thank you for showing me the many ways to be a researcher, and for teaching me how to pause and look at the big picture.

I feel so fortunate to have found more mentors both in and out of UW. Hal Daumé III, Megan Hofmann, Heather Evans – thank you for guiding me through instrumental parts of my PhD, and for teaching me so much about language and access and disability along the way. I am grateful to the many mentors in the accessibility research community who

welcomed me and supported me through conference chats and sporadic zoom calls: Cindy Bennett, Raja Kushalnagar, Christian Vogler, Katta Spiel, Hernissa Kacorri, Ed Cutrell and many more. To my undergraduate advisors: Debra Yarrington, Chien-Chung Shen, Giovanna Morini, thank you for seeing potential and getting me started on this path.

Each research project has been an ocean of adventure and growth. I am thankful for all my collaborators in dissertation research and beyond: Sanika Moharana, Christina Harrington, Daniela Masicetti, Shuxu Huffman, James Waller, Naomi Caselli, Annemarie Kocab, Maartje De Meulder, Julie Hochgesang, Saad Hassan, Oliver Alonzo, Lloyd May, to name a few. I am immensely grateful to participants in all studies for sharing their stories and insights with me, and for the many access providers without whom conducting this research would not have been possible. Thank you to Dimitri Summers for steadfastly coordinating access requests through the years, and for funding agencies for supporting this work.

I believe it is a mark of how foundational the UW community has been to my growth that I cannot imagine having done my PhD anywhere else. My lab mates at make4all and colleagues at CREATE and fellow d/Deaf and disabled friends – you showed me what collective access could look like in small everyday interactions. I have enjoyed stumbling through the theory and praxis of access with you – thank you all for languaging with me. To the many disabled activists and leaders who transformed my thinking, including the late Patty Berne, Alice Wong, and Jon Henner – thank you for showing me how to imagine better.

Of course, most important are the friends I made along the way: Jesse Martinez, Rahaf Alharbi, Stacy Hsueh, Aaleyah Lewis, Gina Clepper – thank you for making this journey feel a lot less lonely, and for bringing many joys (big and small!) to my life. To Avery Mack, for taking me under their wing since the very first day of my PhD and sticking till the very end – I am forever grateful for your friendship and mentorship. To Emma McDonnell, for listening to every half-formed thought about language I've had through the years – I am so thankful

I got to think with and learn from you. To every single friend who has ever coworked with me during the depths of a deadline, a pandemic, a weekend (including Daisy!) – you know this wouldn't have been possible without you.

To my family, I am so immensely grateful for your constant love and support – thank you for anchoring me and reminding me what is important. To my parents, Mitesh and Amishi Desai – your endless curiosity and desire to learn as you move through the world have molded me into the person I am today. To my sister, Shloka Desai, thank you for being my co-conspirator in most things in life, including this. From proofreading emails to soothing away my tears and panic, you have been ready to support me through it all. To Ishita Aunty, Ujjval Uncle, and Bhrugu, thank you for making Seattle feel like home.

As fortunate as I feel to have found such belonging here, I keep thinking about this line from Sarah Kay: *"This land, someone else's. This language, someone else's."* I would be remiss to talk so much about language in this dissertation and not recognize colonial histories that shape the language I am writing in, or that much of this work was conducted while living and working on the unceded ancestral lands of the Coast Salish people. I would like to express my gratitude to this land and all of the creatures on it. One of the things that has truly sustained me through this work is this city – the greenery, the endless water, the mountain and of course, the birds.

This list is far from comprehensive – to the many more friends and family and communities I am a part of: thank you, thank you, thank you.

Chapter 1

INTRODUCTION

Deaf and hard-of-hearing (DHH) communities represent a range of experiences with language and disability and thus communication accessibility. Many DHH individuals identify as disabled from living in a world that centers hearing and spoken ways of being. They use different access practices and technologies to navigate communication in daily life and provision access to sound and spoken language. At the same time, many DHH individuals identify as linguistic minorities through their use of sign languages [163, 212]. These sign languages connect them to their communities and culture, but have faced oppression throughout history [27, 162]. It was not until the 1960s that sign languages were recognized as full languages (and not just analogues of spoken languages) [205]. This history has impacted their presence in digital spaces, and many DHH signers are called to operate in a second language online. Despite being situated as both disabled and linguistic minorities, research and design of technologies for DHH communities often does not engage with both perspectives.

Current research designing technologies for DHH communication primarily takes a disability lens. Most works focus on examining how technology might meet sensorial needs (e.g., sound detection [129], speaker identification [12], captioning [110]) without further examining how language might influence access needs and experiences. When language is considered, research mostly focuses on translation i.e., ameliorating communication barriers between DHH individuals and hearing individuals by translating between signed and spoken language [253, 14]. While this is certainly an important task, it is not the only framework under which we can consider the impact of language needs on communication accessibility.

Consider the diversity of DHH individuals' experiences with language – some grow up bilingual in a sign and spoken language, some never learn how to sign, and some are multilingual (navigating multiple sign and spoken languages) [160]. This complexity of DHH individuals' language lives is an opportunity for us to explore the multitude of ways technologies can be designed for language access, and how language, more broadly, can inform communication accessibility technology.

Consider this series of examples: A Deaf signer using captions while watching a movie with their family. A late-deafened adult interacting with their signing colleagues with an interpreter. A hard-of-hearing person using speechreading while chatting in their second language with a vendor at a marketplace. For each of these individuals, it is not simply their access practice (captioning, interpreting, or speechreading) that is determining the quality of their communication experience – but language as well. Here I mean not only which languages are present (movie in Korean, conference in American Sign Language (ASL), marketplace in Hindi), but also many language-related factors. For example, consider how well they know these languages (e.g., can they read Korean captions? are they a novice signer? how familiar are they with different dialects of Hindi?). Consider the social relations and political histories of these languages (e.g., which other languages dominate the region/space? could they switch between languages? why is access to this language important to them?). Consider the digital footprint of these languages (e.g., do digital resources for learning this language exist? are translation technologies readily available?). These examples highlight that language needs frequently collide with access needs for DHH individuals – engaging with language and language-related factors then can help us design technologies that better match the needs of the DHH community in the real-world.

My research seeks to unveil the impact of language in communication accessibility and investigate how technologies might be designed to support a diversity of language and access practices. To do so, I examine DHH individuals' experiences with a range of access practices in online environments. In each of the scenarios listed above, individuals might combine multiple practices together to make communication accessible – for example, pointing and

gesturing at the marketplace, or speechreading the signer's mouth movements (or the interpreter's voicing), or discretely signing with a sibling to clarify meaning of a phrase in the movie dialogue. However, most prior research studies DHH access practices in homogeneous contexts (e.g., monolingual, monomodal scenarios) and rarely examines how multiple access practices might be combined to facilitate better communication access. I take a holistic approach to DHH communication instead, exploring multilingual and multimodal realities. In doing so, I hope to better understand how language impacts the use and efficacy of different access practices (e.g., speechreading, captioning, interpreting) and explore how these access practices and technologies can mediate language access and communication access. My dissertation is guided by the following research questions:

- How do language and language-related factors shape experiences of communication access and effectiveness of different access practices in DHH communication repertoire?

- How can communication access technologies be designed and built to meet both language needs and access needs of the DHH community?

Through the following chapters, I demonstrate my thesis statement: *Language and language-factors are critical modulators on how DHH individuals navigate and utilize the access practices in their communication repertoires. Considering these factors and designing DHH technologies for the dual goals of language access and communication access can better support the diversity of DHH communication.*

1.1 Research Approach and Contributions

My research features empirical, design, and technical contributions to account for the many ways language and disability collide in daily life. My approach is interdisciplinary, drawing on methods from both human-computer interaction and machine learning. In the following chapters, I describe formative research understanding communities' needs and experiences

with existing technologies, iteratively co-designing new technologies, and developing models to power new technologies.

In Chapter 2, I offer a background on DHH communication practices and then discuss the critical disability and language perspectives that inform my work. I briefly review relevant accessibility and language technology research – each of the following chapters offers a deeper dive into key related work.

In Chapter 3, I begin by exploring DHH individuals' experiences with speechreading online in video calls. This chapter features a three-part study, consisting of interviews, design probes, and co-design sessions with 12 individuals who speechread. The empirical accounts showcase the strengths and limitations of speechreading in mediating access to language and to communication – on one hand, the ambiguity of mouthshapes and cognitive load limit language access, but on the other hand, facial expressions and other paralinguistic cues foster emotional connection central to communication. We find that several linguistic, environmental, and sociocultural factors impact the choice between speechreading and captioning. In line with these findings and holistic approach, our proposed designs explore supporting speechreading alone as well as combined use of speechreading and captioning.

In Chapter 4, I explore captioning technologies – to dive into the intersection of language and disability, I focus on multilingual individuals' experiences with captioning for accessibility. I conduct interviews and diary logs with 13 multilingual and disabled caption users. This chapter showcases that language is fundamentally entwined with participants' experiences of access. We find that factors such as comfort with a language, linguistic affordances of a language, and cultural perceptions all impact what users desire from captions. To navigate disparities in captioning across languages and still create communication access, participants combine languages in unique ways (i.e., multilingual practices). This chapter concludes by suggesting immediate directions for development as well as guiding principles for future language-centered captioning technologies.

Having demonstrated that language is a critical component shaping DHH communication access, my next two chapters further explore how we might use both language and disability

perspectives to inform technology design. For this, I focus on emerging sign language technologies. In Chapter 5, I investigate development of video-based sign language dictionaries – specifically, examining how we might sustainably scale and update the vocabularies of these dictionaries. By focusing on both video-based search (multimodal access) and dictionary expansion (language growth and change), this work demonstrates combining language and disability goals can be technically generative. In Chapter 6, I explore automated interpreting technologies, studying how we might evaluate both the *quality of translation* and *quality of access* provisioned by these technologies. Through a multi-stakeholder study with 20 DHH signers, hearing interpreters, and Deaf interpreters, we decompose the different facets of an *accessible* interpretation – highlighting linguistic, cultural, temporal, contextual, and relational dimensions that guide translation choices. Based on this analysis, we outline implications for the design, development, and deployment of sign language interpreting technologies. We additionally propose improvements to remote interpreting technologies.

Overall, Chapter 3, 4 and 6 explore three different access practices from the DHH communication repertoire – speechreading, captioning, and signing respectively. The empirical accounts in each of these chapters provide evidence that language and language-related factors are critical determinants in how DHH individuals navigate and utilize their communication repertoires. Each chapter additionally showcases how these language-related factors can inform design and development of DHH technologies – for speechreading, multilingual captioning, and sign language interpreting. Chapter 5 takes primarily a technical approach, demonstrating that designing sign language dictionaries (a DHH technology) in accordance with both disability and language access values results in a technology that meets real-world needs as well as illuminates new directions for research inquiry.

Finally, in Chapter 7, I contrast findings from these different chapters to tease apart the different goals of language access and communication access guiding DHH technologies. In addition to discussing the impact of a language lens, I discuss the limitations of my dissertation and directions for future work.

1.2 Authorship and Positionality

The work conducted in this dissertation is the result of collaborations with several wonderful colleagues and mentors. While I led all of these projects, this work would not have been possible without the sustained engagement and contributions of my collaborators. Each chapter begins by listing co-authors – I use "we" pronouns to credit our collaborative work, and "I" pronouns when referring to my own perspective or connections to the dissertation as a whole.

My experiences with language and disability have greatly influenced my research questions and my approach. I grew up speaking Gujarati and English, along with exposure to many other spoken languages. I also began learning ASL during the start of my PhD. Each of these languages connect me to my culture and community, but my access to these languages has been strongly influenced by my evolving experiences with deafness and disability. Through the course of this work, I have used a multitude of access practices and technologies to communicate including automated captioning, CART, speechreading, and signing. To actively engage with how these experiences inform my research, I use reflexive thematic analysis through my qualitative work, which foregrounds researcher subjectivity in theme development. In each chapter, I additionally reflect on how these identities informed research methods (e.g., accessibility considerations for interviews) along with discussing collective authors' positionality.

While my lived experiences have often sparked my initial interest in different research projects, I have also been transformed by doing this work. Gathering empirical accounts from DHH participants helped me recognize the diversity of communication practices and ideologies in the DHH community. My interactions with Deaf peers and colleagues, disabled friends and mentors, and my very multilingual family have continued to inform my understanding of communication access and how I approach it in my research and day-to-day life. While I share some identities with my participants, there are also many ways in which I am positioned away from them – I am privileged in my access to education and spoken language

(particularly English). My perspective on disability and deafness have been shaped by my exposure to disability studies and disability justice through the course of my PhD, and drive me in my goals of designing technology to support and celebrate all kinds of languaging.

Chapter 2

RELATED WORK

2.1 Background on DHH Communities and their Communication Practices

Deaf and hard-of-hearing individuals (DHH) have a mix of receptive and expressive communication access needs, and use a range of communication practices – sign, speech, touch, writing, gesture – to meet these communication needs. Advances in technology and policy have increased communication options in today's world. For example, in the United States, the Americans with Disabilities Act requires entities 'communicate effectively with people who have communication disabilities' and provide 'auxiliary aids and services' – such as access to real-time captioning (CART: Communication Access Realtime Translation) or a qualified sign language interpreter [1]. In addition to these human access providers, DHH individuals are increasingly turning to technologies like speech-to-text and text-to-speech apps to navigate real-time communication [136]. Writing and speechreading also continue to be common access practices. Overall, DHH individuals' communication preferences vary based on their background and contextual needs.

This variety of communication practices is representative of the range of lived experiences in the DHH community. The community includes individuals who identify as culturally Deaf – a group of people bounded by use of shared sign language as well as shared values, traditions, and history [211]. A core belief of this community is that deafness is not a deficit, but rather part of the diversity of the world [161]. They value sign languages as cultural cornerstones, and approach communication as a collaborative with shared responsibility for access. Unfortunately, this perspective on deafness has not been shared by the hearing world – historically, many have aimed to "fix" deafness [162], believing that people who can speak and hear or act like they can speak and hear are superior [27]. This belief is known as

audism, and has resulted in perpetuating oralism, thereby oppressing sign languages and the people who use them [162].

This history has had a lasting impact on contemporary DHH communication practices as well as availability of sign language resources [158]. While audist and ableist beliefs still influence deaf education and access policies, Deaf activists and scholars have worked tirelessly to ensure deaf people's continued access to sign languages. A core part of this work is having sign languages recognized as full languages (not just analogues of spoken languages) and articulating the critical role they play in shaping Deaf people's ways of knowing and being (e.g., through Deaf gain [26]).

However, this has its limitations – Deaf people who are born to Deaf parents are only a small portion of the contemporary DHH community [196]. Many DHH individuals are born to hearing parents, or lose their hearing later in life, which means they may or may not learn sign language. They may have a variety of experiences with different spoken, written, and tactile[1] languages (e.g., [160, 170, 207]) and unique approaches to navigating their communication needs. These groups may identify with deafness differently, with some considering themselves to be a part of a linguistic minority (with sign languages) [163, 212], some considering themselves disabled, and some identifying with both (e.g., [46, 59, 23]), and some neither. Given this range of experiences, in my work, I engage with both critical language and critical disability perspectives. I use the term DHH, concurrent with participants' descriptions of their identities.

My work follows the trajectory of Deaf studies, toward descriptive and away from prescription language ideologies. While I seek to highlight the fluidity of deaf communication (i.e., how DHH individuals switch between languages and access practices as needed), I emphasize that this should not be misconstrued to undermine recognition of sign languages as real languages or deny DHH individuals' access to sign languages in any context they might choose them [66]. DHH individuals' fluid communication practices are often driven

[1]Such as Protactile, used by DeafBlind individuals.

by necessity, motivated by sensory asymmetries and power imbalances [66].

2.2 Theoretical and Critical Foundations

I draw on theories from both sociolinguistics and Deaf Studies to understand DHH individuals' experiences with accessibility, and technology. From sociolinguistics, I draw on the concept of repertoires. First introduced by John Gumperz [102], repertories are linguistic resources individuals' draw on while communicating – including words and sounds and pictures and gestures and signs [38, 48]. Unlike languages, which are externally defined, repertoires allow us to look internally to a person and understand how they leverage all the communication resources they have – and sometimes cross boundaries of defined languages [210]. This perspective puts the emphasis on languaging rather than languages (i.e., viewing language as a practice rather than an entity [217]. Given the wealth of communication and access practices available to DHH individuals in today's world, repertoires give us a meaningful lens by which to understand DHH individuals' experiences with language and technology. To better align with multimodality of deaf communication, I draw on Kusters' *et al.* [159] work, proposing *semiotic*[2] repertoires – thus no longer making distinctions between linguistic and non-linguistic resources.

In addition to academia, my research is also informed by community based and activist perspectives. In order to take both language and disability perspectives to DHH communication, I engage with *both* disability justice [34] and language justice principles [2]. Language justice centers the right for everyone to communicate, to understand, and to be understood in our languages and strives to create a world where no language dominates over the other [2]. As put by the Praxis Project: "Valuing language justice means recognizing the social and political dimensions of language and language access, while working to dismantle language barriers, equalize power dynamics, and build strong communities for social and racial justice" [78]. From disability justice, principles most relevant to my work are interdependence

[2]semiotic: relating to signs and symbols, gesture and body orientation

(everyone relies on one another) and intersectionality (each person has multiple identities, and each can be a site of privilege or oppression) [34]. I additionally draw inspiration from accessibility researchers who have described how these disability justice principles can inform our approach to accessibility technology research and design (e.g., interdependence framework [29]), as well as deaf studies researchers who explore diversity of DHH community [84] (moving beyond "homogeneous community of fluent signing white deaf people" [266]) and intersectionality (e.g., [57, 232]).

Disability justice and language justice are closely entwined, although they have not often been put in conversation with each other (some exceptions: [244, 165]). Both frameworks are aspirational, and intend not for a fixed goal but ongoing and sustained engagement towards a better world – "demand[ing] we move in mutual respect for all people regardless of whether or how they sign, speak, or otherwise convey what's on their mind or in their heart." (from "Language Justice is Disability Justice" [244]). Deaf spaces are an exciting place to engage with both these concepts. Particularly relevant is Henner and Robinson's work on a crip linguistics manifesto [113], which discusses how language use both disables and represents disability, and calls us to celebrate all ways of deaf and disabled languaging.

2.3 Technology and Communication Accessibility

2.3.1 Access technologies

Accessibility technology research is a growing field – a systematic review revealed that the number of accessibility works published per year has increased from 22 in 1994 to 95 in 2019 at ASSETS or CHI, and 11.3% of these works focused on DHH communities [177]. Research to support DHH individuals with technology varies widely, including technologies for captioning (e.g., [122]), sign language (e.g., [135]), speechreading (e.g., [92]), sound detection (e.g., [129]), sound visualization (e.g., [10]) to name a few. The works also explore a variety of contexts, including educational settings (e.g., [155]), social scenarios ((e.g., [215]), workplace (e.g., [251]) alongside in-person and online modes (e.g., [15]).

This body of research has shown how technology has the potential to both support and constraint DHH communication practices. For example, captioning technologies, powered by speech recognition, can offer spontaneous, accurate, and customizable access to speech [110]. However, the text format frequently misses nuances of emotion and tone, lacking paralinguistic information [63]. Similarly, sign language technologies (such as sign recognition models and avatars) might support greater access to information and communication resources [253, 221, 14] – but unfortunately, most digital platforms (video conferencing, social media) are not designed for 3D language and visual communication, and tend to constrain practices of DHH users [15, 176]. Having identified limitations, these works also seek to improve these technologies to better align with DHH users needs and communication practices – for example, for captions exploring how to improve caption readability [154, 31], reduce visual dispersion [128, 126, 215], improve richness [209, 265, 96], engage hearing stakeholders [190, 240, 274], and more. I review research related to specific access practices (e.g., speechreading, interpreting) in corresponding chapters.

While the breadth represented in this body of research is promising, there are notable disparities. The first of these is that not all modalities and languages have received equal attention: research on captioning has extensively explored everything from accuracy to display to augmentations; fewer works focus on sign language recognition, generation and translation; and rarely works focus on study of Protactile. A review of DHH papers in Mack *et al.* 's [177] literature review work shows that American Sign Language (ASL) and English (technologies) dominate CHI and ASSETS. Second, the different practices and modalities are studied in silos – almost no works look at DHH communication holistically. For example, while few works acknowledge speechreading practices of DHH individuals (e.g., [126, 154, 223, 256]), there has not been an in-depth exploration of the interaction between captioning and speechreading in environments where individuals have access to both. Overall, there is not much of an understanding of how different access practices are situated among each other in the DHH communication repertoire, and how we might design technology to support multiple practices.

2.3.2 Language Technologies

The world loses approximately 9 languages a year [242]. These language deaths are often guided by colonial violence, racism and ableism [152]. Technologies have played a mixed role in this endangerment and minoritization of languages spoken/signed by 94% of the world's population [187].

On one hand, technologies make it possible for dispersed communities to connect with each other (e.g., with communication technologies), as well as access information and cultural resources in their own languages (e.g., with broadcast media) [167]. There is also a rise in translation technologies that make it easier for users to communicate and engage with content across language differences such as Google Translate which supported 243 languages in 2024.

On the other hand, technologies have also reified disparities. Resources are not equitably available or maintained across languages (e.g., Wikipedia [230]) and platforms only support some languages – thus digitally disadvantaging users of these languages [293]. Even with English, there are disparities in performance for different kinds of English, (e.g., biases against African American Vernacular English [150] or 'accented' speech [220]). Some technologies actively erase non-normative language practices: e.g., Bender *et al's* work highlights how language models flatten variation and contribute to reifying linguistic hierarchies [28]. Schneider [236] similarly notes how technologies reproduce notions of language as bounded and disadvantage multilingualism and non-normative language practices. There is a deep need for language justice in technology.

While growing research attention to biases and performance disparities across these language technologies (e.g., [37, 195, 112, 167]), there has yet to be a study of sign languages through this language lens. Sign languages experience a digital disadvantage not only because they are used by a smaller part of the population, but also because they are visual languages – this necessitates rethinking paradigms for input and output and digital representation. Training sign language models requires advanced computer vision techniques and large-scale multimodal datasets, which are limited [41, 40]. Unlike written language corpora

which can easily be scraped, sign language datasets need to be thoughtfully curated and annotated [41, 40].

Even beyond signing, language technology disparities significantly impact DHH communication and research. While inequitable performance across languages are recognized, research has not traced how this impacts downstream accessibility tasks (e.g., captioning technologies). Most accessibility research has focused only on monolingual contexts (e.g., English conversations), not considering contexts with language switches. There needs to be a closer exploration of the DHH communication context through a language lens to understand how multiple marginalizations may materialize.

Chapter 3

UNDERSTANDING SPEECHREADING IN VIDEO CALLS

In this chapter, I dive into the access practice of speechreading, exploring how it is transformed in the online environment and ways technology might support speechreaders. This work has been previously published as "Understanding and Enhancing the Role of Speechreading in Online d/DHH Communication Accessibility" at CHI 2023, co-authored with Jennifer Mankoff and Richard Ladner [73].

In the context of the broader dissertation, this chapter presents my work situating speechreading as a part of the larger DHH communication repertoire – examining why DHH individuals prefer speechreading and in which contexts. By focusing on video calls, where automated captioning is easily available, this chapter contrasts two different parts of the communication repertoire (speechreading and captioning) – showcasing users' practices of switching between and using both in tandem.

Under the framework of language access, speechreading has a contentious role. Due to ambiguity of mouthshapes, speechreading does not provide full access to what is said – yet, historically, it has been promoted over sign language learning for DHH children (leaving them without full access to language during critical years). However, for late-deafened adults, speechreading *is* about maintaining access to language – the spoken language they feel the most comfortable with, that connects them to their families and cultures. Availability of other access practices (e.g., captioning) is not guaranteed across all languages. Speechreaders therefore present an interesting space to examine how language influences access choices in DHH communication.

3.1 Motivation

Any language used in communication has semantic (i.e., meaning of words), paralinguistic (i.e., pitch, volume, intonation) and non-verbal (i.e., facial expressions, body language) components. These components play a vital role in rich interaction – it is important to know not only *what* was said but *how* it was said. Bolinger aptly summarizes why: *"Much of the time [the aim of speech] is to cajole, persuade, entreat, excuse, cow, deceive, or merely to maintain contact . . . The importance [of what is said] can be underscored by the words we choose. . . or it can be underscored by the tone."* (Bolinger, 1986). However, with hearing loss, access to semantic and paralinguistic components of spoken language is reduced.

Sign language can convey paralinguistic nuance through the speed and flow of signs and accompanying facial expressions. However, of the estimated 48 million Americans with hearing loss, only 500,000 use sign language [171, 197]. Sign language use is similarly low in other countries due to the deleterious impacts of oralism and audist beliefs [27]. While hearing aids and cochlear implants attempt to improve access to semantic and paralinguistic components of speech, and captions provide semantic information, only speechreading focuses on *how* things are said. With speechreading, movements of lips, teeth and tongue are used along with nonverbal and contextual cues to "hear" what is being said visually [139]. Although speechreading offers this important context, its cognitive demands are high and its accuracy variable due to the inherent ambiguity of mouth shapes (for example, /p/ and /b/ look identical on the lips). Thus, speechreading is most effective in settings where communication includes a rich variety of contextual and non-verbal components. With the increasing importance of online communication, especially during the early months of the pandemic when few in-person conversations took place, captioning has risen in prominence. While a few commercial tools allow captions to be placed near a speaker's mouth, video conferencing offers limited support for speechreading, thus forcing d/DHH individuals to chose between speechreading and captioning, between semantic and non-verbal support.

Previous research has explored factors that impact speechreading ability [174, 239], visu-

alizations to support speechreading [93, 182, 185, 286], and approaches to improve captioning experiences[143, 110, 164, 273, 190, 31, 265, 96, 215, 154, 240, 274]. However, most prior work has looked at these technologies in isolation and fails to capture sociocultural and environmental factors at play. In addition, research into technologies that may improve the accuracy of speechreading is still nascent and has focused on individual words rather than continuous speech. With the growing popularity of video-calls and the different visual affordances of online environments, we need to understand how to create an accessible communication space online and support a variety of communication strategies including speechreading *and* captioning.

To understand the richness of d/DHH approaches to accessible communication, we present a three-part study, including semi-structured interviews, design probes and design sessions, with 12 d/DHH individuals who self-identify as speechreaders. Although our focus is on *speechreaders*, our study is about communication in general, not just speechreading: We take a broad approach, exploring lived experiences in formative interviews, grounding possible solutions in a design probe of mocked-up speechreading supports designed to encourage diverse ideation, and integrating participants' life experiences with concrete designs in design sessions. Our mixed methods approach allows us to study speechreading in the context of multiple communication technologies and to better understand its role in communication. Throughout our findings, we emphasize the complexity of communication and the value of providing information that goes beyond captioning in supporting d/DHH communication goals.

Our contributions offer a better understanding of the richness and variety of techniques d/DHH individuals use to provision access. They include:

- Empirical accounts of communication in online environments, including the variety of ways speechreaders interact with other d/DHH accessibility aids such as captioning; and the importance of interdependence in ensuring access.

- Study of design probes of *continuous* speechreading supports in video-calls, including

video mockups to support disambiguation, and static mockups to offer contextual cues, inspired by [94, 93]. While prior works have focused on efficacy and learnability of speechreading visualizations, our probe findings explore unwillingness to invest time learning such visualizations, and preferences for bottom-up vs. top-down speechreading supports.

- Holistic exploration of the design space of speechreading supports, including codesign sessions to enhance speechreading and address environmental and social barriers to access. The designs highlight creative ways to practice access such as offering speaker feedback, extracting keywords, and supporting tandem caption use in addition to reducing barriers to speechreading such as speaker variability and poor setup.

3.2 Background and Related Work

3.2.1 Background on Speechreading

Speechreading is a holistic approach to understanding communication that combines contextual information, bodily cues, and physical motion of the lips. A spoken language can be broken down to the individual sounds (i.e *phonemes*) and corresponding mouthshapes (i.e. *visemes*). When phonemes and visemes inform guesses during conversation, it is called bottom-up or analytical speechreading [139], sometimes colloquially referred to as lipreading. Speechreading also involves using contextual information, such us body language, location, linguistic information, and topic of conversation to guess what is being said. This is known as top-down or synthetic speechreading [139]. For example, it is easy to choose between a guess of "juice" and "shoes" at the shoe store. However, it is estimated only 40% of sounds in the English language can be seen on the lips [1]. Most research on speechreading focuses on acquisition of speechreading skills [95, 93] and factors that modulate speechreading ability [174].

[1] `https://www.cdc.gov/hearing-loss-children/treatment/how-people-with-hearing-loss-learn-language.html`

Figure 3.1: English Phonemes and their Corresponding Mouthshapes (Visemes)

Speechreading has a contentious history. In 1880, the resolution at the Second International Congress on Education of the Deaf [162] severely limited the use of sign language in schools for the deaf in many countries, where young deaf people first learned it as a natural language. They were expected to learn how to speechread, even though it was a cognitively demanding and inexact practice because of audism [27]. Speechreading was used to propagate oppression of sign languages and Deaf culture.

In no way do we condone oppression of a culture or forcing individuals to learn how to speechread. Instead, our work is rooted in supporting all the ways of being human and all the ways of listening. For those who did not have access to sign language learning, or who lose their hearing later in life and do not know sign language, speechreading is simply finding a new way to listen. It is a way to maintain touch with their communities and cultures – especially for those who are multilingual or native speakers of languages for which accessibility technology is limited (e.g., automated captioning for Indic languages). Understanding the needs and experiences of those who speechread, and how speechreading is integrated into their overall approach to communication, is thus important to improving design of accessible communication technology.

3.2.2 Communication Supports in Video-Calls

A growing body of research aims to support d/DHH individuals communication online, recognizing that most videoconferencing technology was designed with hearing norms in

mind [15]. Of these, few works have considered experiences of speechreading. For example, Kushalnagar and Vogler [157] offer guidelines for videoconferencing accessibility (e.g., ideal background and lighting), including supporting those who might speechread. Kozma-Spytek also explores the impact of frame rate and audio-video synchrony for lipreading [151].

Unlike with captioning and sign language, explorations of lived experiences of speechreading online are missing from literature. One notable exception is [127]. There is increasing work on the use of non-verbal and social cues over video call [97, 81, 234] but does not look at d/DHH individuals specifically. While [239] quantitatively studied preferences for hearing people's behaviors in reference to enunciation, speech rate and voice intensity, they not examine how d/DHH people navigate inaccessible behaviours. The interaction between captioning and speechreading has also not been explored in depth [126, 154, 223, 256]. Previous works discuss the visual dispersion [154] often experienced by d/DHH people, and eye tracking studies have examined eye movements while using captions in movies or prerecorded videos [223, 256]. How this plays out during video-calls with dynamic social interactions and delayed captions remains unstudied.

3.2.3 Approaches to Support Speechreading

Below we describe relevant research for improving speechreading accuracy.

Cued Speech

Cued speech aims to reduce viseme ambiguity in speechreading [60]. Sounds that have the same viseme (Figure 3.1) are assigned to different "groups". Group information is conveyed by the speaker to the listener using different handshapes and positions ("cues") around the mouth. Combining these standardized cues with a viseme makes it easier to recognize a sound. While synthetic speechreading alone achieves 50% accuracy, cued speech can improve this to 90% [206], and has been adapted to over 50 languages [61]. The timing of cues in relation to speech [20] is critical. While designed to be "synchronized", cues actually appear before the onset of the consonant. The anticipatory nature of the cue provides possibilities for

the phoneme that will be uttered, and the viseme helps narrow down to a specific phoneme. However, it requires training and a knowledgeable communication partner.

Digital Complements to Speechreading

Previous work has explored a variety of visualizations to enhance speech understanding. For example, lip visualizations [286, 185] can mitigate the effects of poor camera angle on speechreading. Other work [218] visualizes pitch and volume along with phonemic characteristics to enrich understand of intonation. There are also spectrogram visualizations, which show a spectrum of audio frequencies over time, and can be used to identify spoken words [298, 116, 275]. Similarly, mappings of phonemic and prosodic features to different textures and colors have been explored [277]. However, these approaches require expert knowledge [298], vary with each utterance and speaker [98], and are designed to be used alone and thus divide attention while speechreading.

Some notable works focus on viseme disambiguation without dividing attention while speechreading. Duchnowski *et al.* [76] created an autocueing system to support wider use of cued speech, which was further improved using hyper-realistic animation by Attina [19]. The system made it possible for cuers to speechread non-cueing speakers, but it left out a significant portion of the speechreading population who did not know cued speech. Massaro *et al.* [182] used three LED lights on a pair of glasses to uniquely identify phonemes. These lights corresponded to voicing, nasality, and friction of the spoken phoneme. However, user studies indicated a high learning curve and difficulty using the approach with continuous speech (when ambiguous phonemes are presented as part of a longer sentence or paragraph of speech). Gorman *et al.* [93, 92] took an intuitive approach to design with video overlays in PhonemeViz. Phonemes were placed in a semicircle around a speaker's face, and an arrow pointed to the phoneme corresponding to the beginning of a spoken word. It achieved 100% accuracy when used for individual words, and training time and cognitive load were minimal. However, the design did not translate well to continuous speech.

There are some notable omissions in this body of work. First, there is a lack of visual-

izations that have been shown to work with *continuous speech* rather than individual words. Second, these prior works attempt a range of goals from viseme disambiguation to intonation extraction, and vary significantly on intuitiveness and learning time. These designs are often evaluated by proxies or by speechreaders post hoc and thus miss the opportunity to set priorities for speechreading supports that align with speechreading practice in early stages of design (e.g., type of support, goals, learning time, intuitiveness).

3.2.4 Summary

While we are seeing a renaissance in the study of d/DHH communication technologies, which is beginning to appropriately incorporate a rich and holistic set of communication concerns as well as centering both DHH people *and* their communication partners, little of that work has included speechreading. To date, speechreading supports have primarily focused on improving accuracy and supporting learners. This work has not been translated to real-world settings with continuous speech, and there is a dearth of formative research that explores the role of speechreading and speechreading support technologies in *online* settings. To fill these gaps, we must articulate the lived experiences of d/DHH individuals who speechread, understand the affordances of an online context, and explore the design space of speechreading supports.

3.3 Design Probe and Study

This design probe and study contextualizes speechreading use and adds to the body of work on visualizations to complement speechreading via design probes (inspired by cued speed and PhonemeViz/ContextCueView [94, 93]) and design sessions. Specifically, we conducted a mixed methods study to understand the online speechreading experiences of d/DHH individuals; gathered their reactions to a speechreading support design probes for continuous speech in videos; and ideated new supports for speechreading during video-calls. In designing this study, we took inspiration from McDonnell *et al.* 's observation of the importance of social, environmental and technical factors in d/DHH communication [190].

Team Positionality The first author of this work, who conducted all interviews, identifies as hard-of-hearing, and is a frequent speechreader and a beginner in American Sign Language (ASL). Participants were aware of this identity, which may have impacted their contributions. For example, in the design session with the participants, the hard-of-hearing author actively participated in brainstorming, and used insights from her experiences. The second author has a non-hearing related disability, and is also studying ASL for use with family members and colleagues. The third author has lifelong experience interacting with speechreaders. These identities impacted the directions this research took — from the design probes to the dynamics in all interviews.

3.3.1 Overview of Procedure

The study process consisted of two semi-structured interviews conducted approximately a week apart with a design probe presented between the interviews. Interviews were conducted remotely with a single participant, conducted by one researcher, and supported by Communication Access Realtime Translation (CART) captioners. All interviews were audio and video-recorded. Captions and automated transcripts were saved for analysis.

First Interview: During the first interview, we collected demographic information and discussion focused on formal speechreading training, reliance on audio *vs* visual senses for communication, experiences speechreading online, and participants' vision of "rich" interaction. For example, participants were asked to recall instances where speechreading was really easy or difficult during a video call, and reflect on what aspects of the experience made it easy or difficult to speechread, and whether they found speechreading online different from speechreading in-person.

Design Probe: Next, participants engaged with an asynchronous design probe that would help them to reconsider how digital video systems could support speechreading beyond typical captioning approaches. Our goal was to explore cued-speech-like supports for speechreading in continuous speech contexts. Thus, we focused on presenting a variety of cued-speech-like designs (six total), rather than comparing cued speech to other communi-

cation options. To our knowledge, no prior speechreading support system has been tested with speechreaders using full sentences (as opposed to individual words). Thus, a second goal of the design probe was to give speechreaders an opportunity to provide feedback about the potential value of such an approach.

These designs were presented using pre-recorded videos augmented with information about vowels and consonants, and reactions were collected using a Google form. We conducted this probe asynchronously using a form to allow participants the freedom to navigate between designs and pace themselves. Details of the designs and corresponding disambiguation task are described in section 3.3.2.

Final Interview: The third session brought participants back into an interview setting to brainstorm their vision of online speechreading supports. Participants met 1:1 with the interviewer again to share their thoughts on the speechreading designs and to brainstorm new approaches to supporting speechreading in video-calls. The interviewer brought up common problems the interviewee previously mentioned to focus discussion around how technology could help. Additionally, participants were asked about designs that previous participants came up with.

3.3.2 Design Probes

We created two types of design probes: six bottom-up and three top-down speechreading supports. Below we describe each, as well as the details of our asynchronous data collection process.

Video Mockups of Digitally Cued Speech Supporting Bottom-Up Speechreading

We developed video mockups of digitally cued speech to simulate what a speechreading support system might do in a videocall, inspired by [92]. While traditional cued speech uses handshape and placement near the mouth to encode phoneme information, our approach used abstract symbols, color, and position, as illustrated in Fig. 3.2. Like PhonemeViz, our visualization shows these as overlays on videos around the speaker's head [92], which is

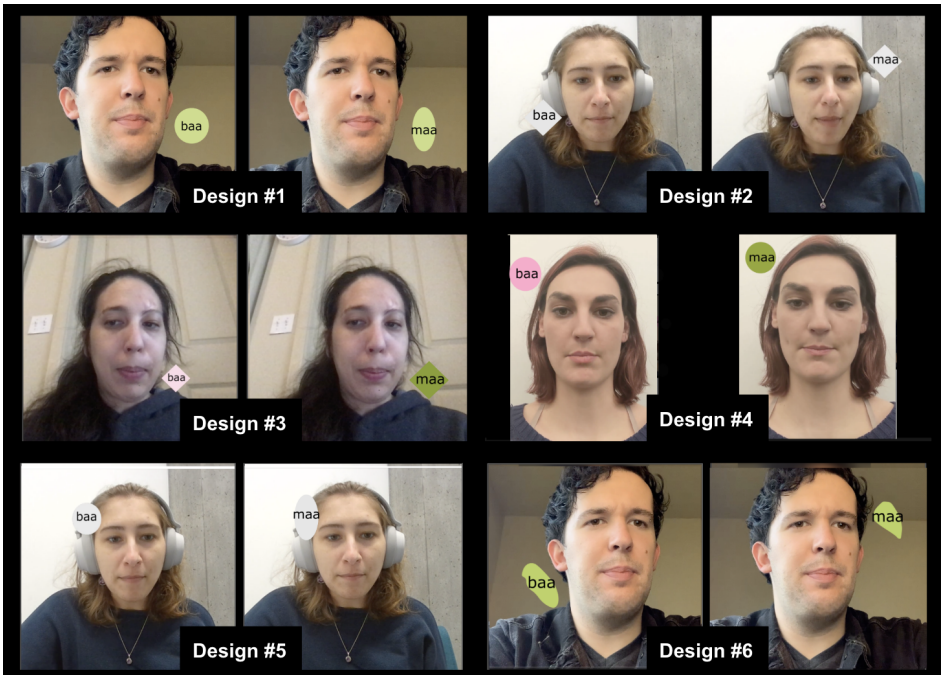


Figure 3.2: Screenshots of Bottom-up designs applied to syllable 'baa' vs. 'maa'. Note how mouthshape is ambiguous.: **Design #1** Consonant Shape, Vowel Color; **Design #2** Consonant Position, Vowel Shape; **Design #3** Consonant Color, Vowel Shape; **Design #4** Consonant Position, Vowel Color; **Design #5** Consonant Shape, Vowel Position; **Design #6** Consonant Color, Vowel Position.

better suited than traditional cued speech due to the limited 2D space in video-calls. We use the idea of "grouping" phonemes from cued speech, where phonemes that are visually ambiguous are assigned to different groups. For example, /p/ is assigned to group 1, /b/ is assigned to group 2 and /m/ is assigned to group 3. Group information is conveyed using the design dimensions described above. When group information is combined with the viseme (mouthshape), it is enough to uniquely identify the phoneme. All phonemes are visualized as they are inherently ambiguous; at any given time, the system shows one consonant and one vowel.

To use such a system, the speechreader must learn to quickly recognize meaning from

color, shape, and position, eventually with peripheral vision. This is supported by assigning each consonant and vowel group a unique color, shape, and/or position. For example, in Design 1, as defined in Figure 3.2, consonants group is indicated using shape and vowel group using color (position is defaulted at bottom right). If a speaker says "*bath*", the first syllable, `/b/aa`, is cued using light green (as the vowel is `aa`, as specified in Table 3.2) circle (as the consonant is in the `b n wh` group, as specified in Table 3.1). If they say "*math*" instead, a light green (`aa`) oval (**m t f** group) is shown. In contrast, Design 2 uses position for consonant group, and shape for vowel group (Fig. 3.2, top right) so both images show a light grey (default color) diamond (`aa`) that moves from bottom left (for the `b n wh` consonant group) to top right (for the **m t f** consonant group), as specified in Table 3.1. When group information is combined with the viseme (mouthshape), sufficient information is present to uniquely identify the phoneme.

Colors, shapes, and positions were chosen to maximize differentiability. The heterogeneity of grouped sounds required the mapping between phoneme groups and these dimensions be arbitrary, making it potentially harder to memorize. It is possible that some mappings may be more memorable than others, but we felt that the differences were unlikely to be salient for beginning visualization users. Instead, to make recognition easier for first-time users in our study, our visualization shows phonetic text (the consonant and vowel currently being spoken).

We created final versions of each design as followed: We videotaped four speakers each speaking all eight of the following sentences. Each sentence pair below has one word that is ambiguous visually (when speechreading) and semantically (the correct word cannot be guessed from the surrounding words). We embedded the ambiguous phoneme in a sentence to offer a sense of a continuous speech use case and highlight the advantages of bottom-up speechreading support:

1. a There is a mat in the house.

   b There is a bat in the house.

Consonant Group	Shape	Color	Position
d p zh	Pentagon	Sage Green	Top Left
r h s	Triangle	Red	Mid-top Left
b n wh	Circle	Light Green	Mid-bottom Left
th k v z	Star	Turquoise	Bottom Left
m t f	Oval	Yellow	Top Right
y ch ng	Rectangle	Brown	Mid-top Right
w sh l	Square	Cobalt Blue	Mid-bottom Right
g j TH	Heart	Lavender	Bottom Right

Table 3.1: Group representations for Consonants. Colors, shapes, and positions are chosen to maximize differentiability, but mapping between groups and dimensions is arbitrary.

Vowel Group	Example	Shape	Color	Position
a	a	Triangle	Red	Top Left
aa	father	Diamond	Light Green	Top Left
o	no	Circle	Cobalt Blue	Mid Left
oy	noise	Oval	Orange	Mid Left
e	there	Pentagon	Yellow	Top Right
ei	bait	Clover	Pink	Top Right
u	full	Star	Brown	Mid Right
aw	down	Heart	Turquoise	Mid Right
i	is, mini	Square	Sage Green	Bottom Right
ai	buy	Rectangle	Lavender	Bottom Right

Table 3.2: Group representations for Vowels. Colors, shapes, and positions are chosen to maximize differentiability, but mapping between groups and dimensions is arbitrary.

2. a The night is beautiful.

   b The light is beautiful.

3. a The fan is making strange noises.

   b The van is making strange noises.

4. a There is a dent on the miniature hill.

   b There is a tent on the miniature hill.

We hand annotated these videos to illustrate the different designs. We made four videos using Design #1, corresponding to the sentences 1a, 2b, 3b, and 4b. We made four videos using Design #2, selecting different variations of the four sentence pairs, and so on for each design. Each of the four videos showed only one variation of each ambiguous sentence and each one featured a different speaker, as seen in Table 3.3.

#	Design	1	Who	2	Who	3	Who	4	Who
#1	Shape Color	1a: Mat	S4	2b: Light	S2	3b: Van	S3	4a: Dent	S1
#2	Position Shape	1a: Mat	S1	2a: Night	S3	3b: Van	S2	4b: Tent	S4
#3	Color Shape	1b: Bat	S3	2b: Light	S4	3a: Fan	S1	4a: Dent	S2
#4	Position Color	1b: Bat	S4	2a: Night	S1	3a: Fan	S2	4b: Tent	S3
#5	Shape Position	1b: Bat	S1	2a: Night	S2	3b: Van	S3	4a: Dent	S4
#6	Color Position	1a: Mat	S2	2b: Light	S3	3a: Fan	S4	4b: Tent	S1

Table 3.3: Ordering of speakers and answers by design

Static Mockups of Contextual Information supporting Top-Down Speechreading

Knowing that our participants might have different styles of speechreading (bottom-up *vs* top-down) and to support diverse design brainstorming later, we also created three sketch-only mockups of top-down speechreading supports. Our top-down speechreading probes are

shown in Figure 3.3 and are similar to ContextCueView proposed in [94, 93]. Our goal was to explore the potential value of keyword information, time information, and transition indicators, and to seed participants with a broad set of ideas going into the design ideation session during our final interview.

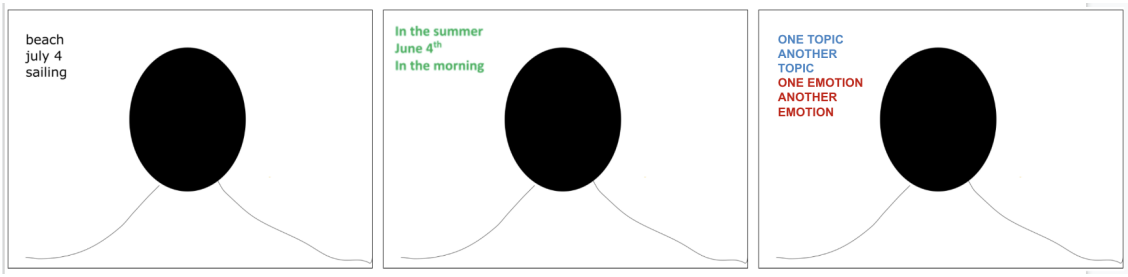


Figure 3.3: Mock-up Stills of Top-down Probes: (left) keyword information, (middle) temporal information, (right) transition indicators.

Design Probe Task

Participants engaged with the probe through a google form. For each bottom-up design, we explained the design through sketches. We chose not to train participants for two reasons: First, participants would have had to relearn things for each design (which would cause more confusion). Second, we wanted to avoid burdening participants with learning the approach before we had their feedback on its value in the broader communication context.

After familiarizing themselves with a design, participants disambiguated the above-mentioned visually ambiguous sentences using the design: they viewed each video mockup then specified which sentence in a sentence pair that design corresponded to. To reduce the need for training, we added phonetic text. After viewing all four videos for a design, participants were asked to respond to a Likert scale statement about the design's usefulness in disambiguating utterances and its distraction from speechreading. We also asked participants to respond to an open-ended query about their initial impressions and critiques of the design. All participants saw the same version of each video for each design. After viewing all six designs,

participants were asked about their overall impressions of phoneme-level annotations and the various design dimensions (shape, color, position) used to encode phoneme information in our designs. Lastly, participants viewed each of the top-down speechreading support images in Figure 3.3 and shared their impressions of these contextual cues and other alternatives to speechreading support.

3.3.3 Accessibility Considerations

The study was designed with accessibility to both participants and researchers in mind [178]. We offered CART for each interview to accommodate both the participant and researcher. However, some participants (P6, P8) preferred to use automated captioning to minimize lag in conversation, which we addressed using Google Meet. In situations where automated captioning failed, the interviewer used chat to correct captions. We had one instance (P5, Session 1) where the CART captioner suddenly canceled. In that case, the participant approved proceeding with automated captioning only. Occasionally, the interviewer would correct captions using ASL/fingerspelling when the participant knew ASL. For CART captions in Zoom, participants were provided with a StreamText link. Realizing some of our participants were not familiar with StreamText, we opted to also stream captions through Zoom.

3.3.4 Analysis Methods

We analyzed the transcripts using reflexive thematic analysis as described by Braun and Clarke [43]. Transcripts of interviews were read and re-read for immersion, important quotes were highlighted and initial ideas were noted. Following this, codes were generated using a mix of inductive and deductive approaches. The reflexive thematic approach foregrounds researcher subjectivity in coding and theme development process. However, to reflect on and deepen our process, the coding was shared with other authors and codes were discussed between authors for prioritization and grouping into themes.

3.3.5 Participants

We recruited 12 participants through several mailing lists of prominent disability and hearing loss specific centers and through snowball sampling. We required that participants identify as d/Deaf or Hard-of-Hearing, speechread frequently, and be fluent in English. We assessed participants' baseline speechreading ability using a recorded sentence without annotations. All participants were correct in choosing the corresponding sentence from options, making us confident in their speechreading ability.

Participants included five men and seven women with a mean age of 53 (SD=20). Table 3.4 lists demographics in more detail. Many of our participants had gradual or fluctuating hearing loss, which progressed variably through the years. Several participants learned to speechread through time and experience. P6 and P8 took speechreading classes early in their hearing loss journey, but attribute most of their learning to real-life practice. P1 attended speech training until he was 7 years old, which partly included speechreading, and P2 taught a speechreading class several years ago. P10 took speechreading classes at Gallaudet University, and spent time learning how to interpret body language. None of the participants were cued speech users, although some were familiar with the concept, and P4 and P8 had attempted it briefly. All participants frequently engaged in online speechreading, since the study took place during the shift to increased online communication due to the COVID-19 pandemic.

Participants used a range of accessibility technologies (AT): audio (e.g., hearing aids or cochlear implants), speechreading, automated captions, CART with professional captioners and ASL. Choice of AT was influenced by availability, efficacy and social factors. For example, using ASL requires adoption by people in your community.

There are also large societal events, such as the COVID-19 pandemic and subsequent use of masks and videoconferencing, which significantly influenced technology use. Finally, our participants had diverse language backgrounds including varying fluency in ASL, French, Hindi, Japanese, Russian, Spanish, German and Chinese.

Participant	Identity	Age	Onset
P1	Deaf	24	Birth
P2	Hard-of-Hearing	75	Birth
P3	Hard-of-Hearing	72	57 years-old, gradual
P4	Hard-of-Hearing	26	Birth
P5	Hard-of-Hearing	23	9 years-old, gradual
P6	Hard-of-Hearing	73	18 years-old, gradual
P7	Hard-of-Hearing	70	64 years-old, sudden
P8	Hard-of-Hearing	62	30s, gradual
P9	Hard-of-Hearing	39	Birth
P10	Hard-of-Hearing	50	5 years-old
P11	Hard-of-Hearing	72	Early 20s
P12	Hearing Impaired	53	26 years-old, gradual

Table 3.4: Participant Demographics

3.4 Findings

Our primary themes are derived from the first interview and parts of the design session: 1) factors that impact speechreading online such as the interactions between speechreading and captioning; 2) reorientation strategies in inaccessible situations; 3) social and cultural factors impacting access provision. The second two phases of the study provide insight into potential technology solutions for speechreading. These are reported in the final two subsections of our findings, where we share reactions to our design probes and ideas from our design sessions, highlighting design considerations for future research.

Participants' speechreading experience was impacted by a variety of factors, most of which has been noted in prior literature [174, 93]. For example, poor lighting and obstructions (e.g., hand, objects, facial hair) make access to visual cues for speechreading difficult. Being amidst a pandemic, all participants remarked on difficulty posed by masking and how it made them realize how much they speechread. Body language, expressions and eye gaze are crucial, and knowledge of context and topic lays the foundation for guesses in speechreading. P11 remarked that in knowing the topic, *"I find myself going faster than the speaker and guess where they are going."* (P11). All participants emphasized the variability of lip movements and body language across speakers. While familiarity with a person's speech pattern, such as knowledge of accent and word choice helps, *"[S]ome people I can lipread 100% basically. Other people I can do zero. And most people of course are in between."* (P2). Finally, multiple participants highlighted the impact of language and accents on speechreading. Each language has its own set of visemes and phonemes that are ambiguous. There are nuances in pronunciation, such as aspiration (Hindi) or inflection (Chinese) that have no visual counterparts, thus making speechreading difficult. Fluency in language and dialect affects speechreading, as it comes with a large vocabulary and knowledge of pronunciation to support guessing. For novices, it is hard to get sense of pronunciation just from writing. For those who are multilingual, there is a different problem that originates from code-switching: As P4 commented, *"I would need to know whether the person is actually speaking in that*

language or not." (P4)

3.4.1 Speechreading Online

While in-person and online speechreading share some affordances, there are significant differences in how mentioned factors play out online. For example, good quality acoustics can complement visual information, offering *"all the cues that I needed in order to read the lips."* (P12). In an online context, this translates to having a good quality microphone, not moving away from it and minimizing background noise. Similarly, one participant remarked on minimizing visual noise:

> *"I have a really busy background . . . and that's terrible. . . because it's distracting. . . I like it when they have like a very neutral background because then I am not distracted looking at other things instead of trying to read what they are saying."* (P12)

Body language also plays out differently online. *"Reading somebody's lips is about the same but it is more difficult to understand somebody's body language."* (P9). The video set up obscures posture and movement of hands and feet that are often key to speechreading. For example, any fidgeting and shaking, such as tapping feet might reveal the speaker is nervous or eager to get away. One participant also mentioned how video-calls can impact speaker's comfort and ease:

> *"People tend to be more rigid on a video call rather than in-person. . . because you lose that human to human interaction."* (P1)

In addition to changed speech pattern, lack of eye contact makes it harder for listeners to give off non-verbal feedback cues to get the speaker to slow down or repeat themselves. Additionally, there is lack of control over a speaker's video setup, or the ability to reposition yourself to get the best angle. Some participants thought that made speechreading in 2D

inherently different from 3D. Also, poor Internet connection can disconnect audio and video streams, making it harder to speechread.

However, video-calls have pros as well. First, there is often a close-up view of the speaker's face from a good angle. The frontal view of all participants and speaker identification makes group meetings easier to navigate. Second, video-calls can also circumvent social norms that are in conflict with access needs. As P6 explains,

> *"Well ironically you can get closer to the person on a video call than you can in-person because social protocols, you're not going to be in their face."* (P6)

Third, video-calls offer control, as described below:

> *"It is much more difficult to control the environment in in-person meetings. The room may be big and there may be other outside noises... Here at my home I can make sure it is completely quiet..."* (P10)

Listeners can adjust the size of speaker window, volume and remove background noise. Also, norms and settings prevent parallel conversations and overlapping speakers, making online more of a *"controlled"* (P9) environment.

Perhaps the most significant difference between online and in-person interactions is easy access to automatic captioning. In online interactions in some platforms, it is often easier to use captions without identifying oneself or interrupting the conversation. Participants varied on their feelings about captioning versus speechreading in video-calls—some preferred only one, or the other, some used both.

For those who were preferred speechreading to captions, they found it to be more accurate than automated captions. While we cannot evaluate compared accuracy, these participants *"trust their speechreading more than captioning,"* (P9) and this might be because speechreading is the default for in-person conversations. Additionally, a key factor in this preference is the ability to perceive the speaker's facial expressions, body language and other nonverbal

cues that are missing from captioning. These are important to get a sense of the speaker's tone and interpret the utterance well, as shared below:

> *"You get emotion. You get those hidden cues and human communication. You can tell maybe someone is being sarcastic, maybe someone is joking and whereas if you are just reading words, you lose all of that information."* (P1)

Captions only offer access to semantic components of language, (*"objective in transcription"* (P1)), so with speechreading, participants get a better sense of the speaker's accent and background. Access techniques are often evaluated by their ability to accommodate impairment, but in articulating speechreading experiences, participants shared the connections and communication fostered through speechreading:

> *"It also sort of helps you build a deeper connection and a bond with the other person because you know exactly how they talk and it makes them more real than just reading it from a piece of... two-dimensional text."* (P4)

As speechreading is synchronous, listeners can perceive conversation in real-time and leverage the audio stream with their residual hearing which is hard to do with delayed captions. Despite recent technological advances, automatic captioning often fails with highly technical vocabulary or unfamiliar words, like proper names. One participant, P6, has a name that automated captioning struggles with and becomes frustrated at being unable to recognize when she is called on by name. Due to the diverse language background of our participants, we also found language of the conversation factored into the choice of speechreading. Captioning is not available in all languages and requires literacy.

For those who preferred captioning, it was largely due to the innate ambiguity in speechreading. Even with an ideal setup and great skill, there are multiple possibilities for a given utterance. Speechreaders prune these choices using context and keywords, but if they get it wrong, the only way to know is by responding incorrectly. In group conversations online,

captions are preferred because they include speaker identity. Additionally, speechreading is very dependent on speaker's speech patterns and set up. Captions make speaker variability and speed non-issues. Most importantly, captions are low effort compared to the cognitive demand of speechreading. One participant expressed,

> *"I would probably go for the captioning. If it's good captioning, then it would probably calm me down."* (P6)

Participants who use captions and speechreading in tandem found captions to be an additional information stream for speechreading, offering context. They described dropping their gaze to read captions whenever they miss a word or a phrase. By offering feedback and corrections for speechreading, captions inform participants if they misinterpreted the speaker. With the pros of speechreading and captioning mentioned above, using both together allows participants to maximize access: *"I use both tools to try to get as much information as possible."* (P9)

However, this simultaneous use is largely dependent on the delay of captions and their positioning. If captions are too far behind, switching between two streams can be disorienting. Bouncing gaze between speaker's face and captions takes practice and divides attention, especially when the captions are further away from the speaker's mouth.

3.4.2 Reorientation Strategies

We asked participants how they deal with inaccessible interactions, and what strategies they use to reorient themselves when they miss something in a conversation. Often, participants' approach was simple: asking their conversation partner to repeat themselves as many times as it takes to understand them; and if they are obstructing their face, ask them to remove it.

Several participants remarked that they had to learn to be assertive, and request conversation partners follow accessible practices or repeat themselves. They needed to keep reminding people because they forget. Some were hesitant to keep asking because it got

"tiring and frustrating" (P11), while others said *"it is a no-win situation if you don't"* (P3). A few participants made sure they really needed a repeat by letting a conversation run-on for context and only then asking if they felt lost:

> *"Before asking for any repeat or anything, I wait until I really don't know. I also let the person keep speaking thinking that if they continue it will give me a clue on what I missed."* (P6)

When the conversation is sensitive, P9 found it hard to ask for repeats because she knew how hard it was for the speaker to share.

Additionally, participants had access strategies. Zoom requires meeting hosts to configure captions before the meeting begins. Many participants experienced hosts who did not know that captioning was available or how to enable it. P3 asked to be host for meetings to avoid any such confusion, and P4 used another captioning app in the background. In situations where captioning lagged behind, P2 would interrupt "to let the captions catch up".

To reduce inaccessible interactions, some participants set communication or access norms from the start. This would be *"a short lesson in what works for me"* (P3), such as avoiding obstructing facial features, speaking slowly and enunciating. One participant requested content previews for meetings to support speechreading, and asked speakers to keep checking in with him during the meeting. While this took a bit of effort, he said *"I make sure my time is valued and so is theirs"* (P4). Another participant, P3, who is a teacher, "looped" her classroom and had students speak into a microphone. This improved audio quality, but also acted a prompt for students to speak slower and enunciate.

The people involved in the conversation also impacted reorientation strategies used. As a member of Hearing Loss Association of America (HLAA) chapter, P3 found that setting access norms easier with group of people with hearing loss as they are all empathetic. Participants also had allies who helped negotiate or provision access, as seen below:

> *"My husband, I could lip read him a hundred percent with no sound.... If we were out socially, and I would give this look, I make eye contact to let him know that*

> *I wasn't following, and he would tell me what was being said . . . without sound."* (P6)

By mouthing missed parts, P6's husband acted as an ally to provision access. Allies might be family members, spouses or friends. P5 would check in a trusted friend in group conversations to reorient herself if needed. Other forms of allyship noted were friends who shaved mustaches and beards to make lipreading easier. P2 also has a spouse with hearing loss and they have evolved their communication practices over time to be accessible to both of them. For example, they always capture each other's attention before they start speaking and consider lighting.

3.4.3 Sociocultural Factors

In asking participants to discuss above reorientation strategies, they also shared important sociocultural factors that impacted access provision. P2 commented on division of access labor:

> *"Communication is a two way street. . . . They think I'm supposed to try harder and I'm already giving 100%, 150% – they've got to do their share."* (P2)

Cooperation plays a key role in access negotiations, because conversation partners need to follow through on any request to speak up, repeat, remove obstructions, reposition themselves, or write out words. But speakers could always choose not to provide this support. Some people react badly when asked to clarify: *"Sometimes I'll ask. It depends on if they take your head off or not – you know?"* (P8). Some might renege on access norms as they forget over time, and others refuse to educate themselves even knowing they are working with the d/DHH community.

Reorientation strategies are complicated by dynamics and context as well, as described below:

> *"So yeah, these – every time I interact with somebody, there are a thousand different dynamics that go into how I am going to interact with that person. It's very very exhausting. It is time consuming... and you have to choose your battles."* (P9)

To elaborate, participants would perhaps speak up for a repeat or access adjustment in a small group, but would hesitate in a larger group or situation where they are a fly-on-the-wall participant. Deciding whether to set access norms with a person depends on length of future relationship, and accounting for different dynamics. *"I will invest time into telling you how to best interact with me if it will be long lasting relationship"* (P9) but not for a one-time interaction. It depends on how critical the conversation is (e.g., business *vs* transactional) and the authority or role they have in the situation. Along with hearing loss, other sociocultural factors (such as race or gender) also affect power roles.

Hearing loss largely is invisible. This invisibility plays out in different ways, as shared by two participants:

> *"So when I meet a new person for the first time, I sort of give them a lesson of what works for me. And often times because you can't see my implant or my hearing aids, they will accommodate me for a few minutes, and then forget."* (P3)

> *"When you are in-person with someone, they would get a little worried for their personal space if you were you know, real close to them. I am just watching your mouth [and they would think] "What are you looking for, something in my teeth? Or what's the deal?""* (P7)

On one hand, conversation partners might forget about participants' hearing loss and accessibility practices. Visible cochlear implants can act as sign of their d/DHH identity and as a reminder of access practices. On the other hand, this invisibility means disclosing hearing loss is a choice. Access practices and technology play an interesting role in maintaining invisibility. For example, gaze on lips while speechreading might give you away as shared by P7.

Hesitance with disclosure partly stems from stigma and misconceptions around disability. P9 had coworkers who would yell while speaking to her after she disclosed her hearing loss, and it was counterproductive as yelling is not easy to understand.

Additionally, some participants found the visual attention demanded from *both* reading captions and speechreading conflicts with 'social norms' of looking at the speaker. For example:

> *"I was just taught that you look someone in the eye when you talk to them, and so when I'm – I literally do feel very disrespectful when I am looking at their mouth. It's just the way I was raised, so it's breaking some of those barriers even in my mind, that it's okay to look at someone's mouth."* (P12)

Advocacy is a crucial aspect of participants' lives, and how much they advocated for their needs changed over time. Some used to bluff their way through a conversation, but now prefer spending time with those who are willing to accommodate their needs. P2 had hearing loss from birth, and in school he knew he did not like teachers standing near windows because it was hard to understand them. But it was difficult to articulate the problem: being back-lit. P11 shared the impact of finding a community of others with hearing loss:

> *"I went from not advocating for myself to now advocating for myself and everybody else who have hearing loss. I've gotten much better."* (P11)

Advocating for access needs can help others who are not comfortable doing so in the same situation. A few participants are also involved it larger scale advocacy efforts such as free Zoom captioning, open captioning at movies and transit accessibility. In both individual interactions and community service, they tried to increase awareness because *"most people don't know what hearing loss is like"* (P3) and *"even though the technology is there, it's not well thought out"* (P10).

To summarize, living in a hearing and ableist world means all participants have to navigate inaccessibility frequently. Here we have described access practices (such as setting

communication norms) and access hacks our participants shared. In the long term, advocacy plays a huge part in improving access. This includes both self-advocacy in relationships, and advocacy for others in the community. Underpinning any of these access practices and reorientation strategies is the listener's analysis of dynamics in the moment. Some key factors are disclosure of hearing loss, cooperativeness from conversation partners, their own role in the conversation and power dynamics (from work hierarchies or sociocultural respects).

3.4.4 Design Probe Findings

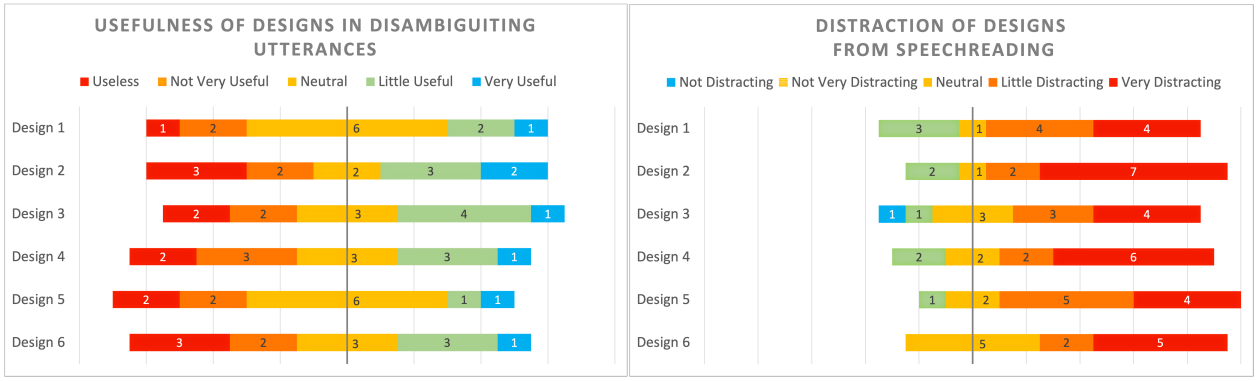


Figure 3.4: Results of Likert Scales on (left) Usefulness for Bottom-up Designs; the scale "Useless" to "Very Useful" is represented left to right in each bar. (right) Distraction from Speechreading for Bottom-up Design. The scale "Not Distracting" to "Very Distracting" is represented left to right in each bar. The line indicates neutral sentiment.

As described in Section 3.3, after a series of speechreading disambiguation tasks, participants were asked to share initial reactions and rate usefulness and distraction of each design. Individual participant statistics can be found in the appendix (Table A.1). There are mixed, lukewarm reactions for usefulness, and all designs were rated as distracting (Figure 3.4). Here we discuss the qualitative results, which were mostly negative. P2 and P12 both commented on cognitive load: *"Too much information to process at once. Old eyes and old brains don't process information as young eyes/brains. It's hard enough to speechread without trying to*

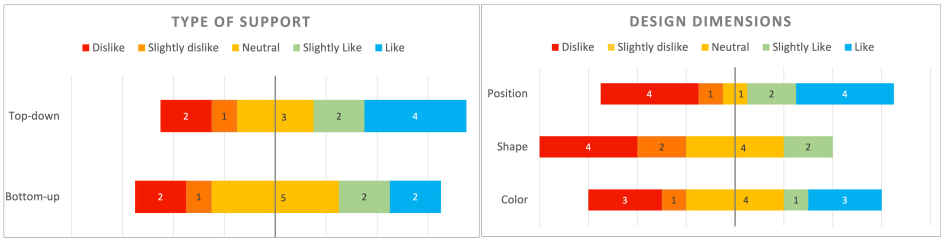


Figure 3.5: Results of Likert Scales on (left) Type of Speechreading support and (right) Design Dimension in Bottom-up Designs. The scale "Dislike" to "Like" is represented left to right in each bar. The line indicates neutral sentiment.

focus on two things [the design and speaker's face] at once." (P2). Additionally, some participants (P1, P6, P9) were not convinced about the promise of phoneme-level annotations over things like cued speech and captioning. Participant P11 found the bottom-up designs promising, but thought they would be strengthened by a less-is-more approach, where annotations would be displayed only for sounds he struggles with. P8, who has tried to learn cued speech before, remarked that she would not invest time in learning a system in this stage of her hearing loss journey. In contrast, top-down designs received higher ratings than bottom-up designs (Figure 3.5). Commenting positively on these designs, P12 said: *"Sometimes speechreading reminds me of putting together a puzzle. If I can figure out two or more words then usually I can piece together the sentence."* (P12)

These results suggest that we would benefit from contextualizing and understanding the nuances of what worked and did not work. We can learn from the aspects of the designs that participants liked. Further, P5 noted the idea of encoding information using these design dimensions does not have to be limited to phonemes, but could be used for other factors such as speaker identity and direction.

Position: The use of position, and the movement associated with it, can be distracting. This was made worse by the fact that participants were trying to read the phonetic text. *"The way the design moves around the screen is very difficult to follow. It's almost like playing*

whack-a-mole with my eyes. After a few minutes of this, I would definitely have a headache and be super exhausted." (P9). Despite this, it was the most liked design dimension, perhaps because it is easiest to distinguish.

Color and Shape: P1 thought color was easier to identify peripherally compared to shape, which requires foveal (sharp central) vision. Shapes were also harder to distinguish, with pentagons and hexagons merging together. P5 said larger shapes might be easier to distinguish but liked how color was "not so dull to look at". This is reflected in ratings of design dimensions, where color has much more positive ratings than shape (Figure 3.5).

To conclude, the bottom-up design probe was mostly unfavored by participants. While we cannot be certain that whether participants' negative response to the design probes was due to our designs, the fact that participants were not trained, or the overall concept of bottom-up speechreading visualizations, we argue that participants' response is valuable for several reasons: First, they provided clear guidance about factors that should be ameliorated in future designs. For example, they were concerned about the visual and cognitive overload posed by the bottom-up designs, particularly based on position. Second, this probe represents the first opportunity that we know of for d/DHH people to respond to bottom-up visualization designs in the context of full sentences of continuous speech, and in comparison with top-down designs. Our results provide compelling evidence that top-down designs (such as ContextCueView [94]) should receive much more attention than they have in past research. This allows us to redirect future research in this space to consider the complexity of communication and access provision. Third, as we will see in the next section, participants entered the design ideation session with multiple ideas that go beyond viseme disambiguation. Thus the contribution of our design probe is not just to the probe designs themselves but rather in participants' reactions to the designs and the brainstorming this led to in the design session phase of our study.

3.4.5 Design Session Results

During the final session, participants shared a diverse set of ideas, codesigning with the first author. The author actively participated in brainstorming, leveraging her own perspectives as a technologist and speechreader, suggesting scenarios where technology could help based on the initial interview, and by building on the participants ideas. Similar to the initial interviews, many of these ideas go beyond speechreading alone to address interactions with other AT, environmental affordances and social dynamics. While some of the features described below may exist in one platform or the other, there are idiosyncrasies in how captioning is accessed and implemented in different videoconferencing platforms. Our goal in presenting these ideas is to share speechreaders' often omitted views on videoconferencing accessibility options and offer insights future iterations of videoconferencing technology.

Speechreading Online:

Many participants set access norms in their relationships. However, conversation partners often forget or are new to these norms, and there is a lack of infrastructure to "normalize" this discussion. Having these guidelines come from *"generalized source"* (P6) was important to negotiate access without disclosure. Further, with multiple and varying disabilities present, participants discussed having a set of *"accessibility guidelines"* (P1) automatically compiled from different people in meeting. These would include support for different disabilities and access needs ("Please describe any images presented during the meeting") giving a lesson on *"Zoom etiquette"* (P9).

Importantly, unlike many technological solutions, access norms frequently involve both conversation partners: access is a cooperative activity. Infrastructure that reinforces this cooperative work can be beneficial. For example, participants described having to remind conversation partners to speak slowly and clearly. Built-in *conversation partner feedback* supports, to prompt a speaker to slow down and speak clearly during a video call (such as reaction buttons) could remove the need to interrupt. Another way would be to automate

speaker feedback that notifies them if they speak beyond a certain speed. Participants P9 and P11 liked how this would reduce their advocacy burden as it gets tiring to keep reminding people. Participants P2 and P3 were skeptical, mentioning that the speaker would actually have to pay attention to the feedback (cooperation) and not become desensitized to it. In their experience, people revert to their speech habits sooner or later. Participant P8 mentioned "those who care will welcome feedback". Technology can also help to support non-speech access improvements. For example, P3 often has friends use cues such as "On another topic,..." so she can follow transitions. Extracting keywords and displaying them separately (e.g., nouns and verbs; subject object) would be valuable in providing context. Participant P11 mentioned from all of the design probes, this would be the easiest to adjust to and adopt.

Another alternative for supporting speechreading is simulation using an animated face. Speaker variability i.e. movement of lips, speech patterns, accents and use of body has a big impact on speechreading ease. Participant P2 brought up the idea of a "Talking Head" to mitigate issues with variance. Inspired by oral interpreters, this would be synthesized human-like head that would repeat what was said in a lip-readable manner. Having this same head for all conversations would let listeners "tune into" this speech pattern. Other participants built on this idea by considering what lip-readable speech means. One aspect is clear lip movements and another is "accent-free" speech (i.e. matches listener's accent). Such an animation could also correct for issues with lighting, angle and proximity. The aesthetics and realness of the synthesized face and positioning are also important considerations.

While much of brainstorming focused on partner communication, participants P7 and P12 discussed wishing for *speechreading learning* tools when they initially lost their hearing. This could have familiarized them with what speechreading entails and what information they could leverage. Whether onset of hearing loss is sudden or gradual would influence this wish. In contrast, P6 found no need for formal training: *"I was losing my hearing slowly so was like riding a bike with training wheels, and you keep moving the wheels up and up [over course of 50 years]."* (P6).

Combining Speechreading and Captioning:

The simplest way to reduce the divided attention while using speechreading and captioning in tandem would be considerate *caption placement*. Captions are often streamed in separate windows (e.g., through StreamText), and overlaying multiple windows on a small screen to allow for simultaneous use is time consuming and tedious. Participants P1 and P12 thought the first step would be to integrate captions in the video calling window, followed by allowing participants to move it closer to the speaker as they wish. Ideal placement would vary person to person (near chin, cheek, above or below head). Dynamic captions that adjust to the speaker's movement would reduce the need to re-position captions to ideal position relative to speaker's mouth. Caption placement for group conversations in grid view warrants additional thought as placing captions in speaker's frame (P1) may complicate speaker identification.

Degree of captioning use offers additional considerations. First, for those who rely significantly on captions, it is difficult to leverage audio with disconnects between audio, video and captioning. As captions are already delayed, and can't always be sped up, we discussed delaying the audio and video stream for it all to be synchronized. Some participants loved the idea while others had reservations. Participant P5 wanted to see information at the same time as everyone else in the conversation, and P10 thought it would make it difficult to participate without being interrupted. Participant P4 felt he had adjusted to the lack of synchronization after years of caption use. Second, for those who only glance down to captions when they need them, like P9, it is often difficult to tune back into speechreading.

> *"If I missed something and then I am trying to figure out what was said – then I am kind of busy trying to fill in that blank. I miss what's said immediately after."* (P9)

To reduce the need to search captions, we considered highlighting the relevant phrase in the captions. This feature could be eye-gaze triggered, where the moment the listener stops looking at the speaker's mouth and shifts to captions, the corresponding phrase could be

highlighted when captions appear. Alternatively, this could manually be indicated using a mouseclick.

Lastly, participants shared ideas centered around supporting multilingual speechreaders or speechreaders who are new to a language. For multilingual conversations, a simple but useful feature would indicate the language of the conversation or the most recently uttered phrase (a valuable type of top-down context). Additionally, being able to set multiple languages for captioning could be promising. Current systems often eliminate unrecognizable words, not realizing they might be a different language. This multilingual captioning could switch between languages as needed. P4 emphasized the need to recognize there are different levels of understanding: audio and literacy, and people may not be equally comfortable with both. So phonetic transcription, where any language pronunciation is rewritten in another language (e.g., Hindi words written in English or vice versa), could support accessibility in any language. Participant P9 suggested having two sets of simultaneous captions: one uses original orthography and the other has a phonetic transcription. This could support novice literacy and language learning.

Other solutions:

While our brainstorming focused on speechreading in video-calls, participants made suggestions to improve richness of captions alone. In scenarios where speechreading is not possible or difficult, some participants (P4, P1 and P5) brought up the idea of embedding emotions in captions for rich connection. Emotions could be matched to color and integrated with captions (e.g., red for anger, blue for sadness, yellow for happy), keeping in mind cultural context and color theory. As this idea was run by other participants, there were concerns about algorithmic accuracy in interpreting human emotion. P8 thought that such extraction would be redundant, preferring to rely on video for nonverbal cues. Participants also envisioned technology to support wider range of experiences. For example, to bridge the gap between online and in-person modes, some participants brought up augmented reality, which would take some of the above described technologies, and bring them to in-person. Simi-

larly, P8, who misses body language and the 3-dimensional nature of in-person interactions, thought VR could be promising in bringing that to video-calls.

Finally, participants underscored the need for awareness and policy to complement design solutions. In light of their advocacy work with the community, they find there is a need to improve visibility and awareness of i) access supports offered and ii) access needs of the d/DHH community. This was reinforced by two participants:

> *"I do think that the general public would care about these issues if they knew what the issues were. I think a lot of it is just out of ignorance and lack of exposure."* (P9)

> *"Leaving it to the person who needs it, bad idea. Sometimes the person who needs it, doesn't even know what's available or what could help."* (P6)

3.5 Discussion

Through our study, we have shared lived experiences of speechreaders and why they value speechreading. The effort and skill that goes into attending to all streams of communication – words, body language, context – highlights the value placed on truly receiving and understanding what is communicated. Yet many accessibility technologies and communication platforms focus heavily on hearing: the auditory and the semantic part of language. While we do not dismiss the value of the semantic, we encourage technologists to consider all ways of *listening*.

3.5.1 Technical, Environmental and Sociocultural Considerations

Looking beyond functional goals, such as hearing or listening, we argue that accessibility technology cannot negotiate access without understanding and contextualizing it. Our conversations with participants highlight the myriad of considerations that go into provisioning access. Expanding on McDonnell's suggested framing [190], we reflect on technical, environmental and sociocultural aspects of technology design.

Technological Constraints on Speechreading Many participants use multiple ATs (hearing aids, captions and speechreading) in concert. Research has focused on these individually, as evident with improving ASR and hearing technology. However, interactions between these technologies remain unaddressed. We highlight how captions offer feedback and a fail-safe for speechreading, but also visually distract from the face. Addressing the tandem-use of captioning and speechreading requires its own set of access modifications as seen in our design session–demonstrating that it is not enough to evaluate an AT based on individual use. Some participants consider audio to be a part of speechreading, and others consider it disjoint. Designers need to consider interactions between AT that come from real-world use.

Environmental Impacts on Speechreading In exploring the differences between speechreading online and in-person, we highlight the affordances offered by both modes. Importantly, differences across modes are non-trivial, and future study of access practices should study both environments in depth. Hybrid meetings may enhance affordances or exacerbate challenges in unique ways. Augmented reality, as suggested by participants, has promise in offering the best of both worlds. Often implicit in design of these videoconferencing platforms is choice of language of conversation. By presuming English-centric and monolingual speakers, we leave out an estimated 50% of the world population that is multilingual and code-switches often [9].

Sociocultural Influences and Access Provision In eliciting reorientation strategies and access practices from participants, we discovered sociocultural aspects that impact access provision. For example, disclosure of hearing loss has a huge impact on reorientation strategies used; any access technology would similarly mediate disclosure. In navigating inaccessibility, technology has the power to decide the division of labour for access. For example, embedding accessibility guidelines and speaker feedback features into videoconferencing infrastructure can emphasize that accessibility is co-created, and thus cooperation can

be prompted. The divided attention caused by captioning and gaze on lips while speechreading are both in conflict with social norms of maintaining eye contact while listening, thus impacting adoption.

Together, the complex interactions of technical, environmental, and social factors underscore the importance of contextualizing AT research and engaging with the interdependence framework [29]. *"Communication is a two way street..."* (P2) – the role played by conversational partners, whether they are allies, strangers or others with hearing loss, cannot be separated out in speechreading accessibility. How AT interacts with other people in the environment i.e. through visibility can reinforce credibility or stigmatization of disability [82].

3.5.2 Summary and design recommendations

In this chapter, we present an exploration of the design space for speechreading supports through design probes and design sessions. By engaging speechreaders in early stages of design, we hoped to set priorities for design in this space. Our participants had the choice to iterate on the bottom-up design probes, or come up with their own ideas for speechreading supports in the design session. Most participants chose to do latter, offering some valuable insight.

We followed prior approaches by focusing on viseme disambiguation in our design probe [94, 93] . Thus, our cued-speech-like supports focused on increasing the upperbound for speechreading accuracy. In contrast, the speechreading specific designs brainstormed by participants (talking head and topic extraction) instead focused on reducing the variability of speechreading accuracy across speakers and contexts. The other design ideas also spotlight a myriad of ways technology can support speechreading beyond accuracy: reducing cognitive load, mitigating variability, enhancing access to richness.

Bottom-up supports may still be valuable for some speechreaders, but may require a different approach. Most participants found it hard to imagine integrating more annotations to the visual field. They learned speechreading incidentally i.e. through necessity and

exposure, and were not interested in investing time to learn complicated annotation systems. Thus, the intuitiveness and learning time for speechreading supports is a crucial consideration in design.

Additionally, addressing environmental and sociocultural challenges is important. In designing their own technology, many participants favored accessibility guidelines and awareness campaigns for their potential to reduce the advocacy burden. Our empirical accounts of speechreading highlight examples of allyship and interdependence. These acts can be seen as a form of linguistic care work, i.e., the work done to ensure all conversation participants can understand and be understood [113]. How technology offers infrastructure for such linguistic care work and access intimacy is another crucial consideration in design.

Many innovative systems have already been explored in research (e.g., topic changes [123, 56], dynamic caption placement [119], caption highlight [137], AR [166, 208], emotion embedding [209, 265, 96]), but could benefit from considering speechreading needs. For example [123] helps d/DHH users track topics and speaker identity. They found that topic changes were distracting for participants while our participants specify these transitions are exactly what they need visualized. One explanation for this difference could be participant speechreading practices. In another example, our participants suggest using caption highlights to quickly find missed words when speechreading during video-calls. In contrast, prior work finds value in highlighting *important* words [154]. The combination of missed words and important words could be very powerful for speechreaders. Similarly, our work adds to prior discussion about speaker feedback features [190, 239] and caption placement [215, 128] by offering insights specific to video-calls and speechreaders.

In interpreting these prior works, it is hard to make direct comparisons to our own results because there is a lack of clear information about participant communication practices. Most prior papers include a range of d/Deaf and hard-of-hearing participants and do not specify the degree to which participants make use of speechreading, sign language, and captions in-person or online. As seen in our interviews, there is large diversity in access practices and experiences across speechreaders and caption users and even signers. It would be valuable

to see future research include these nuances when presenting participant demographics.

Lastly, our results suggest new spaces for research that have not been explored such as video synchronization, language detection, and multilingual phonetic transcription. Prior work has examined the impact of audio-video synchrony on lipreading [151], and could similarly study video-caption synchrony in dynamic conversations over video calls. Language detection may offer valuable support in code-switched conversations, but use may be complicated by accuracy and speed of existing algorithms. Exploring phonetic transcription may reveal complex interplay between literacy, fluency, and speechreading accuracy across languages. A growing body of literature studies the use of non-verbal and social cues over video call [97, 81, 234] and can inform exploration of design spaces described in our work.

3.5.3 Limitations

While our 12 participants had a wealth of diverse experiences to share, a larger sample size could draw broader insights. In addition, our sample did not include cued speech users. While this is a small subset of d/DHH people, their expertise with cued speech could have led to new insights and design directions. Additionally, we only recruited from the United States of America. While our participants had a range of cultural and language backgrounds, recruiting from more countries would provide a valuable global context and provide insight on speechreading in other languages.

Our design probe study was limited to a small number of static sentences without any training. Overall, the results were more negative than positive. A significant amount of new research would be necessary to implement a field-ready deployable system that could annotate video on the fly based on recognized phonemes; however a next step might be to hand annotate more complex videos and provide additional training. We are still assessing whether this is the best option given the negative reactions of participants.

3.6 Conclusion and Future Work

Human communication is rich and nuanced. *How* we say say things matters just as much as *what* we say. Many videoconferencing platforms feature automated captions to support d/DHH individuals, giving access to semantic components (e.g., words), but not much is known about the interaction of captioning with speechreading. Through our conversations with 12 d/DHH individuals, we have demonstrated the intricacies of speechreading, and the myriad of ways to provision access. We identified technical, environmental, and sociocultural factors that contextualize communication accessibility, and emphasized importance of the interdependence framework. Our exploration of the design space of speechreading supports address a mix of technical (disambiguating visemes), environmental (multilingual conversations), and sociocultural (accessibility guidelines) issues. Our work also explores the interactions between speechreading and captioning. Further, we argue that future studies should take a contextualized approach that spotlights diversity in communication practices and how they impact technology use. Finally, our work highlights the role of interdependence in provisioning access and argues for a culture of considering diverse experiences and co-creating accessibility.

Chapter 4

EXPLORING MULTILINGUAL CAPTIONING FOR ACCESSIBILITY

The previous chapter highlighted the many linguistic factors that impact speechreading accessibility, and design of proposed speechreading technology. In this chapter, I dive further into the intersection of language and disability, exploring captioning experiences of multilingual and disabled individuals. This work has been previously published at CHI 2025 "Towards Language Justice: Exploring Multilingual Captioning for Accessibility", coauthored with Rahaf Alharbi, Stacy Hsueh, Richard Ladner, and Jennifer Mankoff [70].

This chapter explores captioning, which is typically studied through a disability lens in HCI, through a language lens. The empirical accounts and analysis highlight the different (language-specific) factors that impact communication accessibility – such as fluency/comfort with a language, linguistic affordances, and cultural perceptions. In line with the framing of repertoires, this chapter showcases how DHH individuals navigate disparities in captioning across languages by using practices of multilingualism (translations, transliterations, translanguaging) – thereby leveraging their full communicative repertoire. It concludes by outlining immediate directions for development and guiding principles for future language-centered captioning technologies.

4.1 Motivation

Captions are a synchronized text representation of audio in pre-recorded media or real-time communication, often used to provide access to those with receptive communication access needs such as d/Deaf and hard-of-hearing (DHH) and neurodivergent individuals [153]. A growing body of research in HCI is dedicated to understanding and improving captioning ex-

periences (e.g., [110, 154, 10, 65, 11]). However, much of this research explicitly, or implicitly, focuses only on English captioning, erasing the needs of disabled people who communicate and consume content in "othered" languages. Over 50% of the world's population is multilingual i.e., understand or express themselves in more than one language [99]. Even in the United States, about 67.8 million people speak a language other than English at home [75]. These statistics suggest a need to explore individuals' captioning and accessibility experiences across different languages.

Captioning is not the only technology with language disparities. Only about 500 of the world's 7000+ languages are represented online, with English (and more recently, Chinese) dominating infrastructure and resources [271]. Speakers of these digitally disadvantaged languages are then themselves disadvantaged, with impacts on their communication practices as well as their adoption of technologies [292]. In addition to English centrism, researchers have called out monolingual bias present in technologies (i.e., supporting only one language), which stifles multilingual practices, such as switching between languages at the word, sentence, or document level [249]. Monolingual design is typical in current captioning interfaces as well: e.g., Zoom only offers automated captions in 35 languages (many of them still in *beta*) and only one at a time [297].

While research in natural language processing and speech recognition has begun to include low resource languages [226] and multilingual contexts [200], this research has not yet influenced the user experience of captioning, perhaps because these technologies are mostly not deployed. We argue it is necessary to understand how caption users are navigating variability in accessibility support across languages today. While some scholarship has explored captioning experiences in different monolingual contexts and regions [79, 24, 215, 258], not much is known about how captioning experiences differ across linguistic contexts and how multilingual practices are represented by captioning. Thus, in this work, we sought to understand the current practices and challenges experienced by multilingual and disabled users of caption technologies – how can we design captions to support multilingual futures and all ways of communicating?

Given the intersectional nature of this work, we sought out theoretical frameworks that could help guide our analysis of the results [109]. In particular, we draw from the concept of language justice. Often used by interpreting agencies and community activists, language justice advocates for the right for everyone to communicate in their own language(s) and strives to create a world where no language dominates over any other – where we 'communicate across differences without erasing differences' [2]. Our analysis questions whether the goals of language justice align with multilingual caption users' needs, and explores how multilingual captions can be a site of language justice.

We conducted a two-part study consisting of semi-structured interviews and diary logging with 13 multilingual and disabled caption users. Our participants represent a range of disabilities, and many languages were a part of their lives. Through elicitation of multilingual experiences in interviews and logs, participants reflected on multilingual accessibility – how it has been shaped by (un)availability of captioning, their comfort with the language, and sociocultural perceptions of access – and how it might be improved.

In this work, we contribute:

1. Empirical accounts of disabled captions users' experiences navigating accessibility in a variety of languages, unearthing how fluency, linguistic affordances, and cultural perceptions impact caption design and adoption,

2. Descriptions of how participants' practices of multilingualism – such as translation, transliteration, and translanguaging – are supported or hindered by captions, thus articulating what it means for captions to *be multilingual*,

3. Suggestions of how we might address existing inequities– offering both immediate directions for development as well as guiding principles for future work informed by language justice.

Overall, we argue for more integrated perspectives on language and disability in communication accessibility i.e., rethinking captioning as a language technology (as well as an

accessibility technology). We believe this reframing allows us to recognize the multitude of ways language needs and access needs of users interact and collide with each other – highlighting the need to engage with language lives of caption users and decenter spoken English in the design of captioning technology. It also illuminates exciting new avenues for research, and we invite researchers and technologists to take up this work.

4.2 Related Work

In this section, we begin by giving a brief overview of empirical research on captioning. Then we discuss multilingualism and related terminology that is foundational to our work. Lastly, we highlight how languages and technologies are intertwined, outlining ways technology research reinforces ideas of monolingualism and English supremacy, emphasizing the need for multilingual captioning research.

4.2.1 Captioning Research in HCI

Captioning is defined as "the process of converting the audio content of a television broadcast, webcast, film, video, CD-ROM, DVD, live event, or other productions into text and displaying the text on a screen, monitor, or other visual display system" [204]. Captions can be generated by humans or by using automated speech recognition software. In addition to captions, which transcribe text in the same language as audio, subtitles are also prevalent in prerecorded media. These are translations to another language, sometimes referred to as interlingual subtitles.

Deaf and Disability Perspectives. As mentioned in Chapter 2 there is an extensive literature on captioning as an accessibility technology for d/Deaf and hard-of-hearing (DHH) individuals. Captions can also offer access to neurodivergent individuals, those with auditory processing disabilities, etc. However, research on captions with these populations is fairly limited. While research has more broadly explored neurodivergent individuals' communication needs (e.g., [296, 62, 295, 131]) and established captions as an access support for

some neurodivergent individuals (e.g., [6, 62])–captioning is the main focus for only a few works that include neurodivergent perspectives [189, 243]. Both works examine captioning of TikTok videos, including content creators' [243] and consumers' [189] perspectives. There is still a need to explore how captions support access needs of this broader group, and how they can be better designed to do so.

Language Perspectives. A closer look at captioning papers published through 1994-2019 at CHI or ASSETS [177] reveals that English dominates given the geographical context for much of this work. While a small body of work has begun to explore different languages captions (e.g., French in classroom contexts [79], Greek in theaters [24], Chinese in augmented reality [215], Japanese in museums [255]) and accessibility in different regions (e.g., India [251], China [49]), there is not much discussion of how language or the broader cultural context shapes the practices of captioning. Takagi *et al.* [258] note the over 4000 characters in Japanese add complexity to real-time human captioning– highlighting that language can significantly impact constraints and affordances of captioning. Notably, advances in speech recognition are slowly increasing the availability of automated captioning on platforms (e.g., Google Meet, Zoom, YouTube, TikTok) – but the users' perceptions of quality and practices of use have not been explored. Additionally, captioning has mostly been studied in the context of fluent users. Proficiency in language also impacts caption use – some research has explored how literacy impacts DHH individuals' experience of captioning, noting that individuals with lower literacy are are less critical of technology probes [32], and examining how we might support reading ease (e.g., [156, 137]). There is room to further explore the impact of fluency on caption adoption and desirability.

Captioning has also been examined from a media studies and audiovisual translation perspective, such as captioning practices in TV shows and films. These works shed light on how to tailor captions for an audience that might not have linguistic access (i.e., through translations) or auditory access (e.g., subtitles for the DHH [3, 291]). A notable example is Gonzales' case study of multilingual content creation [88], which surfaces unique con-

siderations in captioning for linguistic accessibility vs. sensory accessibility and how they might inform each other. Other works of interest include those that explore captioning media containing multiple languages and study how to represent multilingualism in captions [257, 227]. However, the intersection of audiences that know multiple languages (at varying proficiencies) and have access needs has not been explored in depth.

4.2.2 Multilingualism and Related Terminology

There are over 7000 languages in the world [77]. Boundaries between languages, and distinctions between languages and dialects, are not objective but rather socially and politically determined [45]. There is a wide variation in the sounds or phonemes belonging to each language as well as practices of transcribing these sounds. The set of symbols used to represent the language is known as the script. There is not a 1-1 relationship between languages and scripts (e.g., English and French both use the Latin script, Hindi and Marathi use Devanagari, Japanese uses both Hiragana and Katakana, some languages are not written at all).

Multilingualism has been the study of many fields and has a range of definitions [53]. Roughly, a multilingual person is "anyone who can communicate in more than one language" [169]. Multilingualism can also be characterized from a societal perspective, where it refers to the ability of institutions and individuals to engage in multiple languages in daily life [53]. Notably, multilingualism does not require equal fluency in all known languages, and more so refers to an individual's ability to use these multiple languages in everyday life [100].

Practices of multilingualism refer to how the multiple languages in someone's repertoire come together in daily life. They may use **translation**, which involves communicating content from one language in another language while preserving or retaining meaning. They may use **transliteration**, which is the process of writing content in one language using the script of another language (e.g., transliterating Hindi in the Latin alphabet gives us Romanized Hindi). Lastly, they may be **translanguaging**, which often refers to the practice of switching between languages as needed (e.g., during a conversation). Translanguaging

allows multilinguals to use their full linguistic repertoire and communicate without "watchful adherence to boundaries between languages" [210]. These practices of multilingualism are important in an increasingly globalized world – they support cultural diversity and expression of identity. They also weaken forces of language dominance and assimilation by allowing individuals to retain and share different languages [172].

4.2.3 Technologies and Language (In)justice

In addition to lack of support for certain languages (see Chapter 2), digital environments introduce unique considerations and constraints for practices of multilingualism. For example, researchers have unearthed biases in machine translation and language modeling (e.g., [112, 268, 175, 235]) and highlighted lack of support for certain scripts [13]. While some researchers have begun to explore multilingual users' practices navigating these constraints (e.g., [213, 195, 250]), most attempts to make technologies 'multilingual' are disconnected from actual user practices [249, 236].

Within HCI, more and more researchers are exploring multilingualism and arguing for more multilingual research [117, 149]. In their work on non-native English speakers experiences with mobile phones, Karusala *et al.* propose an intersectional approach for language in HCI, that asks researchers to question "how users put diverse facets of identity through languages and the forces supporting or hindering their uses of language" [142]. Similarly, Cardinal *et al.* propose multilingual UX as generative cocreation [50].

However, not many of these works explicitly engage with language justice, besides a few exceptions [283, 107, 106]. Notable is Cardinal *et al's* work, which focuses on how to move from language access to language justice in our community work [51]. Given that there are many ways language justice can be enacted through existing technologies, it would be promising to have it as an explicit goal guiding research and development.

4.2.4 Summary

For captioning technologies, current disparities in availability across different languages raises questions about how multilingual users navigate variability in support and what their access practices are in multilingual contexts. These multilingual contexts additionally bring variations in fluency and literacy, and thus also raise questions of how individuals' language needs and access needs come together. There is an opportunity to understand how captioning design and use can be informed by language and culture, and how captioning might enact a language justice agenda.

4.3 Team Positionality

All members of the research team identify as multilingual. Taken together, we have experience with languages such as Gujarati, Hindi, Arabic, German, Swiss German, Mandarin, French, American Sign Language, and English – each with varying fluencies from native speakers to learners. Some authors use captioning for communication access needs in different languages e.g., when talking to family and friends, watching films, scrolling social media videos, and meeting on remote video software for work. The authors experience in a multilingual world, and their encounters with non-English captioning for accessibility prompted their interest in pursuing this study. Prior to designing the methodological procedure, two authors spent about two weeks documenting the captions they consumed in different languages and how captions' norms, format and (lack of) availability shaped their access needs. This reflexive exercise helped the research team develop informed questions while also keeping in mind how participants' experiences may be different to that of the authors.

4.4 Methods

Given the dearth of research on multilingual caption users regarding their experiences in different languages, we aimed to understand the accessibility of multilingual scenarios, the availability of captioning, and how it might be improved. To do this, we conducted a two-part

study consisting of semi-structured interviews, diary logs, and follow-up interviews.

4.4.1 Procedure

In the initial interview, we gathered the language background of participants to understand where and how often participants encountered different languages. Similarly, we asked participants about their disability identity, their access needs, and how captions supported (or sometimes hindered) these needs. We asked them to share and contrast accessibility of their experiences across different languages and in multilingual contexts. Particularly, we were interested in the role captioning played in these different scenarios, its availability, and quality. We also asked participants to comment on how culture shaped these access mediations in different languages with different people.

Following the initial interview, participants were invited to log their multilingual interactions over a course of a week. This diary study was designed to prompt deeper reflection on accessibility of multilingual interactions and elicit articulation of specific experiences from participants' daily lives (rather than a broad overview). We sent participants specific instructions at the beginning of the logging period that outlined the different kinds of experiences that would count as "multilingual" (Table 4.1). These included personal conversations with friends or family, interactions with acquaintances online and in-person, social media consumption (vlogs, Instagram reels, TikToks) or entertainment (movies, TV shows) – any experiences that involve multiple languages or non-English contexts.

In these instructions, we also incorporated a personalized list of potential experiences for participants to try based on the initial interview. We also pointed to captioning technologies in the participant's languages that they might be interested in experimenting with, e.g., video conferencing software that had introduced automated captioning in different languages [89, 90]. We sent participants a reminder every day to share logs over text or email. In each log, participants were asked to briefly describe the context, the accessibility of the interaction, and the availability of captioning. The logs were intended to facilitate recall in the final interview. Participants were asked to submit a minimum of 3 logs through the

Mode	Activity Type	Example sources	Ways to Log
Online	Media	TV Shows, Movies, YouTube, Instagram reels, TikToks	Share link to media
Online	Video Calls	Video conferencing platforms with monolingual and multilingual automated captioning (ASR)	Name of platform + Screenshot of captions*
Online	Information resources	News, educational lectures	Share link to video or name of particular talk show
In person	Informal conversations	Conversations with friends and family	Short written description of experience or photo/video/Snap/Reel*
In person	Formal conversations	Workplace/professional interactions	Short written description of experience or photo/video/Snap/Reel*
In person	Random	Interactions during errands, day-to-day transactions	Short written description of experience or photo/video/Snap/Reel*

Table 4.1: An excerpt of diary logging instructions sent to participants. *Participants were told to get consent of conversation partners before taking any photos or recordings.

week. They could choose to condense their logs and send them at the end of the day, or send them as each experience occurred.

For the final interview, we created a customized set of questions for each participant based on their diary logs. For each logged experience, we probed them further to understand the degree of technological and sociocultural support present. We aimed to understand the unique features of available captioning, its quality, and the accessibility of the overall experience in multilingual/non-English contexts (e.g., how did you feel about two sets of captions being displayed in that video? or how easy was it to follow the language switches in that in-person conversation?). In contexts without captioning, we asked participants about other access practices they used. Participants were then called to reflect on accessible multilingual futures.

All study activities were approved by the institutional review board (IRB) at the authors' institutions. All participants were compensated for their time and expertise through a gift card. The first author led and conducted all the interviews, sometimes while other members of the research team observed. All interviews were scheduled for 60 minutes.

4.4.2 Recruitment and Participants

Our work is motivated by discrepancies in the availability and accuracy of captioning across different languages. Since this discrepancy likely impacts all multilingual caption users regardless of their specific disability, we sought to gather a diverse set of experiences by intentionally expanding our recruitment from DHH individuals to also include neurodivergent and disabled individuals who used captions for accessibility. Our focus on captioning meant we sought individuals who had experiences with multiple spoken or written languages, rather than multiple sign languages.

We shared our recruitment messages across a variety of disability and multicultural mailing lists, Facebook groups, and community organizations. We also recruited through snowball sampling. This resulted in 13 participants who identified as caption users for accessibility (i.e., identified as DHH, neurodivergent, and/or disabled), and understand or express them-

Language	Count	Language	Count	Language	Count
Arabic	1	Hokkien	1	Russian	1
ASL	4	Italian	1	Serbian	1
Cantonese	1	Japanese	3	Spanish	4
English	13	Korean	2	Tamil	1
French	5	Malay	1	Ukranian	1
German	2	Mandarin	4	Welsh	1
Hindi	2	Portuguese	2		

Table 4.2: Distribution of Languages in our Participant Sample. The counts represent the number of participants that discussed that language to be a part of their lives.

selves in more than one (spoken or written) language. Of these, nine participated in the diary logging portion of the study and final interviews. On average, each participant submitted $\mu = 12.33$ ($n = 9$) diary logs – and most participants logged more online experiences (i.e., TV, social media, phone and video calls)($\mu = 8.375, n = 9$) than in-person interactions ($\mu = 4.25, n = 9$). To support anonymization, we describe the languages covered in aggregate in Table 4.2.

4.4.3 Accessibility Considerations

All study activities were conducted remotely. Interviews took place on Zoom, and diary logging occurred asynchronously. To support accessibility of the research process for both participants and researchers [178], we offered CART, ASL interpretation, translation, and a copy of questions ahead of time. Some participants preferred automated captions, in which case participants used Zoom captions and researchers used CART. Sometimes they used chat to type out specific examples in their languages (P2) or when they preferred to be voice-off for a portion of the interview (e.g., P1). Often participants and researchers both spelled out

words that were unfamiliar to the captioner (e.g., 'Peranakan').

4.4.4 Analysis

Three of the authors met and analyzed the collected data over the course of several weeks, following a reflexive thematic analysis process [43, 44]. The first part involved reading different transcripts and coming together weekly to discuss points of interest and potential themes. Following this initial familiarization with the transcripts, we conducted an open coding pass. Finally, we clustered generated codes together to larger patterns (e.g., practices of translation, practices of language-switching, cultural dimensions of access) and one of the authors reapplied these higher level codes to all of the transcripts. We further refined themes through the writing process. Overall, each of the transcripts has been read by multiple team members and there have been several discussions throughout the process to iterate on the analysis.

4.5 Results

Figure 4.1 offers an overview of our results. The top half of the diagram corresponds to section Section 4.5.1, where we begin by articulating two different sides of communication access for our participants: their access needs and their language landscape. Through the rest of the results, we illustrate how these two sides of communication access (i.e., disability needs and language needs) interact and inform each other, as seen in the bottom half of the diagram. Section 4.5.2 discusses the factors that impact a participant's experience in a specific language– namely, notions of fluency, language characteristics, and cultural perceptions. Section 4.5.3 builds on these to explore participants' practices across languages, discussing practices of transliteration, translanguaging, and translation. Finally, Section 4.5.4 looks beyond a single experience to the systemic impact of (lack of) multilingual access on participant lives, thus emphasizing the need for holistic approaches to communication access that integrate language and disability perspectives.

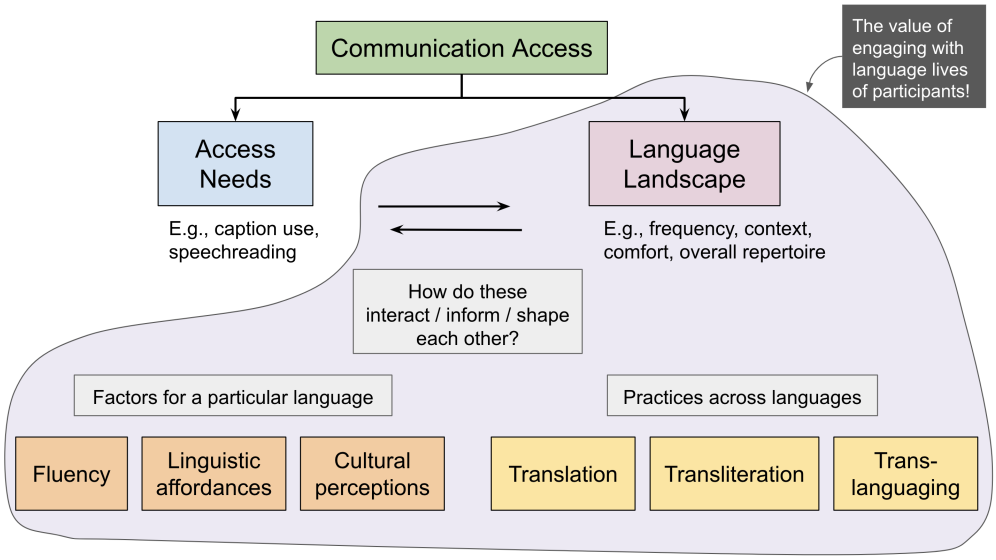


Figure 4.1: A visual map of the results section and our contributions. We illustrate the value of holistic approaches to communication accessibility by highlighting the the many ways disability needs and language needs interact and inform each other.

4.5.1 Communication Access: Considering Both Disability and Language

In this section, we begin by describing the language journeys and accessibility needs of our participants, noting how their relationships with different languages and access often changed through time.

Language Landscape

The participants in our study discussed a variety of languages during the interviews. Since all were in the United States or Canada, they were comfortable with English (except for P3, who preferred ASL). In discussing the different languages they could understand or express themselves in, we found that participants rarely claimed to "know" a language. Instead, they often discussed when they learned the language, their comfort with different aspects (reading, writing, signing, speaking, understanding), how often they encountered it, and in

which contexts.

Participants' language repertoires were often broad and varied (e.g., P2: Korean, Spanish, ASL). The number of languages discussed also varied, with some, like P4, discussed two languages, whereas others, like P5, discussed seven languages. For each of these languages, degrees of comfort with receptive or expressive aspects differed, often shaped by how they learned the language and how they use it today. For example, participants who learned a language from their families only at a young age (e.g., P8: Mandarin, P10: Hindi) were sometimes less confident about reading and writing than those who learned it at home and school. Discussion of literacy was further complicated by numerous scripts available for a language (e.g., P9 could read Serbian[1] in Roman alphabet but not Cyrillic). Additionally, sometimes shared histories between languages meant participants could understand a language they never learned (e.g., P4 understands both Brazilian Portuguese and European Portuguese).

Participants' comfort with a language also changed over time and context, oftentimes with their migration patterns. For example, P2 grew up speaking Korean in South Korea, and only learned English after moving to the United States in middle school. Now, in their mid-twenties, they easily navigate the world in English, yet they feel like they are *"losing Korean"* (P2). Similarly, P12 told us, *"I guess [Hebrew] is my first language, technically, but since moving to [North America] with my family, we didn't really speak it at home [...] it's definitely not at a very comfortable level anymore."* (P12). One participant noted that an emerging fluency of a language is often neglected in Western spaces. P11 explained, *" [Many people's] understanding of how language works is so off-base, [...] they say, okay, you speak French, you should be able to say every word in the entire dictionary, like, you* ***know*** *it, and it's like – that's not the way language works."* (P11).

[1]P9 highlighted that Serbian/Montenegrin/Bosnian/Croatian are considered mutually intelligible language varieties of Serbo-Croatian and attitudes about language are closely tied to the history of conflict in the region.

Access Needs

Our participants included d/Deaf and hard-of-hearing, neurodivergent, chronically ill, and disabled individuals. Some of our participants identified with more than one of these disabilities and some chose not to disclose specific diagnoses (as seen in Table 4.3 and Table 4.4). Therefore, in this section we discuss their communication access needs instead and *how* they use captions to meet these needs.

Many participants mentioned the need for visual components (such as seeing faces and bodies for lip movements and body language) to *speechread* [140]. Speechreading allowed participants to guess what was being said based on lip movements (bottom-up) and on context (top-down). Speechreading often coincided with caption use for our participants, as noted in prior literature [73]. Other access practices included limiting background noise and other sensory input (such as fans, bright lights, or heat). Overall, our participants shared a preference for written modes of communication.

Our participants used captions in a variety of contexts: in-person conversations, video calls, media, and at school or work. Table 6.1 lists some examples of scenarios where they used captioning from their diary logs and interviews. While some DHH participants could request CART as an accommodation at school or work, those with other disabilities were unable to since *"they want you to have a [specific] diagnosis"* (P11).

Captions played different accessibility roles: sometimes they were central to accessing incoming information, sometimes they played a tangential role. For example, for P2, captions were a backup: *"sometimes, I can't really tell what the person was saying. I refer back to the captions"* (P2). For P5 and P6, captions additionally acted as a filter to *"focus on the sound that I'm actually trying to pay attention to"* (P6) and *"[provide] scaffolding for the auditory processing"* (P5). For P7, captions helped hold onto conversational threads and acted as a memory aid: *"retaining information or remembering prompts can be challenging, so I'm actually actively going back to the captions and looking at what you asked for."* (P7).

Often captions played multiple roles for the same participant, depending on their needs in

P No.	Lang.	Disability	Multilingual experience using captions	Multilingual experience without captions
P1	4+	HoH	Watched Tamil movies with Tamil/English captions; uses transcription app inperson.	Conversation with their nieces where they switch between Hindi and English.
P2	4	HoH	Watched a historical Korean drama with Korean captions; used CART in class.	A call with their sister where they switched between Korean and English.
P3	4	deaf	Watched a Chinese drama with Chinese captions.	A chat with friends where they switched between ASL and spoken Chinese.
P5	4+	ND	Watched news in Arabic from Palestinian journalists with Arabic/English captions.	Played multilingual Scrabble with a friend.
P6	4+	ND	Watching Spanish shows on Netflix with Spanish/English captions.	Taking Spanish classes in college.

Table 4.3: Participant Table (1/2) with multilingual scenarios from diary logs and interviews. We only list number of languages to protect participant anonymity. Disabilities are characterized as Hard-of-Hearing (HoH), d/Deaf, Neurodivergent (ND), chronic illness, and auditory processing related.

P No.	Lang.	Disability	Multilingual experience using captions	Multilingual experience without captions
P7	4	HoH and ND	Watched YouTube videos in Mandarin with Chinese/English Captions.	Taking Hokkien classes on Zoom.
P8	2	ND	Watching Mandarin media with English captions with their partner.	Conversations in Mandarin with their family.
P9	4	auditory processing, tinnitus	Went to a movie date watching a Japanese movie with English captions.	A conversation with their parents switching between Serbian and English.
P10	3	HoH, chronically ill	Watched TikTok in Hindi with English captions	A dinner with their mother / call with grandmother in Hindi and English.
P11	3	auditory processing	Watched TikToks in French and German with matching captions.	A conversation with their dog in French.
P12	2	disabled	Watches Russian media with Russian/English captions if available.	In person interactions in Russian.
P13	4+	auditory processing	Watched and listened to Japanese karoake videos.	A chat with receptionist / grandma's nurse in Cantonese and English.

Table 4.4: Participant Table (2/2) with multilingual scenarios from diary logs and interviews. We only list number of languages to protect participant anonymity. Disabilities are characterized as Hard-of-Hearing (HoH), d/Deaf, Neurodivergent (ND), chronic illness, and auditory processing related.

the moment and the communication context. In the absence of captions, participants would either disengage or *"struggle through"* (P5), depending on fatigue and importance. Table 6.1 lists some examples of multilingual scenarios where participants did not use captions, either due to personal choice or due to a lack of availability.

In the following sections, we discuss how different factors for a particular language and across languages impact these roles and shape how captions offer access.

4.5.2 Factors that Impact Communication Access

In this section, we discuss how participant experiences of access are shaped by their comfort with a language, its linguistic affordances, and cultural context. We highlight the impact of these factors on the design and use of captions, demonstrating that 1) we need to move beyond English captioning as the standard and 2) language and culture need to be foregrounded in captioning research.

Communication Access Conditioned by Fluency

As shown in Section 4.5.1, participants have varying levels of fluency across different languages and different types of proficiency within the same language. This provides a rich context within which participants have developed their communication access practices. For example, less familiar languages might demand a slower pace of communication, resulting in participants asking conversation partners to slow down or slowing videos while watching them. For P3, for whom English is their fourth language, *"if we are conversing in English, I need to be a lot slower and they need to be slower with me."* (P3).

Language proficiency can additionally affect speechreading ease. Participants noted that strong knowledge of vocabulary (necessary for top-down speechreading) and the ability to recognize mouth patterns (necessary for bottom-up speechreading) are both crucial for effective speechreading. A lack of familiarity with key vocabulary relevant to the communication context can thus make speechreading difficult in certain languages. For example, P9 watches old Serbian movies with parents and finds *"archaic Serbian"* (P9) hard to follow. Similarly,

P10 sometimes doesn't understand what her grandmother said in Hindi, and attributes it to generational differences in phrases and vocabulary.

While captions can address some of the above concerns, they bring up their own considerations regarding fluency – notably, literacy. Being comfortable with speaking and listening to a language does not necessarily mean comfort with reading it. For languages with scripts novel to participants, reading proficiency had to be attained before captions could offer access: *"[My] confidence with reading [Arabic and Ukranian is] not as strong, so having captions in that language are less accessible or take more time or more energy to parse."* (P5). Not only do captions require familiarity with the script or writing system at hand, they also require a reading pace that matches incoming speech. For example, even French, which has the same Roman alphabet as English, presents different access needs for P10: *"when I was in France and they wanted to watch more fast-paced stuff [...] which I could read [the French subtitles, but] it was a bit tougher, so we would have to watch, like, comedies, or like just change the subtitles to English."* (P10).

In the examples above, participants' lack of fluency or proficiency in language directly affects how they request access or use captions – whether by slowing it down, relying on it more, or avoiding it until literacy is obtained. However, the need for access is not limited to those who lack language fluency or proficiency. In fact, P5 noted feeling like they had more access needs in their native and most familiar language (English) than the other languages, because they are not afforded the leniency often given to those learning a new language: *"Honestly, I think it's . . . the overarching sort of background information of 'you're a language learner' versus, like, 'I expect you to be able to communicate fluently', I think that is a very helpful – at least as far as getting access needs met"* (P5). Communication partners were more willing to alter conversation pace and fluency and be responsive to requests for rephrasing and other access needs when talking to a language learner, rather than someone assumed fluent. This sentiment was shared by P11, who felt there was less stigma for accessing support (such as captioning) in other languages compared to their native language (English), *"I guess there's stigma, and it feels like – I'm obviously an English speaker, why*

can't I just hear what you're saying?" (P11).

These stories show how normative assumptions about fluency impact use of captions *"if you want to be considered fluent in the language, you have to be able to hear it properly, but there are people who can't hear languages properly"* (P13). This raises questions about who is included in the conception of a typical caption user and in what contexts.

Communication Access Molded by Linguistic Affordances

Our participants' perceptions of communication accessibility in a language did not only depend on fluency, but also the *affordances*, or features, of the language itself. For example, while speechreading, P4 found Portuguese speakers moved their lips a lot more than English speakers, making it easier to speechread. Other participants discussed different features of a language that make it feel easier to process or more accessible to them. For example, P11 found the rules for spelling and pronunciation for German made it *"really easy to picture the words, as they're being said"* (P11). Similarly, for P5, *"being able to visually process language"* (P5) and not having to physically speak made ASL feel cognitively accessible.

Interestingly, the match between speech and writing system varies significantly from language to language. Participants remarked on how Spanish is written as it is pronounced (*"really transparent to the phonemes"* (P6)) but other languages such as French and English have less phonetic spelling, which impacts caption readability. P13 discussed Cantonese at length, explaining how writing system impacts descriptiveness of captions: *"there is quite a big difference between what the [captions] say and what is actually being said because spoken and written Cantonese are very different. [... the captions] are usually the more formal version, as opposed to what they say on screen which is a lot more casual."* (P13).

The disparity between speech and text in Cantonese is because of the existence of a standard writing system across all Chinese languages[2] (Mandarin, Cantonese, Hokkien in our participant pool amongst others). This standardized writing system allows written Chinese

[2]Chinese macrolanguage: https://www.kevinhsieh.net/2019/02/27/chinese-macrolanguage/

materials to reach a wider audience. It also support disambiguation of spoken dialects in media with the addition of standard captions. However, this practice of written standardization removes aspects like slang and intonation that are unique to spoken dialects. P13 shared, *"Having subtitles that are a more standardized version of what they are saying is helpful, but at the same time you know it's not exactly one-on-one with what they are saying on screen, so that kind of sucks in terms of accuracy."* (P13).

The above stories highlight the degree and type of information captions encode varies language to language. These linguistic affordances then bring up questions about the inherent ambiguity in captions – if not speech, what do they offer access to? They also call us to reflect on caption metrics like accuracy that may be ill-defined.

Communication Access Shaped by Cultural Perceptions

Participants discussed how certain cultural norms shaped perceptions of disability and accessibility. For example, P7 found that the widespread availability of Chinese captioning on media allowed them to recognize captions *"as a norm that provides access to a lot of different users [...] I knew it was a tool that I could use for myself, even if I wasn't, you know, fully interested in it at the time."* (P7). They found that this norm was lacking in English-produced content.

Other participants offered contrasting experiences: *"I think Western culture is a lot more understanding [...] Indian culture that I've experienced is a little more, like, defensive [...] like, you're kind of treated a little bit like a nuisance if you ask for accessibility."* (P10). These perceptions of accessibility impacted participants' adoption of captions – such as whether they would feel comfortable asking someone to turn on captions while watching media together, if they would openly use a captioning technology in a conversation, or whether they would request captioning as an accommodation.

Cultural norms also influenced how participants thought their access needs would be received. For example, P2 discussed norms of fast communication in Korean: *"Koreans they're just going, doing, and eating, and working just fast and quickly through everything*

without a break. So, whenever I do ask, you know, can you repeat that or [...] ask people to slow down. I do feel a little more self-conscious." (P2). These norms impact whether participants disclose their access needs and thus the role communication partners play in facilitating access e.g., whether they would work collaboratively to make captions work better by slowing down and monitoring errors.

Some participants developed ways to address inaccessibility within specific cultural contexts. P12 was often more covert in expressing their need in Russian, *"I just make light of it, like you know, make a joke around it and it helps the person slow down if I need a repeat."* (P12). This desire for covertness impacted whether they chose to use captions or speechread, since speechreading is less visible and reduces the risks of disclosure.

4.5.3 Communication Access Transformed by Practices of Multilingualism

Given our participants' broad language landscape, the different languages they knew interacted in many ways. Here we discuss how three different practices of multilingualism—translation, transliteration, translanguaging—facilitated or hindered access provision for participants. We also trace how captions embody these practices of multilingualism (e.g., Figure 4.2) and how participants use captions to meet both language and access needs.

Practices of Translation: Depicting Ideas Across Languages

Translation played an important role in participants' lives. While some participants often acted as translators for immigrant parents (P9), less literal translation was a key part of participants lives as well. When P10 and their partner watch Hindi movies together, P10 expands upon the English subtitles, using their knowledge of Hindi to evaluate quality and offer more nuance. In a conversation with Deaf friends signing ASL and a hard-of-hearing friend who was not comfortable with ASL, P3 *"interpreted into spoken Chinese from the ASL that the group was signing"* (P3) and vice versa. P3 was uniquely skilled to offer access in this situation – they know Chinese and ASL, but also needed their hard-of-hearing friend to repeat their spoken Chinese responses to meet their own access needs. These examples make

Figure 4.2: Practices of Multilingualism in Captions from Diary Logs. Left: Screenshot of captions from a video where both English and Hindi were spoken. Captions encapsulate both translanguaging (dynamically moving between languages) and transliteration (converting written script from one to another). The first half of the caption is in English, the second half is in Hindi, corresponding to the audio. The Hindi portion is transliterated to Latin script. Right: Screenshot of captions from a video where Korean was spoken. Two sets of captions are present: translated captions in English (top) and original language captions in Korean (bottom). While the video is monolingual (Korean), the presence of translated captions (English) introduce multilingualism.

clear the ways that language and disability collide – our participants often simultaneously navigated their access needs with regard to spoken language while using their multilingual skill to extend access to others.

Translated captions (or subtitles) were widely available, with media from different languages often captioned in English to increase reach and audience. For those familiar with the language at hand, accuracy of translated captions was sometimes questionable, making them prefer original language captions instead. Along with disagreements about intended meaning, participants found that translations did not conserve aspects that are key to the viewing experience: *"It's either too literal of a translation or it's just the feelings of the translation isn't right. [...] they use a lot of poetic phrases in Chinese, but when they translated it over it felt really flat and boring"* (P8). However, many participants were driven to use these translated captions as they were the only option.

When participants were not comfortable reading a language but were familiar with the language, they used translated captions in interesting ways. For example, when watching Hindi media with English captions, P10 found themselves reading the English translated captions and mentally translating it back to Hindi to match the incoming audio. Similarly, P7 could read simplified Chinese and not traditional Chinese. When they encountered a Mandarin video with traditional Chinese captions, they opted for the English captions and made *"a few mental jumps [...] translating from the speakers [voice] to English [text] and then from English back to Chinese"* (P7). This process allowed them to leverage their full linguistic repertoire and experience all aspects of the media that may have been lost in translation. In choosing type of captions, participants made a calculated decision between the cognitive demands of reading a less familiar language and cognitive demands of this internal translation.

Practices of Transliteration: Adopting Different Scripts

Transliteration is the process of writing words of a language in the script of a different language. For example, Figure 1 (left) includes captions with the words 'Gajar ka Halwa' which

is a Hindi phrase [3] transliterated to the Latin script. Participants encountered transliteration of this kind often in user-generated digital content including text posts, music videos, and open captions on TikToks or reels.

Some participants, like P10, appreciated transliterated captions as they offered access without requiring familiarity with a certain script or literacy in the spoken language. While transliterations removed the need to learn a specific script, they introduce new challenge – sometimes they had to sound out the transliteration to recognize what it meant. Transliteration also introduced a lack of spelling conventions. This hampered reading speed and sometimes comprehension, requiring participants to slow down a video or *"look in the comments to figure out what that means."* (P12). As captions are ephemeral, this impacted the desirability of transliterated captions.

In addition to supporting users with low reading literacy, transliterations aided the process of language learning – such as making pronunciations of unfamiliar words clear, and helping learners familiarize themselves with a new script. However, some participants could not imagine using it outside of a language learning context, or preferred to master the original language script instead: *"Studying so many different languages I can get quite turned around on what sounds certain letters are supposed to make [...] yeah, learning a new alphabet is hard, but after learning the alphabet that a language uses to transcribe itself, like, there's a lot more confidence on my part of, like, what things are supposed to sound like or what things are supposed to be."* (P5). Transliterations also present ambiguity – each language has its own phonemes, and a script different from the original may not be able to depict all the sounds in that language. Reading a transliteration may then be more confusing and require more processing power than reading text in its original script.

[3] A sweet made with carrots

Practices of Translanguaging: Moving Between Languages

Translanguaging, the practice of fluidly moving between languages, allowed our participants to fully combine their linguistic resources. Language switches were sometimes motivated by the need to use specific vocabulary (e.g., P4 switches to English when having technical discussions at work) and often allowed them to foster greater understanding. This practice was not just limited to spoken languages, – for example, when P3 chats with friends who also know ASL and Chinese, they often find themselves fingerspelling Chinese words phonetically in an ASL sentence to refer to a specific concept in Chinese. This practice of translanguaging also supports learning – for example, P2 discussed how he supported his partner who was learning Korean: *"There's like similar pronunciations from Spanish to Korean. If phonetic English pronunciations are not correct for Korean, we'll translate it into Spanish phonetic pronunciations and it will sound a lot better."* (P2). This pooling of resources across languages shows the commitment to collaborative meaning making rather than prescriptive rules of language use.

However, translanguaging is not without downsides – some of our participants found it difficult to follow language switches and identify languages. To navigate this potential inaccessibility, P1 requested multilingual conversation partners to stick to one language during the course of conversation. When jumping into a conversation, P1 used his own responses to clarify the language being used, and sometimes started speaking first to set the language of the conversation. Another participant, P4, took a more direct approach, where he asked conversation partners to clarify whether they are speaking in Portuguese or English when he had a hard time following. This difficulty fluctuates given how expected the language switch is and given how the language and words have been merged together – borrowed words are common in many languages and these often take on new pronunciations in that language: *"[it is easy when] Portuguese pronounce it in a Portuguese way. When it is entirely new [word] [...] I have trouble with that."* (P4).

Participants found that captioning performed in disappointing ways when multiple lan-

guages were present in a piece of media or in a live conversation. For movies or TV shows, human generated captions often depicted the secondary language by name only e.g., [SPEAKS SPANISH]. This approach offered access to neither the phonetic nor the semantic content, making participants frustrated: *"I'm like, so what am I supposed to glean from this?"* (P9).

While some participants understood the choice to not translate might be from an artistic standpoint, e.g., the creator did not intend the audience to understand what was being said in that segment, the choice to exclude both transliterations and translations was deeply frustrating: *"I'm like, excuse me, I'd like to know what they said [...] – yeah, it feels insulting [to the] people speaking and the work and the art itself."* (P10).

While participants were able to toggle captioning language in some contexts (such as Netflix), caption options in all languages had drawbacks, given creators' assumptions of language background of intended users: *"I'm watching this Portuguese show and I have my captioning set to English. And whenever they talk English, the captions disappear because they assume since I put English captions I know English"* (P4). Here, creators are also making assumptions about intended use of captions – to provide language access and not disability access. Some participants attributed this poor experience to not being able to choose multiple languages in captions. This is particularly relevant for automated captioning contexts where the inability to pick multiple languages for captioning caused captions to turn into *"completely gibberish"* (P13). However, this breakdown in captioning was helpful in indicating language switches in a video (e.g., P1 with YouTube). For those who used automated captions in in-person contexts, such as P4, they navigated this breakdown by manually switching the captioning language in their speech transcription app, a cumbersome process. They mentioned that the captioning software *was* capable of detecting language switches automatically – just not for all languages, highlighting inequitable development.

Reimagining Multilingual Captions

We asked participants to imagine accessible multilingual futures and reflect on what they would like to see changed. Many participants desired widespread availability of captions for

all contexts, including media and real-time conversations. For automated captioning, they hoped for better transcription quality and increased accuracy in automatically detecting language switches. In absence of further technical development, they hoped content and media creators would invest in a *"quality assurance"* (P9) process.

Particularly important to participants was the addition of form factors that improved the multilingual accessibility of captions – *"an important metric is how well it handles multilingual settings."* (P4). Some small scale changes with large potential for language switching contexts included indicating the language alongside the transcription *"in brackets above"* (P11) (e.g., [FRENCH] Mais oui...) and allowing users to choose multiple languages for captions on a platform. They also proposed having *"variety of scripts available"* (P7) to support a range of fluencies / literacies. They envisioned being able to use captions from different languages simultaneously – *"show both the translation and the original on the same caption bar."* (P13). This would allow users to leverage information from multiple captions at the same time *"instead of having to open up a new tab, [...] I could look at a translation right away."* (P2). Users could choose which set of captions to focus on moment to moment, based on their comfort with the language, fatigue, and intention.

Participants also sought greater customizability for how language switches are handled and displayed – such agency to pick what would be transcribed vs. translated vs. transliterated based on their comfort in the moment. P13 pointed out, *"It would just be nice to have a lot more learning and accessibility options"* (P13). Customizability would reduce the assumption made about the language and disability background of the audience. Participants also brought up customizability of visual aspects like placement and font size. Even this desire was tied to multilingual fluencies: P10 for example, could catch background audio for Hindi media, *"wherever the captions are I'm fine, [...but] with Japanese I am a little bit more specific just because there's more to [take in] for me. [...] it would impact the size of the captions for me based on my familiarity with the language."* (P10). Lastly, participants emphasized that these technological changes needed to occur alongside social changes. Some participants had experienced negative attitudes around caption use and emphasized the need

to *"destigmatize captions"* (P8) as an accessibility technology. Education on *"how language works"* (P11) (e.g., fluency changing over time and context) would be crucial to address misconceptions and highlight the role captions play in language access as well.

4.5.4 Why Communication Access?

The previous sections have underscored the value of engaging with language in captioning by surfacing factors and practices. In this section, we present parts of our analysis that look beyond a single instance of (in)accessibility to the broader constellation of events, highlighting the systemic impact of (lack of) multilingual access on our participants' lives. This allows us to bring to the forefront *why* communication access is important to participants, and *what* it affords access to. We unearth broader set of contexts and opportunities that participants desire access to – access to language, access to community and culture, and creating spaces where no language dominates or dies. We can see how their current use and vision of multilingual captions align with the goals of language justice.

Multilingual Captioning as Access to Language and Language Learning

Participants emphasized the importance of access to language. Languages allowed them to connect with their *"roots and heritage"* (P1), and learning languages was often necessary to navigate the public sphere (such as school or work or day-to-day errands). However, aspects of learning languages, like identifying new sounds and pronouncing new words were often inaccessible to our participants. Multilingual captioning played unique role in supporting language learning and thus affording them access to language.

For example, participants who were immigrants and learned English after moving to North America (P2, P9, P12) discussed watching TV shows for language immersion in early years: *"we ended up watching a lot of Star Trek of all things because it had English captioning."* (P9). Similarly, P12 found themselves transitioning from using Russian subtitles (translated captions) to English captions (original language) as they became more familiar with English.

Interestingly, the role captions play shifts through their language journey. For example, P2 found that what captions offered access to changed over time – in earlier years they helped to distinguish the rapid stream of sounds from new language, and in later years they offered broader context, *"I think captions will always be helpful for me. [...] I will say though, I think the more fluent you are with a language – least for me – I tend to use it less often."* (P2). P7's language journey with Hokkien was different – they were exposed to the language while growing up and lost touch in adulthood. They took a beginner's class on Zoom to pick up the language again – unfortunately, there were no captions available. Remarking on the experience, they shared, *"I was a little bit ahead of the curve. So I did okay without the captioning but I imagine had I advanced to the third or fourth level, I probably would have needed access to captions."* (P7). At an advanced level, small nuances in sound and speech would have been important to follow necessitating need for captions.

These examples highlight the dual role multilingual captions play in providing *both* language access and disability access against the backdrop of our participants' language journeys.

Multilingual Captioning as Access to Culture and Community

As media offered participants an opportunity to immerse themselves in a language, it also offered a gateway to culture and community – only if the media was accessible to participants, *"language shapes your understanding of the world, right? And captioning [is] also a gateway to that."* (P7).

Inequitable development of captioning across languages and poor quality of captions made participants reluctant to engage with non-English content. For example, for P13, *"a lack of captioning, or maybe a lack of me wanting to find the captions for French, has made me disconnected to a lot of the culture, a lot of the things that I learned."* (P13). This exclusion from pop culture and digital media spaces had far-reaching consequences for community connection as well. P9 shared, *"Cause we don't only watch media by ourselves. [...] we might you know at home [...] but it's part of a conversation you're gonna go 'hey, have you*

seen this movie?' and then if you can't ever see anything like how will you feel like you're close to those coworkers or family or friends?" (P9).

Participants often learned languages to connect with their roots, or to foster deeper connection with friends, family, or partners. By offering/hindering access to language, captions could offer/hinder access to culture and community. For example, P7, who desired to connect better with their Peranakan Malay ancestry remarked on the lack of both original and translated captioning on shows uploaded to YouTube, *"while most people understand that we've experienced cultural loss and that younger generations do not have access to their Peranakan heritage and language, they [still] fail to provide translations. [...] If I'm trying to follow along and I don't understand what's being spoken, it just feels difficult and makes it harder to access culture, especially when I'm already here in [North America] and not surrounded by community."* (P7).

While the examples recounted above show how the absence of multilingual captions has negatively impacted participants' access to community and culture, they also highlight the potential that thoughtfully designed captions have for fostering that access.

Multilingual Captioning as a Site of Linguistic and Cultural Preservation

Our participants were passionate about languages and actively sought out multilingual experiences. Their immense respect for how languages offered access to culture and community made them mindful of how language could be exploited for power. P5 shared, *"[Languages] are these really wonderful keys to, like, history and culture and identity. And as such have been politicized [... to] untether people from their, like, historical or ethnic or religious or cultural backgrounds and connections, like forced assimilation"* (P5). In light of this history, they cared deeply about ways to preserve language and culture, *"it's important to the people and also it's important to the language, to be able to maintain that presence and that sort of steadfastness in existence"* (P5).

Tracing these goals to captions, they discussed how captions often standardize variations in speech in the process of transcribing to written form (also seen in Section 4.5.2). P1

shared, *"Hindi is spoken so differently from different parts of the country and the world. [...] the sound of the word, the way they say it, I think that is sometimes lost when it comes to transcribing"* (P1). They highlighted how standardization could have far reaching consequences by erasing minoritized ways of speaking. They mentioned this had already happened with the use of standard dialects on broadcast media, *"everyone starts to speak like on the TV and radio."* (P11).

Nevertheless, participants highlighted ways multilingual captions could be a site of preservation. P11 discussed how more descriptive or phonetic captioning styles, that push the orthographic rules of the language to better match what is being said (or particularly, how it is pronounced), could preserve these linguistic and regional variations while facilitating access (e.g., with Quebec French transcribed in written Joual[4]) – *"Personally, being a descriptivist with language I think it is really important to keep the diversity."* (P11). Other kinds of multilingual captions, like transliterations and translations both bring tensions and opportunities. As we established in Section 4.5.3, translations may miss cultural context, and phonetic captioning styles like transliterations might be hard to parse. But participants did like the potential for linguistic preservation, especially in contexts like music or comedy, which were central to culture.

Besides Multilingual Captioning

In highlighting participants' current uses and visions of multilingual captions, we have demonstrated participants' hopes for the *outcomes*[5] of communication access: access to language learning, access to community, and preservation of cultural heritage. By bringing these hopes for outcomes of access to the forefront, we show that multilingual captions are only one of many tools for achieving these outcomes.

For example, some participants had doubts about whether captions could facilitate access

[4]https://www.thecanadianencyclopedia.ca/en/article/joual

[5]We draw this terminology from Crip Negativity: "What I'm really after is not the availability of access but rather the outcome of access: What are we hoping access will do for us?"[246].

to language. Limitations of captions as a medium, particularly for languages with mismatch between writing and speech – *"spoken Cantonese and written Cantonese are quite different"* (P13) – made participants hesitant about caption use. Similarly, P4 found that access to Portuguese was better facilitated through speechreading, *"[T]he thing is, my lipreading much better in Portuguese"* (P4) and thus did not use captions with family.

While multilingual captions *could* foster access to community, some participants did not think that the dearth of captioning had impeded their access to community. P3 had *"learned how to work in the world without a lot of accessibility. [...] Captions really are recent and new thing for me."* (P3). They were used to using other access practices like writing notes or asking for repeats. Other participants (P1, P2, P3, P5) sought community and cultural connections outside of the ones they were born into by learning sign languages like ASL. P2 remarked, *"I wish more hearing people would recognize the rich culture that there is within sign language and Deaf [communities]"* (P2). This adoption of sign languages indirectly facilitated preservation, given the history of oppression of sign languages amongst audist[6] world.

Thus, when we look at captions as a *"gateway"* (P7) to different activities and contexts, we can recognize that the use of captions is intentional and goal-driven. These intentions may change over time and context, and sometimes other access technologies and practices are better suited for reaching these goals.

4.6 Discussion

Our results highlight the many ways our participants navigate access in a multilingual world. We articulate notions of multilingual accessibility grounded in our participants' lived experiences – particularly, their language lives, which are often lost in prescriptive ideas of what communication should be.

Our analysis showcases the importance of integrating language perspectives into cap-

[6]Audism is a type of oppression directed at DHH people, privileging hearing senses and ways of communicating [27].

tioning research and examining the interplay between linguistic and sensorial access needs. We underscore the potential for language justice – a practice that examines how language can be a tool for oppression *and* equity – to inform our work in this space. Building from language justice, we call researchers and technologists (1) to attend to factors like fluency, linguistic affordances, and cultural perceptions unique to each language/experience; (2) to foster practices of multilingualism such as translation, transliteration, and translanguaging in captioning; and (3) to reflect on and address the systemic impact of (lack of) multilingual access on communities, culture, and languages.

Informed by these results, we begin by commenting on the state of speech recognition research and offer some immediate directions for development. Looking further down the horizon, we discuss how we might change our approach to captioning research, thus offering guiding principles and preliminary spaces for inquiry. We conclude by discussing how language justice could be valuable beyond captioning research.

4.6.1 Immediate Directions

Addressing the disparities in availability and quality of captions across languages requires significant advancements in speech recognition technologies. For English speech recognition, deployed technologies and state-of-the-art models achieve single digit WER (Word Error Rate [87], which roughly corresponds to the number of errors in transcription for every 100 words). In contrast, the performance of monolingual models of other languages varies widely (e.g., 13-25 for Hindi [35], 16-22 for Japanese [141], 13-30 for Arabic [7]). Efforts to train multilingual models also yield disparate performance – for example, Whisper AI ranges from 5 to 80 WER (e.g., 5 for English, German; 15 for Chinese, Korean; 30 Hebrew) [222]. While NLP researchers are working on innovating new techniques, collecting better multilingual datasets (e.g, [132]), and exploring new metrics (e.g., [105, 147]), it will be a while before these developments trickle down to users and technologies are readily available.

In the meantime, we believe improving caption form and display is a promising avenue of work. Many of the experiences discussed by our participants relate to professionally

captioned media (e.g., TV shows, movies) and user-generated content (e.g., YouTube videos, TikToks, Instagram reels). Our results point to several actionable directions for improvement (Section 4.5.3). Some could be implemented by revising captioning practices and increasing awareness, such as ensuring that the language is identified within the transcription in square brackets. Others require small changes to platform functionalities and leverage existing captions, such as allowing display of more than one caption at the same time (e.g., Korean on top, English at the bottom) and supporting selection of multiple captioning languages for multilingual media (i.e., videos with language switches). For these dual-track captions on multilingual media, we additionally recommend offering users more granularity in customizing font, size, and placement for different languages. Applying some automated approaches to transliteration could further allow for increased support for captions in multiple scripts (e.g., Simple and Traditional Chinese and Pinyin) and thus support users with different literacies. While these are low technical complexity, they would improve multilingual accessibility considerably.

4.6.2 Guiding Principles

In addition to NLP advancements and captioning design, we believe that to truly redress the imbalances we see across languages today, we require a fundamental shift in how we approach captioning research. Below we offer three guiding principles to reshape our ethos: centering language, decentering fluency, and recognizing fluidity of use. For each of these, we highlight how they inform our spaces for inquiry.

Centering Language

We argue language should be a crucial axis by which to design captioning technology, rather than an afterthought. English captioning has become the standard in HCI research, and an implicit assumption of generalizability reinforces the idea of "voice from nowhere" [236] – that there exists a neutral language and that language is English. In contrast, our work shows that captioning norms and practices are shaped by the linguistic, historical, legal, and

structural processes inherent to languages (Section 4.5.2), and that we need to bring these contexts to the forefront. Doing so not only recognizes the *situated* and *specific* nature of our knowledge, but also allows us to begin building tools that support typologically diverse languages and their users.

For example, our participants discussed linguistic affordances like spelling conventions for German or cadence for Portuguese that impacted accessibility of their experiences. P13 discussed how *"spoken and written Cantonese are very different"* (Section 4.5.2), impacting what information captions can encode and convey. Additionally, language and culture are closely entwined, and our study further shows how cultural norms influenced access practices and adoption of captioning – such as P12 who felt the need to be covert expressing their needs in Russian vs. P7 who felt captioning was a norm in China. These results point to the value of language-centric perspective.

How might this inform our spaces for inquiry? One direction for research is extending work on caption style, placement, error display to different languages and exploring how language characteristics impact the efficacy of current best practices. For example, how do unique visual affordances, like vertical reading order and logographic script impact the use of visual space while captioning? Another question rises for languages where writing is significantly different from speech (e.g., Cantonese) – what do accurate captions (and thus metrics) mean in this context? A systematic exploration of how listeners of these languages navigate discrepancies can help us design captions that better meet their access needs. Yet another direction could aim to improve usability of automated captioning in these languages through human-in-the-loop paradigms like crowdsourcing captions and correcting errors. These approaches have often been studied for English in Western contexts – exploring how language characteristics shape efficacy and cultural norms impact generalizability are open questions.

Decentering Fluency

In this work, we engage with participants' experiences in a range of languages – from those they grew up speaking to those they were just beginning to learn. As a result, we could recognize the role proficiency and notions of fluency played in participants' experiences with multilingual accessibility. This is a contrast to most research in our field, which often only engages with fluent (English) speakers as potential users of technology. We argue this implicit prioritization of fluency not only excludes non-native speakers of languages, but also ignores how fluency fluctuates over time. Migration, relationships, and globalization called our participants to pick up (and sometimes drop) languages in different periods of their lives. We call researchers to engage with these complexities of language use by decentering fluency.

Our participants discussed many experiences that occurred in less-fluent languages but were no less important to them – such as learning languages with their partner, watching TV shows with friends, and connecting to their heritage and families. Decentering fluency allowed us to unearth our participants' "subjective, emotional relationships with languages" [160]. They further highlighted how fluency conditioned their use of captions (e.g., familiarity with script and reading speed) thus impacting access. Importantly, our participants highlighted how fluency as an ideology encodes ableism – consider P13's quote in Section 4.5.2, *"if you want to be considered fluent in the language, you have to be able to hear it properly, but there are people who can't hear languages properly"* (P13). These assumptions discount any non-normative communication styles. This is echoed in P11's experiences with caption use in English, where she felt stigmatized as a 'native' speaker. By reducing expectations of fluency, communication and access provision became easier for our participants (such as P5's experience with access needs being lower in a language learning context).

How might this inform our spaces for inquiry? A promising direction for research is actively building technology that supports users with varying fluencies. For example, given the complex interplay between literacy and caption use, how might we redesign captions to support novice readers? Beyond readability, there is also a broader question of supporting

semantic understanding. We could extend research from audiovisual translation literature which largely focuses on analyzing captioners' choices in subtitling media (e.g., [227, 257, 17, 3, 55]) to actually engage with the experiences of audience members who use captions for accessibility and may be multilingual themselves. Similarly, the field of language acquisition could offer valuable perspectives in balancing the sometimes conflicting goals of language learning and access/understanding in the moment. We might also explore how practices of making language accessible in books and other text (e.g., including footnotes and rephrasing content) could apply to captions.

Recognizing Fluidity of Use

Much of this work aimed to highlight the complexities that rise from the multilingual fluidity of the real world. We find that this fluidity extends beyond language (i.e., switching between languages) to the use of access technologies and practices as well. Instead of constraining themselves to a particular access technology such as captioning (Section 4.5.4), participants leveraged different technologies and practices as needed. Their choices were guided not simply by the availability of captioning, but also by *how* it would support access and *what* it would facilitate access to: language, community, and cultural heritage (Section 4.5.4). We argue that recognizing this fluidity is necessary to designing technology that is useful in the real world.

Our participants perform complicated calculus while negotiating both language access and disability access in multilingual contexts since no one set of captions offers full linguistic and sensorial access. Consider a context with language switches – original language captions presume fluency, and translated captions presume the hearing ability and disappear with language changes (Section 4.5.3). Participants' practices of multilingualism, fluidly combining their linguistic repertoires for translation, transliteration, and translanguaging, demonstrate the need to to think beyond static, monolingual contexts, and recognize how multilingual captions may adopt to changing needs.

Recognizing fluidity in use also calls us to be expansive in our conceptions of users. While

prior work on captioning often focuses on the experiences of DHH people and evaluates the efficacy of transcribing audio to text, we opted to include perspectives of neurodivergent, chronically ill, and disabled individuals well. Doing so allowed us to surface unique ways captions supported access – by acting as a scaffolding for auditory processing, a focus tool, and a memory aid (Section 4.5.1).

How might this inform our spaces for inquiry? The different roles captions play in provisioning access for different users raises the opportunity to explore how these functions may inform design and evaluation (e.g., metrics) of captioning technologies. Systematically exploring what motivates use and non-use of captions amongst other access practices in a moment could deepen our understanding of communication accessibility for DHH, neurodivergent, and chronically ill individuals. Importantly, this perspective opens us to research that attends to the fluidity of access and explores what motivates an individual's choice of technology, modality, and language in a given moment.

4.6.3 Beyond Captioning

Through a deep exploration of multilingual captions, we have shown how language and disability shape each other and shape what access means. We invite researchers to explore these ties in broader accessibility research, beyond captioning.

In the last few years, accessibility research has been engaging with disability justice, a developing framework and activist movement led by disabled queer people of color. Disability justice principles have been used as a guiding lens to redesign captioning technologies, such as orienting hearing and DHH people to *collectively* attend to captioning errors [188]. They have also been used to subvert ableist structures in broader accessibility research [254, 280, 279, 118, 120]. We call researchers to incorporate language justice agendas as well, given how closely it aligns with disability justice. In fact, Sins Invalid, a foundational disability justice organization, argues that language justice *is* disability justice – both movements demand we move in mutual respect for all peoples, regardless of whether or how they communicate [244].

For example, consider research on DHH communication more broadly. Language justice allows us to engage with Deaf individuals who identify as minoritized language users, it centers the linguistic richness and cultural importance of sign languages independent of the access they afford [72], and it encourages to respect the multitude of ways d/DHH individuals sign or speak or gesture – the ways they understand and make themselves understood [113].

We believe there is value in taking a "language-first" perspective beyond captioning too. English also dominates research on web navigation [228], audio description [183], technologies for visual access [36, 269, 5], and alternative-and-augmentative communication devices [261] and more. Since languages may encode and describe color, direction, sound, and space differently, diversifying languages explored can deepen our understanding of practices for sensorial accessibility. We believe these differences across languages fundamentally shape how we communicate information and mediate accessibility interpersonally or technologically. Decentering fluency and recognizing fluidity of use is equally important in these broader contexts as well.

More broadly, we encourage researchers to document and investigate accessibility practices across languages and explore how multilingualism is transforming the design and use of accessibility technology. Given that practices of transliteration, translanguaging, and translation can be powerful ways to preserve languages, we call technologists and researchers to reimagine all of our systems to foster these practices.

4.7 Limitations

Getting a representative sample of as a large and dispersed population as "multilingual caption users" is difficult. As such, through recruitment we tried to capture experiences from people with diverse language and cultural backgrounds, hoping to articulate the complexity and nuances and space for future research. However, despite offering language interpretation, our research was limited to multilingual caption users who felt comfortable communicating in English or ASL, which means we did not include perspectives of a large number of individuals with valuable and potentially broader cultural insights. All participants were located in the

United States or Canada. Thus, our findings may not be applicable to other geographical contexts. Additionally, languages represented in our sample are mostly widely spoken. We invite future research to engage with languages at risk of disappearing due to colonialism and other oppressive structures (e.g., Hawaiian).

We conducted our study at an exciting time in the multilingual captioning world. Improvements in speech recognition technologies are resulting in deployment of automated captions in different languages to various platforms. For example, videoconferencing platforms such as Google Meet began adding non-English captioning in mid-2022. As of June 2024, Google Meet supports 87 languages [91], over half of which were added after we began data collection. Our results are situated at a point in time where for most users, lack of high-quality automated captioning support in their (non-English) languages is a reality. However, the shifting technological landscape may succeed in addressing this disparity in availability. This increasing availability of multilingual captioning presents an opportunity for future work to investigate individuals' adoption and long-term use of these technologies, and track their accessibility dimensions.

We attempted to include diverse disability perspectives on captioning. However, our sample was limited in size and thus may not capture the multitude of ways captions can support individuals' access needs, particularly those who are multiply disabled. As stated in Section 4.6.2, being expansive in our conception of caption users is integral. Multiply disabled caption users with unique combinations of access needs, such as those who identify as DeafDisabled or DeafBlind, could further point to ways we rethink captioning practices– for instance, by offering braille access to captions, word bubbling around caption text [247], access to complete transcript, or rapid serial visual presentation [85]. Exploring these could be a valuable space for future work.

Lastly, our interest in multilingual captioning for access meant we focused on spoken and written languages instead of signed languages, as most do not have a standard written form. Exploring individuals' use of translated captions with sign languages (e.g., English with ASL) could be insightful. Additionally, understanding how individuals navigate interactions with

translanguaging between different signed languages, and signed and spoken languages is an interesting avenue of future research.

4.8 Conclusion

We live in a multilingual world. Disabled people speak various languages, and experience varying levels of access needs across different languages. Yet, HCI and accessibility research has implicitly or explicitly focused on English in designing technologies like captioning. Our research contributes a language and disability justice agenda to decenter English, and celebrate all ways of communication. Through stories of 13 multilingual individuals who use captions for accessibility, we show how disabled people navigate multilingual access needs, the roles that multilingual captions play in their lives, and their visions for the future. Overall, our analysis allows us to situate multilingual captioning within broader social, cultural, and linguistic structures. By centering language, decentering fluency, and recognizing fluidity of use, we can orient communication accessibility research toward language and disability justice.

Chapter 5

DEVELOPING SIGN LANGUAGE DICTIONARIES

Having explored deployed technologies and communication experiences in existing digital environments, my next two chapters shift focus to emerging technologies – studying how we might center both language and disability needs while building technologies from the ground up.

In this chapter, I focus on sign language dictionaries. Dictionaries are inherently about language access – in how they support learners of a language as well as native users and the ways they document a language. With sign languages, the visual nature additionally calls us to explore information retrieval paradigms that align with these modalities. Sign language dictionaries therefore require we take both disability and language perspectives – a disability lens to ensure the search is accessible, and a language lens to develop approaches that reflect realities of language use and change. Given the diversity of signing practices in the DHH community and evolving vocabularies, sustainable and scalable sign language dictionaries can help us document and preserve minoritized signing practices.

In terms of methods, this work takes a technical approach, exploring how we might train sign language models that are grounded in the values of disability and language access. The results showcase that aligning modeling approaches to these values is technically generative – not only resulting in practical insights for dictionary maintenance, but also illuminating new directions for ML methods inquiry. This work has been previously published at EMNLP 2025 "Investigating Dictionary Expansion for Video-Based Sign Language Dictionaries", coauthored with Daniela Masicetti, Richard Ladner, Hal Daumé III, Danielle Bragg, and Alex Lu [74].

5.1 Motivation

Dictionaries are important resources for sign languages, offering a way to document the many different signs comprising the intricate vocabularies of these languages. In addition to documentation, dictionaries are particularly helpful to novices or language learners, allowing them to easily look up signs they are unfamiliar with. It is crucial that these dictionaries have mechanisms to stay up-to-date with changes in the language, such as the creation and adoption of new signs by signing communities. Here we consider the problem of dictionary expansion, where the vocabulary of an established sign language dictionary is updated to incorporate new signs.

A key factor that makes sign language dictionaries (and thereby their expansion) unique is that sign languages are visual-manual languages. This means that sign language dictionaries typically use video entries to represent each sign in their vocabulary, and need video-based approaches for users to look up signs and query the dictionary. In a video-based dictionary, a user demonstrates a sign to a camera, and the dictionary returns a ranked list of entries from the dictionary that might correspond to that sign. Recent advancements in sign language datasets and modeling show exciting promise in making video-based dictionary retrieval a reality [111]. For example, many state-of-the-art isolated sign language recognition models (which are trained to recognize single signs from vocabularies of 2k+ signs) currently achieve recall@10 above 90% (i.e., queried sign is in top-10 results) – this high performance makes it possible to deploy these models for dictionary retrieval.

However, a key limitation of almost all such modeling approaches is that it is unclear how they might support dictionary expansion. The technologies underlying most video-based retrieval methods are often machine learning classifier models trained on a fixed vocabulary, where each sign corresponds to a label in a fixed label set. Incorporating new vocabulary into these models would typically require retraining them from scratch – however, this is can be computationally expensive and relies on the availability of a sufficient amount of high-quality training data, which may be difficult in practice. It also means that third-parties

can't adapt existing dictionaries to their needs (e.g., representing local dialects or specialized lexicons). An ideal approach would support vocabulary expansion without retraining the existing model, but this class of approach remains largely unexplored.

In this work, we investigate the feasibility of dictionary expansion without retraining for video-based sign language dictionaries. To overcome fixed vocabulary constraints encountered with current classification approaches, we propose a simple method that instead uses the representations learned by deep learning models for dictionary retrieval, rather than just their final classification layer. Doing so allows us to work with unseen signs through nearest neighbor-type approaches.

We explore how this type of approach performs across a range of circumstances, simulating dictionary expansion in ideal settings as well as more realistic and challenging scenarios. We separately evaluate the performance of expanded models on core signs (those in the original dictionary) as well as newly added signs. We find that our method performs well when adding a small number of new signs ($\sim 1 - 100$), each with many examples (~ 15 per sign), to a dictionary with an existing large vocabulary – maintaining performance compared to the pre-expansion for both new and old signs. However, performance degrades substantially when we constrain the number of examples for new signs, or attempt to add many new signs to the vocabulary, which we argue are likely scenarios for real-world dictionary expansion. Our work is the first to systematically explore computational challenges dictionary expansion, providing directions for future methods development grounded in real-world needs of sign language dictionaries and their users. These opportunities for future research not only offer a space for technical discovery, but also further the exploration of sign languages as languages [289, 72].

5.2 Background

Sign Languages and Dictionaries. Sign languages are visual languages, with each sign composed of distinct handshapes, movements, non-manual markers, and other phonological features. There are over 300 sign languages in the world. Our work focuses on American Sign

Language (ASL), which is the most common in North America. ASL is culturally significant to Deaf community in the continent, and is also learned by many as a second language [173]. ASL dictionaries are a valuable resource for this group.

Part of what makes dictionary retrieval in sign language challenging is that two different signs might share nearly all the same phonological features (which can cause them to look visually similar) but have very different meanings. For example in ASL, SORRY and PLEASE only differ in handshape, SUNRISE and SUNSET only differ in movement (examples of *minimal pairs*). Dictionary retrieval methods need to be robust to such dense lexical neighborhoods.

Existing sign language dictionaries support querying in English, through a set of descriptive features, or by video demonstration. Searching by English word or gloss (e.g., SigningSavvy, LifePrint) is a valuable approach for those looking to learn a sign from an English word (i.e., English-to-ASL translation). However, users cannot leverage these dictionaries to look up the meaning of an unfamiliar sign (i.e., ASL-to-English). Allowing users to navigate dictionary search directly in sign languages is complicated as sign languages do not have standardized written forms. One approach is feature-based search (e.g., HandSpeak & [42]) that allows users to search by describing features of a sign (e.g., handshapes, movements). Video-based search allows users to search by demonstrating a sign (returning all dictionary entries that may match the search video). Video-based dictionaries also allow native signers to navigate dictionaries in their primary language– their development is an important avenue of work.

Video-based Dictionaries and Sign Language Recognition. Recent advancements in datasets and modeling have changed the landscape of sign language recognition research. Consider the release of large-scale isolated sign datasets for ASL (e.g., ASL Citizen [71], Semlex [145], PopSign [252], WLASL [168], ASLLVD [18]), each of which contains large vocabularies (over 2000 signs) and multiple samples per sign from different contributors. It is now feasible to use deep learning to train models for single sign recognition, and thus build

video-based sign language dictionaries [111]. While many have worked to surpass dictionary retrieval performance on these benchmark datasets (e.g., [101, 282]), most of these methods are limited by approaching sign language dictionary lookup as a classification problem – thus making it hard to adapt to changing vocabularies. In this work, we propose a basic method for leveraging these classifiers as dictionaries expand their vocabulary, and demonstrate where this method succeeds and where challenges still remain.

Prior Work on Dictionary Expansion. While some work has explored how sign language recognition models might generalize to unseen data (such as different datasets and different languages, e.g., [282]), only few have focused on the task of dictionary expansion. Huamani *et al.* [121] compares different incremental learning approaches for Peruvian Sign Language dictionaries, and Gupta *et al.* [104] explores the same for Indian Sign Language. Both works consider small vocabulary contexts (under 100 signs) under ideal data conditions (multiple examples of the new vocabulary), which might not reflect real-world conditions with larger vocabularies, larger expansion demands and variable amount of examples. While some researchers have explored sign language recognition with data constraints (e.g., limited data for all signs [39, 267] or unequal data across signs [145] or demographics [21]), they again focus on fixed vocabulary contexts. In this work, we look at the combination of the two contexts i.e., expanding vocabularies *and* data-constrained recognition as this is the most reflective of real-world use of sign language recognition for dictionaries.

5.3 Experimental Setup

5.3.1 Task Definition

We consider the task of video-based search in sign language dictionaries. For dictionary retrieval, we aim to map user-submitted video queries $\mathbf{x}$ to glosses[1] y. We assume have access to a pre-trained classifier f that maps videos to a fixed number of glosses N: $f : \mathbf{x} \mapsto$

[1]English translations for isolated signs

$\{1, \ldots, N\}$, outputting probability for each gloss in the vocabulary of the dictionary. These probabilities can be ranked and used to retrieve a list of likely matching signs for the user (e.g. the highest probability gloss is the top-ranked retrieval result).

After training this classifier, we discover that M new signs have become commonplace and we wish to add them to our classifier, to yield $f' : \mathbf{x} \mapsto \{1, \ldots, N + M\}$. We wish to expand the dictionary without having to retrain f, as computational resources or time for retraining are limited. To facilitate the expansion of f to f', we assume that we have access to varying number (m) of videos for each of the M new signs (e.g., through crowdsourcing contributions).

Given this new model f', we care about how well it does on *both* the original N signs (i.e., we do not want its performance to degrade dramatically on the old signs in comparison to f), *and* the new M signs (i.e., we want it to correctly recognize the new signs).

5.3.2 Our Dictionary Expansion Approach

In this work, we propose a method to support dynamic vocabulary expansion of f to f' using learned feature representations. Deep learning models, particularly those trained for classification like f, learn rich intermediate representations that capture semantic and visual structure in the data. These representations (typically extracted from the penultimate layer) can often generalize beyond the specific labels used during training. We adopt a similarity-based retrieval approach using these feature representations. Given a query video of a new sign, we extract its feature representation using f and retrieve the most similar video from a database using a nearest-neighbor search in the feature space. This database consists of feature representations from both the core vocabulary and an expansion set of new signs.

In practice, to expand the dictionary, an administrator would simply need to process new sign videos through the trained (frozen) classifier to extract their feature representations and add them to the retrieval database. No additional training or fine-tuning is required. This provides a lightweight, scalable solution for incorporating new signs into a fixed-vocabulary dictionary by leveraging the generalization capacity of learned embeddings and similarity-

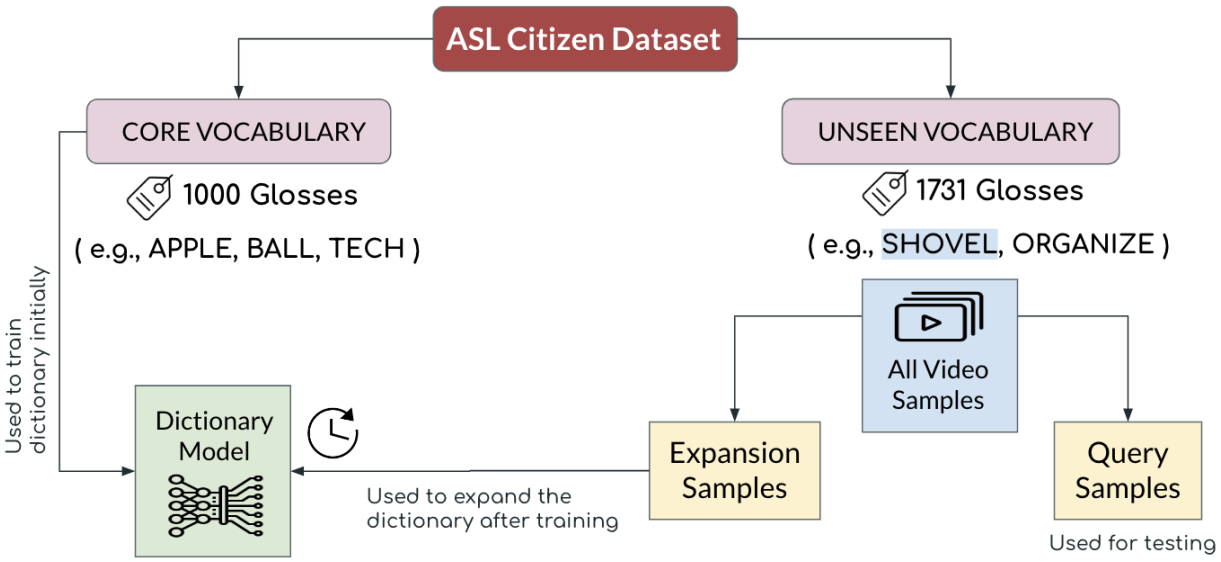


Figure 5.1: Schematic of Data Setup for Dictionary Expansion

based retrieval. Below, we discuss how we simulate a variety of realistic dictionary expansion scenarios, carefully slicing a large ASL dictionary dataset into appropriate subsets.

5.3.3 Data Setup

We use the ASL Citizen dataset [71] as it was collected to support research and development of ASL dictionaries. This dataset contains about 84k videos of 52 d/Deaf and hard-of-hearing contributors fluent in ASL performing single signs. It also contains videos of the seed signer (a highly proficient ASL signer) whose videos were used to prompt data collection. Each video is labeled as 1 of 2731 glosses. The dataset provides standardized train, validation, and test splits by contributor, allowing for testing of model generalization onto unseen users, mimicking real-world use.

To simulate dictionary expansion, we split the ASL Citizen dataset by gloss into two non-overlapping subsets: a "core vocabulary" and an "unseen vocabulary". The core vocabulary is intended to simulate entries that are initially present in a dictionary. We assume that

the dictionary administrators have already trained a classifier, using the training dataset associated with the core vocabulary. Held-out test recordings of the core vocabulary can then be used for the purpose of simulating a dictionary query.

The unseen vocabulary are glosses that we reserve to test dictionary expansion under various scenarios. The unseen vocabulary is intended to simulate new signs that will be added to the dictionary at a future date, after machine learning models have already been trained on the core vocabulary. We further split the unseen vocabulary into expansion and query samples. We consider the expansion dataset to be examples the dictionary administrator is in possession of, and can use to expand their machine learning model. The query dataset is then intended to simulate users querying the expanded model, and used to test performance of the expanded models.

Mindful of how the composition of the vocabulary might influence task difficulty, we randomly generated three different core-unseen splits for our dictionary expansion experiments[2]. We repeat our experiments with each split and report average performance and standard deviation across these runs. We note that we did not segment the core and unseen vocabulary by contributor, meaning that signers in the unseen vocabulary may have already been seen in the training dataset of the core vocabulary. We consider this to still be reflective of real-world usage as crowdsourced dictionaries (like ASL Citizen) often sustain repeat contributors.

5.3.4 Dictionary Settings

To understand how the difficulty of dictionary expansion interacts with various data acquisition challenges in the real-world, we begin by declaring a base setting for dictionary expansion. We then vary parameters of this base set-up in controlled ways to simulate these various data challenges.

---

[2]Splits can be found here: `https://github.com/aashakadesai/asl-dict-expansion-splits/`

Base setup In this set-up, we assume we begin with a reasonably large core vocabulary with many examples per sign. We then simulate a scenario where only one new sign is added to the vocabulary, with ample expansion samples for this sign. In our case, our core vocabulary consists of 1,000 signs and total of 14,691 training videos (i.e., ∼15 videos per sign, each performed by different contributors). We then expand the vocabulary by 1 sign (going from 1000 to 1,001 signs) using all available expansion samples for that sign (i.e., ∼15 videos per sign, each performed by different contributors). We repeat this procedure over all 1,731 signs in our unseen vocabulary and report averaged metrics (see section 3.6), testing on a total of 20,902 videos.

Adding a Sign with a Single Example Gathering multiple samples for a new sign involves considerable effort. For example, for dictionaries that rely upon crowdsourcing, newly added vocabulary may be recently invented signs that have not fully disseminated through the community yet. In these cases, only a small number of contributors may initially provide videos of the sign. To simulate this scenario, our core vocabulary consists of 1,000 signs (with ∼15 videos per sign, each performed by different contributors). We then expand the vocabulary by 1 sign (going from 1000 to 1,001 signs) using *only one expansion sample for that sign* (in contrast to ∼15 samples used in the base set up). We consider two settings for how the single sample per sign is obtained: In the first setting, we assume that signs are organically contributed by different contributors. We simulate this by randomly sampling a video from the unseen expansion set. In the second setting, we assume a long-term contributor (i.e., someone who has likely recorded examples of most of the core vocabulary) contributes a new sign to the dictionary. We simulate this by using a specific contributor's video to represent core and added signs throughout the dictionary.

Adding Multiple Signs with Multiple Examples In the real world, it is likely that multiple signs will be added to a dictionary. Dictionaries are expected to grow iteratively with contributions over time. This larger expansion context might have a more significant

impact on dictionary performance for the core vocabulary. To simulate this scenario, our core vocabulary again consists of 1,000 signs and newly added signs have 15 expansion samples each. Unlike the previous two settings, instead of adding one sign to the dictionary, we sample *a set of signs* (n = 100, 500, 750, or 1000) to add to the dictionary. This allows us to test expanding the vocabulary by 10%, 50%, 75% and 100% respectively. For each stage of expansion, we randomly sample three different sets of signs from the unseen split to be added to the dictionary, as differently composed sets might impact the difficulty of the task uniquely. We report the average performance across these splits as well as corresponding changes in performance of core vocabulary after expansion.

Adding Multiple Signs with Few Examples Next, we were curious about the interaction between adding multiple signs and having limited expansion samples for a new sign on dictionary expansion – a very likely real-world scenario. To simulate this, we used used aspects from each of the above settings: Our core vocabulary consists of 1,000 signs (with ∼15 videos per sign). We then test expanding the the vocabulary by 10%, 50%, 75% and 100% (i.e., adding sets of 100, 500, 750, and 1000 signs respectively). But in this setting, newly added signs have only one expansion sample each. We again consider two settings for how the single sample per sign is obtained: In the first setting, we assume that signs are organically contributed by different contributors. We simulate this by randomly sampling a video from the unseen expansion set. In the second setting, we assume a long-term or hired contributor adds new signs to a dictionary. We simulate this by using the seed signer's video to represent core and added signs throughout the dictionary.

Exploring Smaller Dictionaries Certain sign language dictionaries may be more specialized or in early stages of documenting a lexicon, and thus smaller in vocabulary. To simulate this scenario, we decided to create three new core-unseen splits (100 and 2631 signs respectively) for our experiments. Each of these splits is uniquely composed to reflect various underlying rationales of smaller dictionaries. First, naively, we randomly sampled set

of glosses. Second, reflecting a dictionary that might be aiming to be comprehensive but is newer, we sampled a set of semantically distributed glosses (details in Appendix A.6). Third, to reflect a more specialized dictionary (such as one targeted towards language learners), we sampled a set of glosses matching that use case (from HandSpeak's list of first 100 signs for ASL learners). For each of these splits, our core vocabulary has 100 signs and total of ∼1490 training videos (i.e., ∼15 videos per sign, each performed by different contributors). We first test performance adding a single sign with multiple examples (i.e., newly added signs have ∼15 samples). Then we test adding multiple signs to the dictionary– exploring adding sets of 10, 50, 75, and 100 signs to the dictionary, following the same sampling approach as the large dictionary setting.

5.3.5 Model Architecture and Training

We use the spatiotemporal graph convolutional network (ST-GCN) [287], following previous works on this dataset. We extract keypoints using MediaPipe Holistic, and use the same 27 keypoints outlined in OpenHands [241]. We preprocess the keypoints using the same procedures outlined in prior work [241, 71]. The model is trained on the core vocabulary for 100 epochs using a learning rate 1e-3, an Adam optimizer and a Cosine Annealing scheduler. We select the best performing checkpoint on the core validation set for our expansion experiments. To extract feature embeddings for our experiments, we use the encoder of the ST-GCN model[3]. We use cosine distance to calculate similarity between embeddings in nearest neighbor search for dictionary retrieval.

5.3.6 Metrics

We use metrics consistent with prior work on sign language dictionary retrieval to evaluate the returned ranked list of glosses: Discounted Cumulative Gain (DCG, which evaluates overall ranking of the correct sign in the list) and recall-at-k (which considers if the correct

[3]as the decoder is a single fully connected layer this corresponds to the second-to-last layer of the model

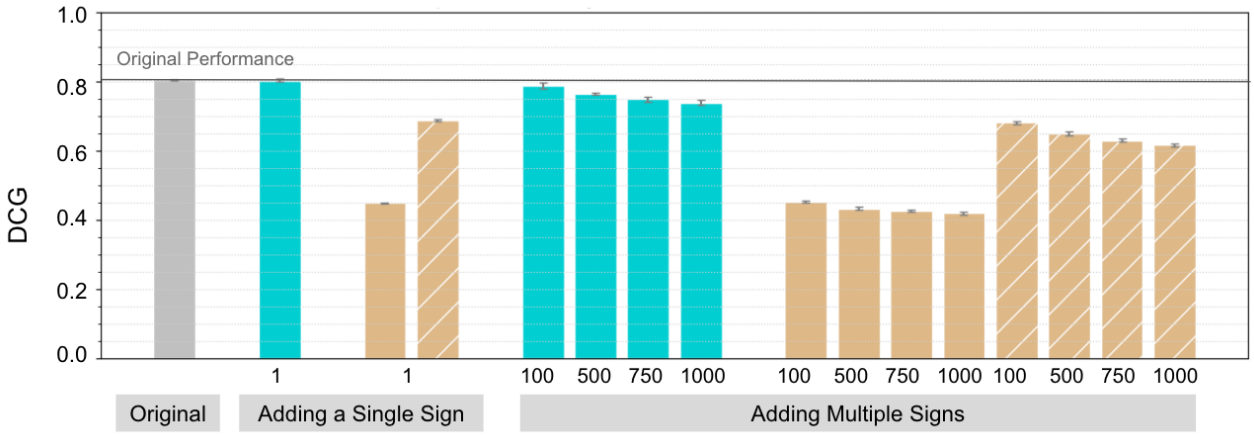


Figure 5.2: Performance of added vocabulary across different dictionary expansion scenarios. Bars are clustered by scenario and represent DCGs of added signs using centroid representations- scenarios with multiple samples (blue), one random contributor example (solid gold), one long-term contributor example (striped gold).

is in the top-k rankings).

5.4 Results

Overall, we find that the performance of our dictionary expansion method varies greatly across different dictionary settings – Figure 5.2 provides a snapshot of our results. In the following subsections, we walk through the results in each setting and our corresponding analysis. The ranking of settings was consistent across all metrics so we report DCG, but full tables reporting each metric for each experiment can be found in the appendix.

5.4.1 Core Dictionary Performance

We provide a baseline to contextualize dictionary expansion by first measuring the retrieval performance of seen signs in the core vocabulary only, with no dictionary expansion performed (Table A.2 in Appendix A.2). The classifier trained on the core vocabulary achieves a DCG of 0.7787 and top-1 accuracy of 62.21% when tested on the test split of the core vo-

cabulary (1000 glosses, ∼12038 videos). We then validate our feature representation-based retrieval method: it achieves a DCG of 0.7719 and top-1 accuracy of 61.11% on the same test split of the core vocabulary. Comparing these two approaches, we see that the feature-representation approach performs as well as the classifier on seen (i.e., core) signs. We were further able to improve performance of the feature representation approach by calculating the centroid of all sample videos for each sign and using this instead in our retrieval database of feature representations. Interestingly, this strategy even surpasses the classifier, with a DCG of 0.8050 and top-1 accuracy of 66.04%. Thus, we retrieve the centroid representation of all core and expanded signs for all subsequent experiments.

5.4.2 Base Setup Performance

Having established that using extracted features from a classifier is an effective strategy for dictionary retrieval, we next focused on testing a baseline dictionary expansion scenario that we consider the most ideal setting for dictionary expansion. In this set-up, we add a single sign, with expansion samples equal to the maximum number in the ASL Citizen training dataset for that sign (typically 15).

When we have multiple examples for each sign, we find that the performance of newly added signs is comparable to that of the core vocabulary (Table 5.1). Specifically, we see a DCG of 0.8042 and top-1 accuracy of 65.38% on the new vocabulary (compared to a DCG of 0.8050 and top-1 accuracy of 66.04% for the core vocabulary). This suggests that our feature representation method could be effective in this dictionary expansion setting.

5.4.3 Adding a Sign with One Example

When adding a new sign with just a single example, however, we see a significant degradation in performance (Figure 5.2). When this sample is randomly selected, the newly added signs achieve a DCG of 0.4492 and top-1 accuracy of 21.74% (Table 5.1) – a large drop compared to a DCG of 0.8042 and top-1 accuracy of 65.38% in the base setup with multiple examples per sign.

Setting	DCG	Rec@1	Rec@10
Many Examples	0.8042 ± 0.0037	0.6538 ± 0.0047	0.9105 ± 0.0040
One Random Contributor Example	0.4492 ± 0.0016	0.2174 ± 0.0041	0.5396 ± 0.0056
One Long-term Contributor Example	0.6881 ± 0.0025	0.4955 ± 0.0006	0.81205 ± 0.0070

Table 5.1: Dictionary expansion performance when adding a single sign to the dictionary under a variety of data settings. We can see that amount of data and choice of sample (in low-data settings) are a significant factor in performance.

This demonstrates that data plays a critical role, and we hypothesize that constraining ourselves to a single example video for a sign limits our method's ability to provide a robust representation for retrieval with the new sign. The feature representation of a given video likely contains both information about the sign along with variation specific to a given contributor – we hypothesized that taking the centroid across many samples "averages out" the contributor-specific variation, providing a more robust representation of the sign itself.

Based upon this hypothesis, we reasoned that using a single contributor for all videos in the retrieval database, as opposed to different contributors, may reduce the impact of contributor-specific variation. We experimented with using a specific contributor's videos as expansion samples for the dictionary. We reasoned this would simulate a scenario where a long-term contributor to the dictionary has recorded many, if not all, signs in the vocabulary in addition to the added sign. A benefit of this is that we can then use this signer's videos to consistently represent all signs in the dictionary, both core and added – and thus minimize impact of user variation. We tested this approach with videos of five different contributors from the ASL Citizen dataset: the seed signer (P52), P33, P11, P50, P27 – all of whom had recorded videos for (almost all) glosses in the ASL Citizen dataset. We find that dictionary retrieval performance for added signs improves significantly with this single signer set up – in the best case (seed signer), achieving a DCG of 0.6881 and top-1 accuracy of 49.55%

after using only one expansion sample per added sign (Table 5.1). However, we also note that performance varies drastically across contributors: e.g., 24.81% top-1 accuracy using P11's videos vs 42.15% using P27's videos (results for each signer can be found table A.3 in Appendix A.3).

5.4.4 Adding Many Signs with Many Examples

Having explored expanding the vocabulary of a dictionary by one sign (i.e., going from a vocabulary of 1000 to 1001), we next studied the impact of larger scale expansion that involves adding multiple signs. The left half of Figure 5.4 and Table A.4 summarize performance of core and added signs at different stages of expansion (adding 100, 500, 750, and 1000 signs).

We observed that as more signs are added, dictionary performance degrades. However, even in the largest expansion scenario we tested, dictionary retrieval performance overall remains reasonable for core and added signs at each stage of expansion (Table A.4). When adding 100 signs, new signs achieve a DCG of 0.7984 and a top-1 accuracy of 64.33% on average – a slight drop in performance compared to the baseline expansion setup. At this tier, the impact on core vocabulary is also minimal, dropping to a DCG of 0.7968 and top-1 accuracy of 64.84% compared to pre-expansion. We find that performance continues to drop for both core and added vocabulary as we add more signs to the dictionary. The last expansion tier, adding 1000 signs, achieves a DCG of 0.7397 and top-1 accuracy of 56.03% for new signs and a DCG of 0.7436 and top-1 accuracy of 57.27% for core vocabulary. This tier corresponds to effectively doubling the vocabulary of the dictionary and shows a $\sim$9 point drop in top-1 accuracy and a 0.06 point drop in DCG. While this is not insignificant, we note that top-10 accuracy remains above 85% post-expansion for both core and added signs, meaning that users of a dictionary will continue to find the desired sign among the first 10 entries. This suggests that expanded dictionaries generally maintain similar levels of usability in our simulated setting, even up to 1000 added signs.

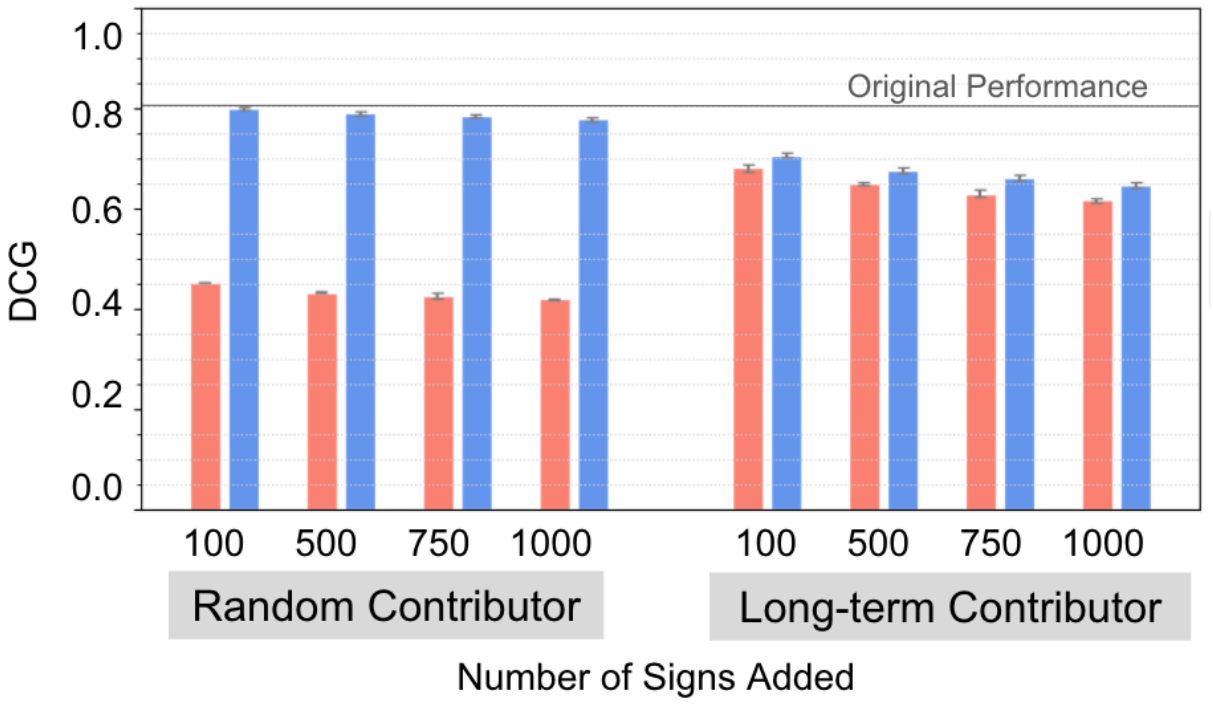


Figure 5.3: Impact of chosen expansion sample when adding multiple signs with one example. Reporting DCG of added signs (red) and core signs (blue).

5.4.5 Adding Many Signs with One Example

We next assessed the interaction between adding multiple signs to a dictionary and limited number of expansion samples per sign. Figure 5.3 summarizes our results multiple stages of expansion and across two different setups– using a random contributor's video for each sign vs. using a specific, long-term contributor's video for all signs.

For both setups, we note a similar overarching trend that performance degrades as we proceed to larger expansion tiers i.e., as we add more signs to the vocabulary. However, we find that performance for core and added signs varies drastically in each setup (Figure 5.3 and Appendix A.5). When using the random contributor's video for added signs (alongside centroid for core signs), we find large disparities in performance between added signs and core signs – added signs achieve a DCG of ∼0.4 and top-1 accuracy of ∼20% in contrast to core

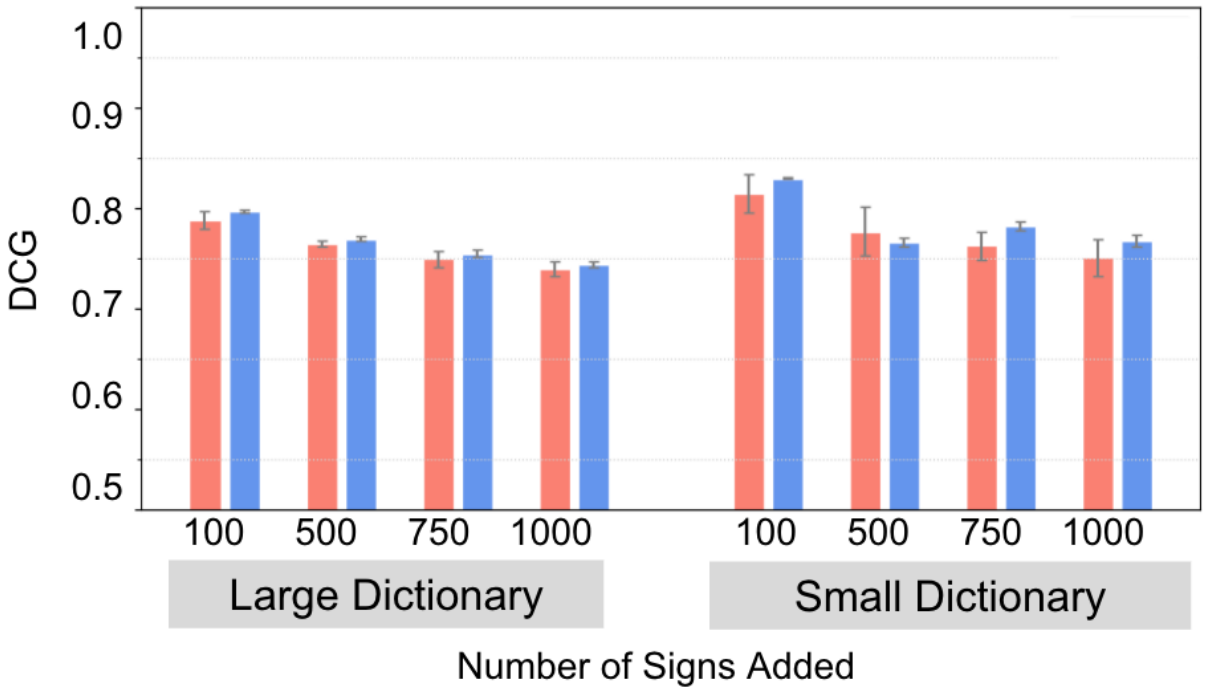


Figure 5.4: Comparing trends in dictionary expansion across large and small dictionaries.Reporting DCG of added signs (red) and core signs (blue).

signs with a DCG of ∼0.75 and top-1 accuracy of ∼62% across different stages of expansion (Table A.5). When using a long-term contributor's videos, we again note the benefit of reducing user-variation in the retrieval space (Table A.6)– added signs achieve a much better performance: a DCG of ∼0.64 and a top-1 accuracy of ∼44%. We note that there are less disparities between core and added signs in this setup; however, we see a bigger drop in core vocabulary performance compared to pre-expansion: achieving a DCG of ∼0.65 and a top-1 accuracy of ∼48% (Table A.6). This is likely due to the fact that the seed signer's video does not approximate the latent representation of the sign as well as calculating centroid across multiple users.

5.4.6 Small Dictionaries

We were curious if the trends we see in dictionary expansion experiments would hold for dictionaries with smaller core vocabularies (such as new or emerging dictionaries). We tested this scenario using models trained on a vocabulary of 100 signs with multiple examples per sign. We then tested expansion using multiple examples per sign.

When adding a single sign to the vocabulary, we find that similar to our results with the larger dictionary, the performance of the added signs is comparable to that of the core vocabulary (first row of tables in Appendix A.6). When adding multiple signs, this trend continues (Figure 5.4)– we find small gaps between core vocabulary performance and added vocabulary performance at each stage of expansion. However, for the smaller dictionary, we note much larger variations in performance across different core-unseen splits (as evinced by large deviation bars in Figure 5.4). For example, with the randomly generated split, adding 10 signs to the vocabulary leads to a 9 point drop in top-1 accuracy for added signs (going from 67.22 to 58.21) (Table A.7); whereas the semantically distributed split has almost no drop (Table A.8) and the HandSpeak split has 3 point drop (Table A.9). We believe this suggests that the lexical composition of the vocabulary (both core and added) is a much more significant factor for small dictionary expansion. Lastly, we note that performance degrades faster with each expansion tier for the small dictionaries. We hypothesize that this is because small dictionaries have not been exposed to sufficient data during training to learn good/robust feature representations for signs.

5.4.7 Visual and Runtime Analysis

Our representation-based method assumes that signs that are visually similar are placed closer to each other in the embedding space. To evaluate this, we examined the visual similarity of errors made by our method (i.e., confusions) in the base setup. We used phonological labels from ASL-Lex [238] to identify key parameters (Handshape, Movement, Location) for each sign in our vocabulary and then counted how many of these features differed between

confused signs. We find that on average 1.85 features differ between confused signs. In contrast, when sampling 1000 random pairs of signs from the lexicon, 2.38 features differ on average. This indicates that the confusions made by the model are between signs with fewer differences in phonology (i.e., more visually similar signs). Appendix A.7 provides some examples of common confusions.

Finally, given the motivation around practical deployment, we calculated the time it takes for retrieval as the dictionary expands. As expected with nearest-neighbor search, retrieval time increases linearly with the size of the vocabulary: 10.12ms for 1000 signs vs. 16.08ms for 1500 signs vs. 21.47ms for 2000 signs. At the current vocabulary sizes, we anticipate this increase in time would not significantly impact user experience. However, as the dictionary grows further, implementing techniques for speeding up nearest neighbor search (e.g., using Meta's FAISS package) would be valuable.

5.5 Discussion

In this work, we explored sign language dictionary expansion across a variety of settings. Our results highlight the feasibility of expanding sign language dictionaries using feature representation-based approach, and demonstrate that performance is contingent on the type and quantity of data available for new vocabulary.

Given enough examples of a sign, we show that incorporating it into an existing dictionary is quite feasible. This is valuable for when small changes need to be made to a dictionary quickly (e.g., recording a new variation or documenting a newly emerging and rapidly adopted sign, like the sign for 'coronavirus').

When adding multiple signs with multiple examples, we show that models are generally robust, albeit at a slightly lower performance than when just a single sign is added. This can inform the frequency at which we retrain models in response to changing vocabularies. A comparison of numbers from the small dictionary experiments (100 core signs) to the large dictionary experiments (1000 core signs) suggest models may need to be retrained at more frequent intervals initially, but less often once they've develop robust representations.

On the other hand, we find that performance degrades when added new signs that have only a limited number of samples. As we would expect, this is the most difficult, but also the most realistic, expansion scenario. This illuminates an important direction for future research: examining new methods for incorporating signs into a dictionary in data-constrained scenarios. We focused on classifiers trained using deep learning in this work; however, alternative training approaches (e.g., meta learning) may result in more robust feature representations for core and added signs. We encourage researchers to take up this line of inquiry.

Looking forward, we emphasize the importance of addressing dictionary expansion for sign languages. Compared to written languages, existing sign language dictionaries represent only a small fraction of the true language lexicon. While many crowdsourcing and documenting initiatives are underway, gathering a snapshot of sign languages that is representative of the many different regional and contextual variations continues to be a challenge. In addition to variations in everyday vernacular, different fields of study (such as STEM disciplines) frequently produce new signs to represent new concepts. Dictionary expansion then offers a crucial pathway to building sustainable and scalable sign language dictionaries.

Our results also offer interesting implications for the maintenance of current crowdsourced dictionaries. Examples available for each sign are expected to grow over time with contributions from users. A representation-based approach like ours (that works with any number of signs and any number of examples per sign) aligns well to the fluctuating vocabulary size and data availability encountered in these dictionaries. Initially, dictionary administrators could work to incorporate a sign with any available sample, and then switch to centroid for these added signs as more examples are collected. While this results in poor added sign performance initially, it allows the sign to be incorporated into the video-based dictionary and solicit further interaction and contributions from dictionary users. To maximize initial performance, dictionary administrators could also work to sustain long-term contributors to record most, if not all, signs being added to the dictionary (as this shows best added sign performance with limited data).

Overall, our work outlines directions for sign language research aligned with community needs and real world settings and furthers the exploration of sign languages as languages in their own right.

5.6 Limitations and Future Work

In this work, we reported results on dictionary expansion using only an ST-GCN model. While this architecture is frequently used in sign language recognition and has the benefit of being lightweight, it is not the only established approach. Appearance-based approaches to sign language recognition (like I3D) might fare differently under expansion scenarios. Exploring dictionary expansion across a variety of architectures as well as robustness of their learned feature representations could be an interesting direction for future research.

To simulate dictionary expansion, we created core-unseen splits from ASL Citizen. This meant that many of the contributors in the unseen split had previously been seen by the model – while this is reflective of a real-world scenario (long-term dictionary contributors), using a completely new dataset for unseen vocabulary would have also allowed us to simulate a dictionary setting with a completely new contributor scenario for added signs.

In our results, we note that the centroid approach outperforms the classifier even in the base dictionary setting – to the best of our knowledge, this has not been discussed in prior literature. While valuable, a core limitation of this approach is that the centroid representation requires multiple training examples to become reliable, which may be difficult leverage with data scarcity.

Our analysis surfaces large variance in dictionary retrieval performance across contributors (section 5.4.3), we do not investigate these disparities in depth. We hypothesize these disparities could be explained by variations in video quality for these different signers or biases in the underlying model (or both). Disentangling these aspects and more systematically exploring dictionary expansion performance across different contributors is valuable direction for future work.

We also highlight differences in dictionary expansion trends at smaller vs. larger core

vocabulary sizes. However, our results are limited to comparing two specific vocabulary sizes (100 vs. 1000) – a more systematic investigation across different vocabulary sizes may have better unearthed overarching trends. Additionally, with our smaller vocabularies we note sensitivity to lexical composition, but our analysis reports aggregates which prevents us from investigating sign-level nuances and conducting a linguistically-informed analysis.

Lastly, we position our work related to evolution of sign languages, but only focus on adding signs to dictionaries and not replacing or removing signs. While our method could be easily adapted to address these, further research is required to investigate impacts on overall dictionary retrieval performance.

Chapter 6

INVESTIGATING TECH-MEDIATED SIGN LANGUAGE INTERPRETING

In this chapter, I continue in the vein of exploring emerging technologies and focus on sign language interpreting technologies. The advances in AI technology are driving research on different applications for sign language AI, one of which is automating sign language interpreting. Unlike the dictionaries, which focus on individual signs, interpreting considers continuous signing and is therefore much more challenging to model computationally.

To explore how language and disability perspectives might inform the development of these automated interpreting technologies, I investigate what makes a good sign language interpreting experience. In contrast to Chapter 4 (multilingual captioning), where I introduce a language perspective to something that is typically studied through a disability perspective, this chapter does the opposite. Sign language interpreting is often understood simply as a translation problem i.e., moving between two different languages. However, communication access is much more than translation – interpreters are negotiating different access needs and refining their translation in accordance.

This chapter maps translation choices to communication access needs. Particularly, it distills the many facets of the access work done by human interpreters, highlights impact of technology on these practices, and outlines implications for development of sign language interpreting technologies. This work is in submission to CHI 2027 "From Translation to Access: A Study of Technology-Mediated Sign Language Interpreting", coauthored with Shuxu Huffman, Amelia Becker, Jennifer Mankoff, Naomi Caselli, and James Waller.

6.1 Motivation

Automated sign language translation has long drawn interest from technologists given its technical complexity and real world utility. As a three dimensional language with no standard written form, translation approaches have to build on both computer vision techniques and language modeling methods. While advances are being made on these computational fronts, open questions remain around how these sign language technologies might be developed and deployed ethically [41, 65, 86]. Historically, sign languages and the communities that use them have faced systematic barriers and oppression [224, 25], and many technologies that aim to bridge communication between Deaf and hearing individuals are embedded with biases against Deaf cultural norms and visual ways of communicating [72, 231, 16].

In response to these concerns, researchers are increasingly collaborating with Deaf communities and grounding their work in sign language linguistics (e.g., [146, 138]). For example, more works are incorporating facial expressions and non-manual markers into models (e.g., [294]), and investigating metrics more suited to sign languages [124, 198, 148, 288]. However, not much research explores how well trained models align with proposed downstream applications, such as automated sign language interpreting.

Human interpreting services are a common tool for signing Deaf and hard-of-hearing (DHH) individuals navigating a hearing world (in education, healthcare, legal, professional domains). However, these services are not always available due to a shortage of qualified interpreters, and economic and legal systems can be barriers to access [68, 47]. Video-remote interpreters (VRI) have been used to address this shortage by expanding the pool of interpreters available and allowing on-demand service, but this too has limitations [245]. Like VRI, automated interpreting technologies could be offered on-demand, but there are a host of questions around whether and how they would provide access.

To develop these technologies successfully, it is critical to have a foundational understanding of what interpreting actually entails. Sign language interpreters are not only tasked with converting content from one language to another, but also with mediating access [194] – the

most grammatically correct and faithful translation does not necessarily create access if it does not meet the needs of the consumer and the context. Yet, most sign language technology research is evaluated on quality of *translation* rather than quality of *access* afforded by these technologies. This presents a problem because access can take many forms, some of which are often in conflict with notions of "correct" language or communication norms [113, 133].

To address this gap, we sought to understand how human sign language interpreting, beyond translation accuracy, creates access. We conducted a study on American Sign Language (ASL) interpreting with 9 DHH signers, 9 hearing interpreters, and 2 Deaf interpreters, gathering their experiences navigating communication access across a range of interpreted contexts. Because participants have limited or no experience with automated translation, we used their experiences with remote video interpreting as a starting place to ground a conversation around emerging sign AI technologies.

Our work is situated among recent research that highlights the human expertise and care in creating access [30]. In our interviews, we dive into interpreters' and DHH signers collaboration practices. Our analysis distills different dimensions of communication accessibility tackled by interpreters (linguistic, cultural, temporal, contextual, relational). Interestingly, we find that this access work and roles adopted by interpreters cannot be divorced from their work as translators – rather, these dimensions fundamentally guide their translation choices. We show how technology reshapes these roles in the remote interpreting environment, and investigate whether participants think AI technologies would be able to mediate access in a similar manner. Overall our work contributes:

- Empirical accounts of sign language interpreting, highlighting the different dimensions of communication accessibility and their impact on translation choices

- Detailed insights into the complex and collaborative practices of interpreters and signers, highlighting constraints of the remote, virtual environment on co-constructing access

- Recommendations for access-centered sign language interpreting technology, offering directions for design and development

6.2 Background and Related Work

We begin by briefly summarizing the history of sign language interpreting and the infrastructure it operates under in the U.S. (where our work is situated), followed by a review of measures of interpreting quality. We then discuss research at the intersection of technology and sign language interpreting.

6.2.1 Background on Interpreting Provision

One of the most common access services provided for DHH people is sign language interpreting: typically bidirectional interpreting between a signed and spoken language. In the U.S., the Registry of Interpreters for the Deaf was first established in 1964 and marked the start of sign language interpreting as a recognized profession. Still, Deaf organizations had to advocate for DHH individuals' access to these services to be formalized [237, 214]. The Americans with Disabilities Act (ADA) (passed in 1990 and still effective today) legally mandates the provision of interpreting services for DHH individuals.

Sign language interpreters work in a variety of contexts: in healthcare, education, workplaces, court, public events and conferences, and so on. Contexts further vary by number of interlocutors present and communication norms (e.g., conversations vs. lectures, interpreting for one person vs. for many). In contexts where interpreting is mandated by the ADA, it is arranged by the organization or institution rather than the a separate government entity or the Deaf individual. In situations where an in-person interpreter is not available or the meeting is online, video-remote interpreting (VRI) may be used. For phone calls, the Federal Communications Commission (FCC) has mandated the option of Video Relay Service (VRS): DHH people are able to sign up for a VRS phone number with one of several providers [225].

Interpreters often work in teams – for longer assignments, two interpreters work together

and can support each other and take breaks when needed. Deaf interpreters (DIs) might join this team in situations where hearing interpreters are unfamiliar with or unable to fully match Deaf consumer's language practices. In this configuration, Deaf and hearing interpreters work sequentially: speech is interpreted into sign language by the hearing interpreter, and then the message may be re-produced by the Deaf interpreter for the DHH consumer. Deaf interpreters also work in other contexts: tactile interpreting for DeafBlind, written-to-sign translation work, interpretation between sign languages at international conferences [248].

6.2.2 Understandings and Measures of Interpreting Quality

Our understanding of what a "qualified" interpreter means has evolved as the profession has matured. Early understandings of interpreting used a "conduit" model, where the interpreter's role is simply to recode a message from one language into another. Over time, the field moved to discourse-based understanding [272, 229], acknowledging that the interpreter has an active role in making meaning [278]. As a result, approaches to evaluating quality also shifted from a product-oriented perspective focused on accuracy to a service-centered perspective exploring listener's needs and speaker's intentions [219]. The latter understanding of quality centers interactivity and successful communication among parties during an interaction [219].

Sign language interpreting differs from spoken language interpreting in that access to "qualified interpreters who can provide effective communication" is mandated under disability and access legislation such as the ADA. Interpreting is therefore situated not only as a translation practice but also as an access practice. This distinction shapes the work interpreters do; as put by Ingram, "[Sign language interpreters] have to interpret every act of communication—intentional or unintentional, human or non-human—that a receptor would normally perceive except for his hearing loss." [125]. Measures of sign language interpreting quality thus need to encompass components of both translation and access. Interpreters' certification processes such as the CASLI exam reflect this understanding, covering both form and function, with many items focused on extralinguistic aspects [54, 260]. A signifi-

cant portion of this exam is also devoted to ethics and decision-making, including ability to prepare for assignments and determine if access is sufficient [54] (i.e., if a Deaf interpreter or interpreting team is needed, or if the interpreter is appropriately certified and capable for the assignment). How interpreters make real-time decisions and determinations about access is an active area of research [259, 233], as is work on how DHH signers and institutions evaluate and select strong interpreters [181, 202]. Quantitative measures developed for sign language translation tasks may be inadequate for evaluating sign language interpreting even if specifically adapted to sign language [148, 124, 199] because of this breadth of practices involved in interpreting.

Creating widely-applicable and holistic measures of interpreter quality remains challenging. Many studies have explored users' satisfaction with interpreting services [202] as well as desirable characteristics in sign language interpreters [67]. While people are generally able to determine if they do or do not understand a message, it often requires linguistic training to describe the source of the miscommunication precisely. In addition, any communicative exchange involves multiple people, so identifying communication failures requires multiple perspectives [202]. Practitioner perspectives highlight a multifaceted understanding of quality including source fidelity, maximizing understanding, and managing interpersonal aspects of communication [192]. Converting these perspectives on quality to valid quantitative measures is challenging and requires a deeper understanding of the entanglement between linguistic and extralinguistic components in interpreting practice.

6.2.3 Sign Language Technologies

Research focusing on sign language technology has grown over the past decade [134] and comprises overlapping bodies of research. One strand focuses on developing computational methods to process sign languages using machine learning, including methods for sign language recognition, generation of sign language videos, and translation between signed and spoken languages [41]. Another examines how existing technologies support or constrain the use of sign languages, as videoconferencing and the internet give rise to newly possible

digital spaces [144] and novel accessibility challenges [231, 157, 176]. Both strands of work are foundational for sign language AI research, as built technologies would encompass both machine learning models as well as user interfaces.

Remote interpreting technologies. While remote interpreting promises easier access to interpreters, research has found that interpreting through a screen (such as in VRS, VRI, video-meetings, and hybrid setups) introduces new environmental and structural challenges. Sign language depends on visual attention and clear lines of sight, requiring ongoing physical coordination between the signer and viewer. When the interpreter is present via a screen, coordination of visibility, attention, backchanneling, and other interactional cues becomes more difficult [231]. Bringing in an interpreter on-demand (such as in VRS and VRI) also means the interpreter often lacks preparation, and cannot determine fit (in content, regional sign) for the assignment beforehand [276, 203]. Research on addressing these limitations has focused primarily on environmental constraints, like improving interpreters' access to visual cues (e.g., with conference room cameras [33], or portable fisheye cameras [184]) and improving users' view of the interpreter and other visual information (e.g., with AR VR glasses [184], videoconferencing design [231]).

AI interpreting technologies. While advances are being made on sign language translation (e.g., [294]), only a few works have explored implications of AI technologies on sign language interpreting. De Meulder warns that current evaluation practices for sign language AI too often benchmark systems against professional interpreters and rely on brief, scripted lab tasks or self-selected surveys—methods that miss real-world comprehension, equity, and ethical risk [65]. A survey of 32 U.S. Deaf participants found conditional interest in automatic sign language translation depending on context, with high performance expectations for sensitive domains (e.g., medical contexts) and major concerns about cultural appropriation and premature deployment [264]. There are a number of advocacy groups that have focused on regulatory checks for these technologies [284, 58, 8, 285]. These call for metrics for

organizational and technological readiness – for example, recommending that technologies offer bidirectional translation, preserve linguistic integrity of sign languages and consider linguistic diversity and intercultural practices.

Overall, our review reveals three interrelated gaps. First, VRI and VRS offer an interesting proxy for considering AI technologies, but both remain largely underexplored within HCI. Second, although many machine learning studies develop translation models for supporting communication [72], they seldom specify the downstream task or usage context in operational terms, which reduces the utility of technical results for applied system design. Third, calls for the ethical development of automated translation do not always define automated interpreting in ways that fully allow us to envision their form, which in turn hinders our ability to concretely guide research and evaluation of emerging technologies.

6.2.4 Team Positionality

The author team includes both Deaf and hearing researchers, bringing a range of lived, cultural, professional, and linguistic perspectives to this work. All authors have prior knowledge of and engagement with DHH communities, and several have close personal or professional relationships within Deaf spaces, primarily within the U.S.. All authors have at least some knowledge of ASL. Our collaboration was structured through an ASL-first process: we held weekly meetings in ASL, with sign language interpreters supporting authors' participation as needed. This working practice shaped our team communication and how we approached questions of access, language, and representation throughout the research process.

6.3 Methods

Our study aims were two-fold: first, to understand how access is conceptualized and created in sign language interpreting and second, to explore how existing and emerging technologies impact these practices of interpreting and access mediation. We wanted to gather perspectives of both users of interpreting (i.e., DHH individuals) as well as providers (i.e., including both hearing and Deaf interpreters). Particularly, we focused on individuals who had expe-

riences with remote interpreting (such as VRI, VRS, or Zoom interpreting) to understand their perspectives on current and future uses of technology in interpreting.

6.3.1 Procedure

We conducted semi-structured interviews to explore participants' experiences with interpreting, their perspectives on what contributes to successful access, and their views on the potential role of technology and AI in interpreting practice. Interviews lasted approximately one hour and were conducted via Zoom, in ASL and/or English (see section 6.3.3). The semi-structured format allowed us to ask a consistent set of guiding questions across groups while also following participants' lived experiences, examples, and concerns in greater depth.

6.3.2 Recruitment and Participants

We recruited participants through social media posts, mailing lists, and snowball sampling [201]. Interested participants were invited to complete a short online screening survey (under 5 minutes) to share their background and verify eligibility to participate. To be eligible, our participants needed to 1) identify as a DHH signer, hearing interpreter, or Deaf interpreter, 2) engage in interpreted contexts somewhat routinely (i.e., at least once a month) 3) have some experience with remote interpreting (e.g., Zoom, VRI, VRS). The last criteria ensured participants had tech-mediated interpreting experiences to reflect on during the interview. Excluded respondents were individuals who did not use interpreting or work as interpreters.

This resulted in 20 participants who either regularly work with interpreters or work as interpreters. Our sample includes 9 hearing interpreters (HI), 2 Deaf interpreters (DI), and 9 DHH people who use interpreters regularly (n=20). We report demographic information in aggregate. Of our DHH participants, three self-identified as female, four as male; one interpreter participant self-identified as non-binary, three as male, and six as female; three participants chose not to disclose gender. Most of our interpreter participants identified as

white (n=9), one as mixed race. For our DHH participants, two identified as Black, one as Asian, one as Afro-Latina and four as white. Two participants chose not to disclose race.

Our participants use ASL and English in their daily lives. Their experiences represent a range of professional and lived experience contexts, including educational, workplace, legal, healthcare, and community interpreting. For some participants, their perspectives on sign language technologies were additionally informed by their experiences training interpreters, working as technologists, and interactions with (other) DHH people more broadly.

6.3.3 Accessibility Considerations

Our research team comprises of authors with mixed levels of fluency in ASL and with different communication access needs. Interviews were therefore conducted keeping both participants' and researchers' communication access needs in mind. All signed interviews had an ASL-fluent author present. Participants could choose between communicating in ASL or in English, and interpreting and captioning were utilized by participants or research team members as needed. For signed interviews with interpreting, we used interpreters' voiced interpretation for initial transcription, which was then reviewed by ASL-fluent authors for accuracy. If the signing was not interpreted live, we requested translation and transcription of the recording through the university's interpreting service. These considerations were important not only for facilitating participation, but also for aligning our research process with the accessibility values central to the topic of this study.

6.3.4 Coding and Analysis

Three authors met and analyzed the interview data using reflexive thematic analysis [43, 44] over the course of several weeks. We began with an open coding pass through a subset of the transcripts to familiarize ourselves with the data and identify points of interests related to interpreting, accessibility, and technology. These formed the basis of our initial codebook , which we refined through discussion. All interview transcripts were analyzed by two researchers each; one researcher applied the codes, and another reviewed these codes,

PNum	Type	Contexts
DHH-1	DHH Signer	workplace, VRS
DHH-2	DHH Signer	workplace
DHH-3	DHH Signer	workplace, VRS
DHH-4	DHH Signer	VRS
DHH-5	DHH Signer	workplace (designated), healthcare, social
DHH-6	DHH Signer	healthcare
DHH-7	DHH Signer	workplace (designated)
DHH-8	DHH Signer	social, workplace, healthcare
DHH-9	DHH Signer	education, workplace
HI-1	Hearing Interpreter	higher education, workplace, legal
HI-2	Hearing Interpreter	legal interpreting
HI-3	Hearing Interpreter	conference, medical interpreting
HI-4	Hearing Interpreter	K-12 interpreting
HI-5	Hearing Interpreter	medical interpreting, government
HI-6	Hearing Interpreter	legal interpreting
HI-7	Hearing Interpreter	higher education, sports, community interpreting
HI-8	Hearing Interpreter	legal interpreting
HI-9	Hearing Interpreter	workplace, conference, spiritual
DI-1	Deaf Interpreter	higher education, test materials
DI-2	Deaf Interpreter	community, legal interpreting

Table 6.1: Participant type and main interpreting contexts discussed.

flagging ambiguities for discussion and suggesting additional codes. The first author analyzed all transcripts to ensure consistency, either as coder or reviewer. Throughout this process, we compared and refined themes to better capture both shared concerns and points

of divergence.

6.4 Results

We begin our findings by describing how interpreting fits into our participants' communication access practices (Section 6.4.1). We then discuss how interpreters mediate communication access, outlining the different roles they play beyond translators (Section 6.4.2). The rest of our results section dives into tensions and considerations in this partnership between DHH signers and interpreters (Section 6.4.3), the impact of remote interpreting on this partnership and these roles (Section 6.4.4), and finally speculations on AI technologies and their capacity to mediate access in a similar manner (Section 6.4.5).

6.4.1 Why Interpreting?

Our DHH participants had many different tools in their communication repertoires, including signing, speaking, writing, and captioning. They frequently switched between these in different contexts, weighing availability, ease of use, and access offered by each. DHH-4emphasized, *"there are many different ways to be Deaf, and we all have different communication needs and preferences"* (DHH-4).

Our DHH participants frequently used interpreting for work, school, healthcare, social contexts, and live performances or events. In describing why they picked interpreting, for some participants it was simply a matter of language access: ASL was their primary language, and interpreting allowed them to engage with the world in the language that was the most comfortable for them. Many liked the human connection offered by interpreters and their ability to capture and communicate *"not just the words but also the feeling and what's being expressed by what is said"* (DHH-1). This emotional context is frequently missing in captions, the most common alternative option for access.

Yet interpreting is not always readily available, and participants frequently turned to captioning, writing, and speechreading to fill the gap. For example, DHH-2 and DHH-8 often used live transcription apps to chat with friends, and DHH-7 and DI-1 similarly used

notes to write and communicate. Captions had the benefit of capturing the specific words used by conversational partners, allowed for spontaneous conversations, and did not need to be arranged in advance. However, these methods required them to be comfortable with English.

Some participants found that a combination of methods aligned best with their language preferences and communication needs. For example, DHH-2 preferred using ASL to understand what others were saying and speech to express themselves in interpreted contexts, stating *"I just prefer to have control over my words and what I say."* (DHH-2). DHH-7, on the other hand, said, *"I want to see the captions. And then, when it is my turn [...] I want to be able to use my natural language and sign, and have the interpreter express my signing that way."* (DHH-7)

For some participants, captions acted as an additional information stream, allowing them to verify what was interpreted into spoken English. For example, DHH-4 shared, *"I will catch that I signed something wrong if the interpreter voices something incorrectly."* (DHH-4) This would then prompt them to pause and clarify. DHH-3 used captions as a backup – *"sometimes interpreters miss information, I like to have interpreters backed up with captioning so that way I get the full story."* (DHH-3)

These combination setups worked well with video calls, where automated captioning is readily available. But for in-person scenarios, participants had to evaluate the environment (e.g., noise and group size) to pick between the two. Captions can go *"awry"* (DHH-1) with multiple speakers, with excessive lag or errors. DHH-8 found that for large groups and noisy environments like restaurants or games, interpreters were very helpful: *"it's just easier for me to get information when I'm on the go [...] an interpreter will have all the feedback loop and I can engage more directly with people"* (DHH-8). DHH-7 similarly appreciated the context interpreters were able to gather: *"it's the human that has the experiential language, all of the knowledge"* (DHH-7).

Finally, we note that sign language interpretation can be an imperfect substitute for being able to communicate directly in sign. Several participants (e.g. DHH-1 and DHH-9)

note for example, that they primarily work in Deaf spaces and with people who know sign language, and thus do not require interpreters in those spaces.

6.4.2 How do (human) interpreters mediate access?

Here we highlight how interpreters approach the process of mediating access and articulate the different roles they play beyond translators. Grounded in multiple parties' perspectives, our analysis distills five dimensions of communication access: linguistic, cultural, relational, contextual, and temporal. We showcase how these dimensions guide interpreters' decisions and translation practices.

Linguistic Access: Mediating Language

One of the main things all DHH participants valued was the interpreters' ability to match their language use. One way to understand this is in terms of speaker variation, and indeed, even our small sample used a range of labels to describe their signing styles (e.g., PSE, SEE, ASL, Black ASL). But another key aspect was the demands of the situation and their comfort with English. Some situations (e.g., a lecture) required DHH signers to be able to track and reconstruct the exact English terminology used. For other contexts, 'full' conceptual translations into ASL may be better. Sometimes signers had their own unique preferred jargon, and interpreters would have to pick up new vocabulary on the fly. The interpreters needed to be able to recognize and match this range of linguistic needs. Some interpreter participants approached this explicitly, asking signers to share their preferences. But for most it was a much more intuitive process, guided by watching the signer in front of them and using that to build a schema of their preferences.

Interpreters begin building this schema right from the start of interactions, and further refine it throughout the interpreting process based on the feedback cues shared by signers (e.g., gaze, nodding) and any communication breakdowns. Describing their process, HI-7 shared, *"[If] I start to realize it's not working for this person, I might adjust and say, oh, okay, I was trying to go a little more visual. Maybe I should go a little bit more closer to the*

English concepts." (HI-7) Thus they modified their signing style to conform more closely to the consumer's needs. Similarly, HI-8 described how they would track what *"triggers a better response"* (HI-8) and use that to guide their interpretation choices: giving more context, referring back to a previous interpretation, or using props as visual supports. They emphasized that there is no single way to interpret something; rather it depends deeply on the audience and their language background and access needs. H1-1 shared, *"sometimes, when there's a repair around a certain sign, it's not necessarily that that sign is incorrect [...] maybe, the sign that I use could work for one group, but not for another. [...] it's more about choices than it is about error, I think."* (HI-1)

This work of "matching" the audience was bidirectional, i.e., necessary for both signed and spoken utterances. Each environment brought forth different expectations regarding register and, therefore, vocabulary, tone, grammar. For example, HI-3 discussed how the expectations for English used are different in formal presentations or small work meetings, vs. casual conversations; *"being clear as possible"* (HI-3) requires different approaches and translation choices for each of these contexts. Some of the clarity comes from using accepted jargon. For example, HI-9 shared their experience encountering a specific sign during a conference: *"I don't know what verb to use, stuck to, connected to, attached to, bind to. [...] in your field, I know you have a word for that."* (HI-9)

In evaluating communication fit, our DHH participants therefore valued interpreters' experience with the topic at hand because it ensured familiarity with context-specific vocabulary (e.g., technical acronyms in software engineering environments). They would also work with interpreters to develop new signs to cover lexical gaps. DHH-5 shared the example of the word *"sprint"* (DHH-5), which means a very specific thing in software development (as opposed to its canonical meaning). They worked with their interpreting team to repurpose existing signs and invent new ones (e.g., name signs for coworkers, acronyms for long terms).

Perhaps the best example of the range of language mediation done by interpreters is Deaf interpreters' work. Deaf interpreters often work in contexts where hearing interpreters are unfamiliar with or unable to fully match Deaf consumer's language practices. DI-1, for ex-

ample, worked with international students who were learning in ASL-dominated classrooms. The teacher signs in ASL, but the students are only familiar with their native sign languages. *"So, I'm kind of teaching ASL, supporting ASL acquisition. [...] It's a triangle approach. I'm sitting with the student. And we're pointing back and forth, making sure that they're following along, clarifying as we go. When they look at me, I will give them kind of the key points."* (DI-1).

Cultural Access: Making the Implicit Explicit

In addition to linguistic mediators, we found that interpreters also acted as cultural facilitators, providing a bridge between Deaf and hearing parties, negotiating different knowledge bases, and making implicit information explicit in their interpretations.

Implicit information can take many forms. For example, interpreters frequently incorporate auditory information from the environment into their translations. This could be background sounds like the sound of a phone ringing, or a doorbell that might be relevant to the conversation at hand, filtering out other sounds like traffic noise outside. It can also be paralinguistic information, like HI-3 shared, *"There's a lot of information in the tone [...] are they being friendly, or mad, or are they sarcastic? [...] I have to figure out how to convey the same message with the same attitude in the other language."* (HI-3) Interpreters paid careful attention to the relationships in the room and the power dynamics at hand, incorporating emotional subtleties into translations as needed.

Interpreters also made cultural knowledge explicit in their translations, updating their translations based on the receivers' "fund of knowledge". HI-8 offered age as an example – *"Someone who's 17 has never lived their life without a cell phone. A 70-year-old Deaf person might not even have a cell phone, right? [...] Their thought worlds are very different."* (HI-8) This means that in a given situation, effective communication requires bridging knowledge gaps in addition to offering a translation. For example, in a medical context, a doctor might discuss words like nutrition or diabetes. Simply fingerspelling these words (or even using a corresponding sign) does not necessarily offer access if the patient does not know those

terms. In these scenarios, interpreters might expand their translation to describe this term conceptually. For example, HI-5 described discussing a ligament tear – *"I set up the joint like this [visually], and we talked about how there are bands that are wrapped around that kind of stretch with the knee, [these are called] ligaments and things, and how those ligaments were worn out."* (HI-5) Choosing whether to offer an expansion required careful consideration from interpreters, so as to not underestimate or otherwise insult the Deaf patient. HI-9 approached it by asking them directly, *"I can choose to say: 'do you want to ask the doctor about this?' Because I know that this is missing [context], but I don't know whether they know it or not."* (HI-9)

This process of making implicit information explicit worked in both directions (e.g., Deaf to hearing, patients to doctors). HI-3 shared an example of adding context about the Deaf experience when voicing a parent's concerns about their daughter locking herself into the bathroom which the doctor might not immediately understand *"the doctor might not automatically think, oh, you also can't hear, and you can't see through the door, so you don't even know what's going... like, is she jumping out the window? You don't know that, but a person who could hear might know that."* (HI-3)

Contextual Access: Advocating and Sharing Domain Expertise

Many interpreters in our study specialized in a specific domain such as legal, medical, or educational interpreting. Routine exposure to interactions in these contexts meant that interpreters became carefully attuned to translation choices and procedural constraints unique to these domains.

For example, HI-4 has spent years interpreting in primary education, frequently working with young children who are just beginning to acquire sign language. In this context, supporting students' understanding and engagement requires *"signing and gesturing and being very big with your facial expressions and big with your body. [...] expanding as much as possible."* (HI-4) In contrast, working in legal interpreting limited how much HI-6 could expand and make explicit. *"In community interpreting, I could ask lots of different examples*

and say, which one is it? Or I can ask more of a leading question, I could set something up and negate it. But in interpreting in court, you have to be very, very careful. Like, if they haven't mentioned a body part, you can't say, did he touch your breasts?" (HI-6) This need for abstraction is prompted by rules of evidence, and requires expertise to navigate while still ensuring communication access.

Additionally, in-depth knowledge of a topic can shape the smallest of translation choices. For example, in ASL many verbs require the signer to specify the direction of the action while English does not. HI-9 shared an instance where familiarity with underlying spiritual philosophy impacted verb directionality in their interpretation in a theology context: *"[in Lutheran theology] it's always God coming to us. It's never us having to go to God."* (HI-9).

Beyond these translation choices, interpreters also mediated access more indirectly by sharing domain expertise and access knowledge with DHH and hearing parties. For example, heading to court is a stressful experience for everyone, but this stress is compounded by communication barriers DHH individuals face. HI-2 shared how they worked to meet these clients where they are in terms of experience with the legal domain. Similarly, DI-2 frequently begins by educating judges about interpreting and their role as a certified Deaf interpreter during legal proceedings, finding that establishing this understanding paves the way for communication access. These practices of sharing accessibility knowledge applies to educational contexts. Working with international students, DI-1 often *"introduce[s] them into interpreter practices, you know, because a lot of countries don't have Deaf interpreters, so the process itself is new to them."* (DI-1)

Finally, interpreters also ensured that the communicative context was accessible and conducive to the interpreting process. HI-1 shared they might pause legal proceedings if effective communication cannot be established, saying, *"I'm not able to understand this [Deaf] person, I'm not able to confirm that they're understanding me, and because this is a legal situation, you know, this is high stakes, and so we'll need to postpone these proceedings."* (HI-1) Often interpreters anticipate accessibility issues before they arise, checking in when visibility might be obscured or the physical space may need to be rearranged. HI-2 explained that they take

initiative warn courts that about issues that may arise – for example, multiple DHH people in a trial could require separate interpreting teams: *"The court doesn't necessarily know to research that... so I try to be proactive."* (HI-2) In another example in an educational context, HI-4 (HI) discusses working with young children who are still learning their own access needs and how to advocate for them.

Relational Access: Engaging Hearing Participants

All interpreters in our study emphasized the importance of preparation in ensuring quality interpretation. While domain knowledge and prior experience helped, having some context about specific content and participants significantly eased their cognitive load. To acquire this context, interpreters actively sought information, relying on their networks and relationships to acquire prep materials ahead of time. The contact person and material varied across contexts: it could be a meeting agenda shared by the Deaf party, slides shared by a teacher, case information shared by a court clerk, or a setlist shared by a musician. However, these requests were not always honored and the degree of detail provided varied significantly.

Beyond meeting materials, interpreters also sought out information about the parties that would be present and the relationships between them. This helped them develop schemas and an understanding of the relational and power dynamics in the room. Connecting with their team in advance let them work through logistics and pool contextual knowledge. One noted, *"sometimes the team's been there longer than I have, sometimes I have more contexts than I can share [...] what do you know about the topic, the person, this place? What can you bring to this table that we can leverage together? "* (HI-7)

Information seeking continued beyond acquiring materials; interpreters also sought out advice on how to best construe an interpretation. For example, if unsure about the right sign for a technical context, they might ask Deaf friends and DIs or use online dictionaries. In performance contexts like concerts and plays, interpreters discussed working with Deaf consultants on translations that convey the meaning while still capturing the artistic quality.

Recognizing that access is co-created, interpreters often engaged hearing participants

in the interpreting process, encouraging them to meaningfully think about how to make things more accessible. Participants described collaborating with hearing participants on their materials ahead of time: working with comedians on their jokes (HI-5), musicians on their songs (HI-6), or pastors on their sermons (HI-9). This process often deepened hearing participants' understanding of the interpreting process, *"that it's not just a word-for-word substitution project"* (HI-9). Over time, they got better at meeting interpreters halfway, such as preparing to share key themes in their performance ahead of time. Beyond prep, interpreters also drew out knowledge from hearing participants in the moment. HI-5 shared, *"Let's say, like, it's an orthopedic doctor. I'll say, hey, could you maybe Google some pictures so that we can visually see this, and then we can expand on it to get a better understanding?"* (HI-5) thus ensuring they have more context for their translation.

Temporal Access: Maintaining and Disrupting Pace

Most of the interpreting contexts involved simultaneous translation, requiring a significant amount of temporal gymnastics while interpreting, as concepts take a different amount of time to express in different languages. For many of our DHH participants, an interpreter's ability to maintain pace with a conversation (both for signing and speaking) was an important aspect of good communication fit.

Good pacing depended on practice and training, but also often also required foresight. Interpreters working in education (HI-4, DI-1) discussed practices of front-loading information to bypass the overwhelm of being introduced to many new concepts and vocabulary during a lesson. Typical practices like elaborating an idea can be challenging to accomplish in real time without falling too far behind. DI-1 shared, *"I find it very helpful to have some prep time and discussion before class. You know, all the possible signs that teachers might use. And then when we get to class, I find myself working a lot less hard."* (DI-1) Similarly, HI-5 discussed the importance of timing for comedy context: *"we want the Deaf folks to laugh at the same time. We don't want them to see the hearing audience and be like, what... what was that said? Why is that funny?"* (HI-5)

Timing was further complicated in DIs' work, where multiple degrees of interpretation are happening. DI-1 described their process of working with international students in an ASL classroom: *"If an interpreter tries to translate at the same time, the student can't watch them. Instead we try to pause the interpretation and take in all the information, and when the student gets a break to look at us we translate the whole idea at once."* (DI-1)

In addition to *maintaining* pace, another component of access was being willing to *disrupt* pace. Most communication follows hearing norms that prioritize speed and flow, but communication access often requires interlocutors to pause, repeat, or clarify. Interpreters played an important role in this, calling hearing people to change their behavior. DHH-8 shared, *"it's really nice and helpful if the interpreter can maybe speak up and ask people to slow down or let people know that the interpreter is catching up to what is being said."* (DHH-8) Judging when to conduct repair (i.e., calling for a pause, clarifying at the end of a conversation, and/or taking accountability for errors) demanded a great deal of reflexivity from interpreters and the ability to recognize when a misunderstanding was an artifact of interpretation or stemmed from the source.

6.4.3 Interpreting as a Partnership: Tensions and Considerations

Many of our DHH participants approached interpreting as a collaborative process: *"it takes two to tango, right? [...] we're communication partners."* (DHH-5) They actively contributed to creating access by sharing materials and context ahead of time, feeding the interpreters vocabulary, matching interpreters' communication styles, or slowing down their signing and clarifying as needed.

DHH participants emphasized a strong partnership built on trust and transparency was key for success; they needed to feel confident in interpreters' skills and willingness to receive feedback: *"[interpreters] having the right attitude and the right heart is very important to me."* (DHH-2). As most parties in the interaction cannot follow both languages and verify output, transparency when navigating breakdowns was critical. Our interpreter participants valued the same: *"I'm very supportive of an open process, so if there's something that I'm*

saying incorrectly, [...] I'd rather that the Deaf person can see my team is saying, oh, no, no, no, they said this instead." (HI-7)

Having a third-party involved in conversation to mediate access also raises new considerations around privacy. DHH-1, for example, preferred to use other access practices for situations involving sensitive information (e.g., bank details). DHH-2 felt that interpreters' role as facilitator reduces intimacy in personal conversation, sharing, *"If I have a friend who I want to have a deep personal conversation with [...] I think that I would prefer to have it be not perfect, but to have it be direct and to communicate one-on-one."* (DHH-2) Sometimes, interpreters had to step back as they were well-acquainted with people in the conversation. HI-9 shared an example of a spiritual counseling conversation where they felt their presence would break privacy, saying *"I'm a member of this community, [...] I can't interpret this conversation between the two of you."* (HI-9)

Interpreting as a collaboration also raises questions around representation. DHH signers often do not share the same identity characteristics (e.g., race, gender, age) as their interpreters, and these differences impact how their words are received by the audience. This mismatch could also impact the degree to which the interpreter could follow cultural references and nuances. However, representative fit was not always the DHH participants' main priority. DHH-8 shared, *"I prefer a white interpreter who has worked with Asian clients before [...] who knows Asian culture versus an Asian person who doesn't know anything, you know?"* (DHH-8)

Finally, preserving autonomy and negotiating power dynamics were critical to a successful partnership. Interpreters needed to recognize the privilege they hold as hearing people, and constantly reflect whether their actions centered the Deaf person and their agency. HI-9 reflected, *"it varies from person to person, and so it's gauging, like, how much autonomy the person has and wants to have in that situation, so that I'm not stepping on anybody's toes."* (HI-9) However, interpreters could use their privilege to subvert larger dynamics at play between other parties. For example, DHH-6 shared an instance where a white interpreter was able to increase the sense of safety she felt during a police encounter.

6.4.4 Remote Interpreting: Constraints and Affordances

Technology-mediated interpreting has risen in recent years. In this section we explore the constraints and affordances of the remote interpreting environment, as well as the impact on the access roles played by interpreters.

Constraints and Affordances

Availability of interpreters varies widely across locations, and the biggest advantage of tech-mediated interpreting is that it expands the pool of interpreters. For interpreters, this modality allowed them to *"dial in from anywhere"* (HI-1), minimizing commute and maximizing interpreting time (HI-6). Additionally, VRI and VRS are services that can be accessed "on-demand", thus offering some degree of spontaneity.

When all parties are remotely connected through the same video conferencing platform, DHH individuals had better access to the rest of their communication repertoire: close up video for speechreading, reduced background noise for listening, addition of automated captioning and transcription, and chat features for occasional written English communication between all parties.

However, our participants echoed many of the constraints of remote interpreting described in prior research [231, 270], such as flattening of sign language, lack of shared spatial information, or difficulty getting attention and using gaze. Video quality varies widely depending on lighting and background, and improving layout and visibility is not always an option (e.g., availability of custom grid size, spotlighting and pinning videos varies across platforms). Unfortunately choice of platform was not always up to DHH signers and interpreters. Many businesses and other institutions (e.g. courts) required specific platforms for privacy and security, standard policy, or unspecified reasons.

Hybrid contexts (where some participants are co-located) were often the hardest, especially when considering situations *"out in the world"* (DHH-6). Some contexts like office meetings may be set up for hybrid, and meeting norms match. But for school field trips

(HI-4), concerts (DHH-8), or restaurants (DHH-2), the environment is far too mobile, noisy, dark and dynamic. Even semi-structured scenarios like doctors' offices present challenges setting up visuals to maintain communication access during common procedures (e.g., lying down, changing into a gown, or having eye drops put in).

As such, most of our DHH participants preferred in-person interpreting, with its access to *"facial expressions, depending on the context, emotions, affective language"* (DHH-7); that is, sign language in all three dimensions. Interestingly, there were also more subtle shifts in communication norms. DHH-1 remarked, *"[When the interpreter is in person, it] requires the other hearing people to maybe act a little different, and the dynamics are different, and the turn-taking, the timing"* (DHH-1). Below we consider how interpreter roles change in these tech-mediated contexts.

Impact on Interpreter (Access) Roles

The constraints and affordances of the remote interpreting environment changed how and whether different interpreting roles could be realized. Some roles saw minimal changes. For temporal access, participants found that turn-taking during a conversation was harder, but the other aspects of keeping pace with translation and conducting repair were largely the same. Similarly, for relational access, as most of the preparation occurs ahead of time, there is not much change for online and hybrid calls. However, VRI and VRS are exceptions: the on-demand nature of this work meant that most of the time interpreters were given minimal context. HI-7 shared, *"I don't know where they're from, I don't know who they're calling [...] you have to be more willing to kind of go with whatever goes"* (HI-7).

On the other hand, the roles of language mediation and cultural facilitation were notably more challenging in remote interpreting. Both roles rely heavily on access to feedback and implicit environmental cues, and with poor audiovisual quality and cameras often turned off, these cues are less accessible. This makes it difficult to build the language schema of conversants for successful language mediation, or read body language, tone, or other contextual information for cultural facilitation. For example, DHH-2 discussed a time when

they were talking to someone in a healthcare situation who was suicidal: *"the interpreter wasn't understanding, because they were on Zoom, and it was very frustrating. [...] If the interpreter is there in the same environment, and is able to see everything that's going on [...] that really helps with facilitating communication better, because there's that awareness of the environment."* (DHH-2) Moreover, most participants show up right when the online meeting begins, reducing opportunities for the small talk that allow DHH signers and interpreters to become familiar with each other. *"you just kind of start really cold, and that doesn't feel as good. And then also, when it's over, everything just, like, closes. [...it's hard to discuss] what you would do differently next time, and stuff like that."* (HI-3) For DIs, this significantly impacted their approach. DI-1 insisted strong relationships had to be established first in-person before they could successfully work with DHH clients online: *"I don't recommend Zoom for the beginning of the semester. Luckily, we've developed a relationship over time, so we could, if we needed to, move it to Zoom occasionally."* (DI-1)

Finally, we find that the role of context and access experts transforms into providing technology support and expertise to both Deaf and hearing parties in these environments. Interpreters are frequently tasked with explaining how spotlighting, pinning, and cohost privileges work – all features that need to be set up correctly to ensure communication accessibility. In hybrid contexts, interpreters had to frequently troubleshoot video and audio feeds, and advocate for onsite persons to help adjust the set up. HI-6 frequently encountered this with courtrooms when the location of the camera and the screen are not always aligned: *"the client is looking at the screen, watching me, but all I can see is the side of his face."* (HI-6) HI-2 found herself 'directing' the court clerk in these situations: *"they'll work with me, but it's awkward, and sometimes the Deaf person has to sit in the witness stand, because that's [the only] place they can make the camera focus on them."* (HI-2)

The Changing Partnership

Overall, despite the promise of increased access, remote interpreting brought forth new challenges. Particularly for VRI and VRS, the maintenance and deployment of hardware was

subpar and was a source of frequent frustration for DHH users. DHH-6 shared, *"Oftentimes the connection is lousy, sometimes the battery will die, sometimes the interpreter will be there, and then all of a sudden it will just shut off and then it's like, oh, man, okay, well, I guess now we have to turn it back on, we have to call another interpreter"* (DHH-6). DHH-2 similarly found that microphones and speakers were frequently broken, with interpreters and hearing parties struggling to hear. Shifting roles bring increased responsibility on DHH users to figure out access hacks with these technical breakdowns, which could be dangerous in high stakes, time sensitive contexts.

Additionally, the choice of VRI for a situation (e.g., doctor's appointments) was not entirely in the hands of DHH users. Requesting in-person interpreters required additional self-advocacy but was sometimes necessary. DHH-5 described advocating against VRI during an eye checkup: *"I can't do that, because when you're putting, you know, drops in my eyes. I'm not going to be able to see the interpreter on a screen."* (DHH-5) However, the shortage of in-person interpreters meant that DHH participants often did not have much choice, pointing to diminishing agency/autonomy.

For interpreter participants, working in VRI and VRS was incredibly stressful and lacked the human connection they hoped to foster. They shared that policies at the centers offering these services are increasingly strict, with interpreters having no breaks between calls, no support from other interpreters during calls, and no opportunities for growth. Switching contexts between calls was taxing. HI-5 described interpreting on a call where someone was in distress. *"That was really stressful and heartbreaking, and then I'm on to the next call, and I have to pretend like nothing happened [...] it's just not healthy mentally, emotionally, or physically for interpreters."* (HI-5)

However, outside of VRI and VRS, we found that interpreters and DHH signers worked to maintain the collaborative nature of interpreting (i.e., "the partnership") in the remote environment. To navigate platform constraints, participants often set up another meeting in parallel (e.g, zoom meeting, or FaceTime) for backchanneling. Participants also used chat, text, Whatsapp, etc. to communicate non-urgent information with their team (e.g.,

who goes next). Similarly, DHH-4 also frequently used built-in chat to offer feedback to the interpreter they were working with, typing in preferred vocabulary and clarifying questions. Shared synchronous documents (e.g., Google Docs) served a similar purpose. *"They'll do that during my interpreting, like, 'I like your shirt! Oh, that word choice was great!' And I'm interpreting, and I'm seeing the Google Doc populate with that information. It's nice. It's a very collaborative process, which is cool."* (HI-1) However, these practices required them to invest in multiple devices and larger screens, frequently causing overwhelm and divided attention.

6.4.5 AI Interpreting: Speculations

The previous section highlighted some of the shifts in interpreting practices as technology is incorporated, drawing on participants' current experiences with remote interpreting. In this section, we turn to the future, and the possibility of automated interpreting.

When we asked our participants about this possibility of automated interpreting, many mentioned that it was a "hot topic" in the community. However, when probed further, our participants were mostly skeptical about these technologies' viability. Notably, our discussions stayed fairly abstract, given that most participants hadn't encountered any sign language AI technologies. The few participants who had seen demos remained unimpressed.

A primary concern was performance of these technologies, and their capacity to master both receptive and expressive aspects of sign language. DHH-7 said, *"it might be easier for AI to understand signs. But for AI, to start signing, I think I don't think it's ever going to be possible."* (DHH-7) They were concerned about its ability to render realistic humans with facial expressions and fluid movements. Even with the receptive side, they were unsure how models would learn contextual nuances of different signs.

Going beyond language understanding and production, participants still had concerns about AI's capacity to *interpret*. H1-1 shared, *"I don't want to assume that technology's going to replace people. [...] Maybe to support and offer some assistance, for sure, but for full access, I'm doubtful."* (HI-1) Some of this stemmed from being skeptical about

technology's capacity to act as a language mediator, and adapt to the diversity of signing styles (reflecting background, race, age etc.). Instead of AI code-switching to different users' needs, participants worried that they would have to change their own signing instead: *"Kids will grow up with AI interpreting, and with their signing their intention is to fit the AI interpreter, signing for AI instead of signing for Deaf people."* (DHH-1) They worried that this would flatten and homogenize sign languages over time.

Participants similarly worried about technologies capacity to detect emotional and relational nuances. DHH-3 asked, *"Could AI understand how seriously I am sick? Or could AI -- translate and convey how upset I really am?"* (DHH-3) They were also skeptical about AI's ability to express emotions, and deliver their messages as they intended and facilitate across cultures. DHH-4 summarized, *"A robot is not good enough [...] the point is the human connection. Humans help read the room. They pick up vocal tone, they pick up facial expression. They help with conversations in a way that avatars absolutely cannot."* (DHH-4)

With our participants' speculations also came concerns about early deployment of these technologies. As seen with VRI and VRS, the choice of modality is not always in the hands of Deaf consumers. Participants worried that those with power would not be informed about the limitations of these technologies. DHH-5 said, *"you have to think of ways to get that information to people, those caveats."* (DHH-5) Others, like DHH-9, were also concerned about the impact on the interpreting community. DHH-6 said, *"my concerns are how AI can steal our interpreters. Steal! They're stealing their hands."* (DHH-6) DHH-4 was frustrated by the burden this would place on the Deaf community, to advocate against underperforming and unequivalent technologies. *"It is a tax on us to educate repeatedly, and still run into the same wall with different people. [...] the interpretation of how [AI] should service Deaf people, it should not be influenced by lack of understanding by hearing people."* (DHH-4)

We found that participants often pivoted away from interpreting, suggesting more low-stakes contexts where these AI technologies could be deployed. Some participants were open to sign-to-text technology that functions as a virtual Alexa [263], or could augment or speed up text-based communication with hearing people (DHH-7). For some, AI interpreting put

sign language technology on the wrong track: *"At the core, we want more people to sign. And not have robots sign for them."* (DHH-4) DHH participants want to live in world with more opportunities to sign with people: other participants also suggested they were more interested in technology that got more people to sign - rather than try to eliminate the need for hearing people to learn sign language or the interpreting profession.

6.5 Discussion

Our results offer an in-depth study of sign language interpretation as an access practice: exploring how it affords access to DHH individuals (Section 6.4.1), the work interpreters do in matching communication access needs (Section 6.4.2), and the partnership between these parties in co-creating meaning (Section 6.4.3). We see how it acts as a specialized tool in the communication repertoire, offering human connection and culturally-nuanced, well-paced, and contextualized access. However, in examining experiences across different modalities (in-person, online, hybrid, VRI/VRS), we note that technology constrains interpreting practices (Section 6.4.4), demanding users reevaluate the partnership while often receiving less access. These experiences with technologies shape participants' concerns with automated interpreting (Section 6.4.5): that DHH users would have a greater responsibility in managing hearing participants and navigating communication breakdowns, while simultaneously having reduced agency and autonomy to guide the interpreting process to match their needs, thus losing much of what makes interpreting special.

Our findings showcase that automated interpreting is often implicitly understood to follow a 'conduit model' of interpreting – where an interpreter is neutral and invisible and simply translates content from one language to another. In reality, a good interpreter does not passively interpret without consideration of what is needed for communication to be successful. Our participants were constantly engaged in meta-reflection of the interpretation process: monitoring needs and breakdowns, engaging hearing participants, and explicating cultural nuances. They were skeptical that a conduit approach to automation would succeed in provisioning access. Interestingly, this parallels trends we see in automated captioning

technologies: they frequently fail in noisy and hybrid environments, fail to incorporate non-speech information (e.g., emotions, background sounds) [64, 186], and need to be redesigned to engage hearing participants actively [191]. Whether captions or sign language technologies, these socio-technical challenges will occur with increased automation and must be designed for.

In this discussion, we reflect on the future of sign language technology, offering three tiers of recommendations: improving delivery of remote interpreting, exploring technologies to support interpreters, and considerations for automated interpreting technologies.

6.5.1 Improving delivery of remote interpreting

We find that even with highly skilled human interpreters, remote interpreting often falls short of meeting users' needs (Section 6.4.4). This suggests that one of the most pressing problems in this space continues to be the form factor of technology, i.e., whether signers and interpreters are visible to each other and have access to the environmental information critical to the conversation. These concerns are relevant not only for remote interpreting but also automated interpreting technologies, as this would shape the inputs available to any algorithm and how the outputs are rendered.

While video quality and intelligibility have improved significantly in the last few decades [52, 262], our participants still report frequently experiencing frozen screens, audio issues, and other technical breakdowns. These issues point to the need for hardware to be more robust. Often, these technologies are being selected and operated by hearing participants (e.g., a court clerk or a doctor) who are unaware of the demands of interpreting. Additionally, video optimization and processing techniques could improve signing intelligibility in poor lighting and mobile contexts.

Negative evaluations of remote interpreting focused on several interrelated issues: the contextual and feedback cues interpreters rely heavily on are missing; and it can feel very cold – it is difficult to build rapport or prepare. Research could consider how videoconferencing features could be improved to address these issues – offering customizable layouts or options

for multiple calls in parallel. Fatigue and non-verbal overload in remote meetings is a common phenomenon [22, 83], and remote sign language interpreting could be informed by strategies to improve remote work and communication broadly (e.g. virtual and augmented reality, representations of space, gaze, and extralinguistic information) [193]. Exploring built-in mechanisms to guide camera, microphone, and screen placement in hybrid contexts could be valuable in improving the virtual signing space. We emphasize the need for continued research on improving video-conferencing accessibility for signers (e.g., [231]) and exploring different output modalities (such as AR/VR [33]).

6.5.2 Exploring technologies to support interpreters

Our results highlight the many ways interpreting is a cognitively demanding task. Our participants had to navigate multiple languages, process a rapid stream of information, and make a multitude of adjustments on the fly. While working in teams eases some of the cognitive load, there were many situations where interpreters were required to work alone, including in VRI/VRS, rural areas, and emergencies. This workload frequently results in burnout and injuries, and motivates interpreters to leave the profession after a few years [69]. This suggests that a valuable direction of inquiry would be to explore how technologies might support interpreters in their work, improving retention of qualified professionals.

The field of computer-assisted interpreting is devoted to this very question [103, 80], but sign language interpreting has been under explored in this field. Interpreting between sign and spoken language has unique affordances and demands (e.g., multimodality) which differ from interpreting between spoken languages and could point to new and interesting ways to leverage speech and sign recognition technologies. For example, interpreters in our study already use automated captions for support – captions act as a backup they can turn to if they miss anything or need spelling of specialized vocabulary and speaker names. Isolated sign recognition technology might be similarly valuable in enabling interpreters to quickly look up unfamiliar or regional signs. Yin *et al.* [290] proposes an interesting application of fingerspelling recognition relevant to this context in which a model could be trained to

detect fingerspelled words and suggest alternative signs (e.g., for STEM vocabulary). Such technologies could support interpreters in their role as language mediators, and refine their contextual knowledge. Another direction could be to support interpreters' prepping processes. Interpreters can only remember so much about each client, and AI technologies could potentially be better suited to tracking client preferences (e.g., for vocabulary) systematically. Other (verifiable) technologies to summarize prep materials, research translations, and connect with other interpreters could also be valuable in supporting relational aspects of interpreter work.

6.5.3 Considerations for AI interpreting technologies

Our participants were lukewarm, at best, about the use of AI in interpreting, and raised many concerns. However, we recognize that the design space for AI is wide, and that built technologies could operate in imaginative ways and unexpectedly address participants' concerns. Below we offer some provocations and recommendations for thoughtful design, development, and deployment of these technologies.

On Design: Interpreting as a Partnership

We find that interpreting is a deeply collaborative practice, with DHH parties guiding the interpretation to best match their needs (Section 6.4.3). DHH individuals continuously shape the interpreting process to offer clarification or context, suggest preferred signs, request a change of pace, and more. However, as automated interpreting follows a conduit model, listener feedback is rarely considered as a critical input required to guide the translation. Approaching automated interpreting through a different model – one that values DHH insights and contributions to access and recognizes successful interpreters as active agents of access – requires creative thinking about how such listener feedback is gathered and incorporated. In human interpreting, much of this feedback is conveyed through facial expressions, gaze, and discreet signing. For automated contexts, user feedback could perhaps be more explicitly stated. One novel direction for future inquiry could be exploring different interaction

techniques and mechanisms for users to share real-time feedback. Additionally, exploring tensions in this partnership – around trust, privacy, representation and autonomy – is a critical aspect of designing these technologies (e.g., how might automated interpreting systems imbue transparency?). These questions overlap with larger body of AIxAccessibility research, which explores explainable and verifiable outputs [281], obfuscation tools to maintain privacy [4], and identity and representation concerns [179], and thus are good opportunities for cross-disciplinary collaboration.

On Development: Interpreting as a Series of Choices

Developing automated interpreting technologies through an access lens requires a way to evaluate the degree to which they consider the different accessibility facets highlighted by our participants (Section 6.4.2). Most existing datasets for sign language offer a one-to-one mapping between a spoken and signed utterance, and trained models are evaluated on whether the produced translation matches the data labels (e.g., through WER or BLEU scores). However, we see that our human interpreters take a different approach in evaluating success, not just checking if the signs are generally correct, but whether they are appropriate for the audience and context. Recall, *"it's more about choices than it is about error"* (HI-1). Interpreters' choices go beyond just picking the right sign. Participants described deciding whether to expand a concept, summarize an utterance, refer back to a previous interpretation, depict or describe a concept, pause for clarification, and so on. The field of interpreting studies has frameworks for analyzing many of these translation choices for human interpreters (e.g., [180]). It could be valuable to approach evaluation of automated systems in a similar manner, where rather than just accuracy, we examine the multitude of different choices a system can make in a given scenario. This would reframe the task from just rendering an acceptable translation to rendering an *accessible* translation that matches the audience and context needs: models would be trained to optimize choices made. This has vast implications for the design of automated tools. Data sets would need to reflect a range of audiences and contexts, capture multiple interpreters' choices, and include meta-data about

rationale. Algorithms would need to include DHH and interpreter context in their training data. Evaluation metrics would need to reflect not just on correctness, but on factors that influence interpreter choices, such as relevance, conciseness, and continuity in the mapping of signs to concepts.

On Deployment: Interpreting as a Reflexive Practice

Human interpreters serve as guardrails in their own deployment: we see that they can intervene when they recognize a lack of resources to provide appropriate access. However, automated technologies do not embody this same level of reflexivity. If the human interpreter is removed from the context, who determines access? We suggest that hearing participants are not equipped to make these decisions, nor should AI systems be entrusted with this task. Rather, the deployment of automated interpreting would have to be under the purview of the DHH individuals themselves. This would be a different paradigm for interpreter provision: rather than automated interpreting being provided in lieu of human interpreting by institutions, it could instead be something DHH individuals could *opt-in* to, affording additional access in scenarios underserved by existing infrastructures (e.g., small, everyday interactions). People with disabilities have long been early adopters of technology, re-purposing imperfect tools to meet their access needs (e.g., [114, 130, 108]). While this is not a reason to not aim to develop high-quality technologies, it points to a way we might mitigate risks of early deployment and shift power back to DHH consumers – allowing them to pick the access practice that best fits their needs.

6.5.4 Limitations and Future Work

This study has several limitations. First, most participants, DHH and hearing, were comfortable using English. This is not true of all DHH individuals, and as we see in our interviews, English language proficiency and comfort is a recurring theme in participants' choices related to interpreting and access. As a result, the findings may not fully reflect the experiences of DHH people with lower English proficiency, particularly those affected by language depriva-

tion. Future work should include participants with a wider range of language backgrounds to better capture how language access shapes engagement with these systems. Second, our research is primarily situated in the U.S and influenced by American Deaf culture and U.S.-centered assumptions about communication, access, and technology. These perspectives may not generalize to other Deaf communities, where norms around sign language and interpreting may differ. Future research should examine how these findings extend to other contexts. Third, our work focused primarily on live communication contexts and did not examine engagement with pre-recorded media, text or other non-live communication settings. These contexts may involve different accessibility challenges and should be explored in future work. Finally, this work focuses only on perspectives of DHH signers and interpreters – hearing interlocutors may offer additional insights.

6.6 Conclusion

In this work, we reframe sign language interpreting from a problem of translation to a practice of access. Through a study of DHH individuals, hearing interpreters, and Deaf interpreters, we show that interpretation is not just a recoding of language but a situated process that involves linguistic, cultural, contextual, relational, and temporal work. Our findings highlight the limitations of current technology-mediated approaches to interpreting, and surface the risks of deploying automated systems that overlook this complexity of access. Rather than replacing interpreters, we suggest that future systems should support the work of access by augmenting human expertise, enabling collaboration, and preserving user agency. By centering the lived experiences of DHH individuals and interpreters, this work calls for a shift in how interpreting is conceptualized, moving us toward technologies that approach access as a dynamic, human-centered practice.

Chapter 7

DISCUSSION

In this chapter, I revisit the research questions that guide my dissertation:

- How do language and language-related factors shape experiences of communication access and effectiveness of different access practices in DHH communication repertoire?

- How can communication access technologies be designed and built to meet both language needs and access needs of the DHH community?

Chapters 3, 4, and 6 examine three different access practices – speechreading, captioning, and interpreting – across a variety of technological contexts (such as online communication in video calls and in-person communication with remote interpreting and transcription technologies). These chapters not only demonstrate empirically that DHH individuals have multilingual lives, but that language is a critical component shaping access in monolingual contexts as well. I additionally explore how we might design and build technologies to meet both language needs and access needs of the DHH community. Chapters 3, 4, and 6 feature implications for the design and development of speechreading, captioning, and interpreting technologies respectively. Chapter 5 explores the development of sign language dictionaries, studying how we might train models for dictionary retrieval that match disability needs (video-based search) and language needs (dictionary expansion). In this discussion, I take a step back and put these different findings in conversation with each other. Particularly, I begin by discussing dual goals of language access and communication access construed through these works.

7.1 Reflecting on Language Access and Communication Access

7.1.1 On access to language

Through these different chapters, we see that each access practice varies in how it affords access to a language. For example, Chapter 3 explores speechreading – the visemes (mouthshapes) do not have a 1-1 mapping to phonemes (sounds), and therefore do not allow a speechreader to fully track the words spoken in a language. Also recall participant examples of tone in Mandarin and aspiration in Hindi – these phonological components distinguish words from each other, but have no visual counterpart. The degree to which speechreading can afford access to these languages is then constrained by these phonological factors. Similarly, consider Chapter 6 on sign language interpreting – the signing style of the interpretation can influence how much of the original utterance can be reconstructed by a user. Fingerspelling and more "English-like" signing helped participants in contexts where they wanted to track the specific vocabulary used (e.g., science lectures), while other contexts called for more conceptual translations. This degree of reconstruction variance also impacted expressive communication, for example with some participants choosing to voice for themselves. Chapter 4 offers similar insights on captioning – participants remarked that captions did not always capture the pronunciations used by speakers. The degree of mismatch between speech and orthography varied across languages – for example, Traditional Chinese (or simplified Chinese) captions are used for both speech in Mandarin and Cantonese. While this allows for a larger audience, a caption user has to work harder to match the incoming speech to the captions.

Through these examples, we see that *access to language* is a goal that guides DHH individuals' use of access technologies and practices. While research on Deaf education and language acquisition emphasizes the importance of access to language through developmental years, here I am talking about the desire for access to a *specific* language articulated by DHH participants (e.g., see section 4.4.4). Each language has a flavor, a history, a culture, a connection – DHH individuals want access to that without things being 'lost in translation'.

But this desire is complicated by their familiarity and fluency in each language as well as lack of access practices that mediate access to language in its entirety. Language access then impacts how DHH individuals choose between practices and navigate their communication repertoires (i.e., picking between captioning and speechreading and signing). While many of these examples focus on spoken languages, I argue this applies to sign languages as well – particularly to the experiences of DeafBlind and DeafDisabled individuals in the DHH community. My research shows that technologies can be designed for this goal of language access – consider the variations of multilingual captioning presented in Chapter 4 or the sign language dictionary presented in Chapter 5. In affording access to languages and supporting learning, DHH technology has the potential to connect people to their cultures and act as a site of linguistic and cultural preservation – thus enacting language justice.

7.1.2 On access to communication

While language access is an important goal, we see that it is not the only goal motivating and guiding DHH individuals use of their repertoires. Communication is about more than language – in fact, communication is not even bounded by language. Throughout these chapters, we see that DHH individuals creatively switch between languages as needed and combine multiple access practices to build pathways for communication and connection. Access to communication can happen in the absence of access to a specific language – for example, consider the examples of translation and translanguaging in Chapter 4. Recall how a participant moved between signing in ASL and speechreading spoken Chinese and English to communicate with a group of friends – access to communication is about bridging a multitude of communication repertoires. Access to communication also pushes the bounds of languages themselves – consider how DHH individuals develop new signs and backchanneling practices to meet contextual and temporal needs (Chapter 6).

In attending to the communication repertoire holistically, my research unearths what makes each practice uniquely specialized. Speechreading offers visuals and emotional connection (Chapter 3), captioning provides communication ease and synchrony (Chapter 3 and

4), human interpreting works well in contexts with groups and movement (Chapter 6). *Access to communication* for DHH individuals then is about being able to leverage this full repertoire without constraints, switching between languages and access practices as desired.

7.1.3 On these dual goals

I argue that both of these are distinct, but equally important goals to guide the design and development of DHH technology. Technologies for language access could address limitations of each access practice, and work to improve phonetic and linguistic information encoded. Technologies guided by communication access could lean into the specialization of each access practice, and support the use of multiple practices in tandem thus allowing DHH individuals to leverage their repertoires holistically. For example, prioritizing language access with captioning technologies might point us to exploring phonetic transcriptions and other methods to better represent the cadence of this language when spoken. In contrast, prioritizing communication access with captioning technologies might encourage us to explore incorporating emotion and tone (e.g., [63]) or better optimize them for simultaneous use with speechreading. Importantly, DHH individuals' goals are fluid – the same individual at different periods of time or in different contexts might choose to prioritize different goals. Designing technologies that can be customized to address these different goals on the fly would be a valuable way to maintain DHH users' agency.

Lastly, identifying unmet needs and recognizing the ways technology can scaffold existing access practices requires we study DHH communication holistically. Echoing Blommaert and Backus, "A repertoire is composed of a myriad of different communication tools with different degrees of functional specialization. No single resource is a communicative panacea; none is useless." [38]. No single communication access technology can be designed to be a "communicative panacea", but it can be designed to meet previously unmet needs, and to be used in conjunction with existing access practices. Through all my research studies with DHH participants, I consistently began each interview by asking about *all* of the communication practices they use in day to day life, and sought to understand use across multiple

modalities and contexts. I argue that even for works that focus on a specific technology or access practice, this approach is valuable in situating this access practice for the person and understanding their repertoire holistically. Informed by this, we can ensure that designed technologies do not erase the specialization of the access practice and we can evaluate these technologies in ways that better match their real-world use.

7.2 Limitations and Future work

While I have sought to engage with a variety of access practices through my dissertation, there are notable gaps in my exploration of the DHH communication repertoire. I have not explored tactile communication practices (such as Protactile and tactile ASL) or the perspectives of DeafBlind individuals through my work. Similarly, while multiply disabled individuals were not excluded from my studies, my research does not fully engage with how multiple disabilities impact the communication repertoire and its use. Additionally, my work predominantly focuses on receptive communication access needs of the DHH community (with the exception of Chapter 6) – expressive communication needs and practices (such as writing and text-to-speech technologies) are a meaningful component of DHH communication. Examining these all further through the lens of language and language-related factors is a valuable direction for future research.

Most of my research presents qualitative accounts of DHH individuals' experiences of switching between access practices and languages. Quantitative explorations of the same phenomena would deepen our understanding of how to design and build for use of multiple access practices in tandem. For example, behavioral measures (e.g., gaze analysis) of DHH users experiences of video calls where interpreting, captioning, and speechreading are readily available could help us understand the degree of mixed use and create metrics to measure ease of the same.

Building on the concept of language access, another interesting direction for future work could be accessible language learning technologies. Supporting the language learning goals of DHH individuals – for example, learning a sign language or acquiring literacy could mean-

ingfully change the kinds of access practices someone has in their toolbox (e.g., reading captions). Existing technologies focus on non-disabled users, rendering traditional pedagogical approaches like listening exercises or immersion through media inaccessible. One direction could be to incorporate multimodal representations of language into learning tools and spaces (e.g., mouth movements, phonetic transcriptions) and investigate how they support DHH language learners. Emerging AI and language technologies could further allow us to generate customized course content and offer dynamic feedback– personalizing learning to each users' unique goals and access needs. Overall, this direction of research could examine how technology might help people to grow their communicative repertoires (instead of constraining them).

Similarly, AI technologies have the potential to support the goal of *communication access*. Emerging multimodal language models allow users to combine multiple inputs (images, text, speech), and thus allow users to move between different communication practices in their repertoire. Sign language processing research is advancing as well, promising the addition of sign-based inputs to these models. While these advances have the potential to allow users to leverage their full communication repertoire, critical consideration is still required so that the work of languaging is not fully outsourced to technology. Throughout my disseration research, DHH participants have described the thought and care that goes into switching between access practices and languages such that it meets the demands of the context and conversation partners. An important direction of research for these multimodal AI technologies is ensuring users remain active agents in decision making around languaging, i.e., how different modalities and languages are combined.

Finally, while my research has focused on DHH communities, I believe that language and language-related factors deeply influence accessibility experiences of many others with communication access needs– such as neurodivergent, chronically ill, and otherwise disabled individuals. For example, consider multilingual captioning from Chapter 4 – many participants in that study identified with different disabilities. Notably, these differences shaped *how* captioning offered access. Similarly, while not represented in Chapter 5, many with

expressive communication access needs also use sign languages to communicate. For these groups, captioning or interpreting services are not easy to request or mandate, impacting availability and how it fits into their communication repertoires. These differences in access provision combined with language-related factors points to an exciting area for future research: exploring these communities' communication repertoires through a language lens and redesigning access technologies (such as alternative and augmentative communication devices) to account for a range of language experiences.

7.3 Concluding Remarks

Through my dissertation, I have highlighted how language and disability collide in communication accessibility. While many technologies are built to support communication and foster connection, they often constrain multilingual and multimodal communication practices. I am excited for a future where technologies reflect the full diversity of communication – where technology is no longer a site of erasure or oppression, but a space where communities and languages can flourish. I look forward to designing and building technologies that enact access and embody both language justice and disability justice.

BIBLIOGRAPHY

[1] ADA.gov Resources. Ada requirements: Effective communication. https://www.ada.gov/resources/effective-communication/, 2020.

[2] Antena Aire. How to build language justice. https://antenaantena.org/wp-content/uploads/2020/10/AntenaAire_HowToBuildLanguageJustice.pdf, Oct 2020.

[3] Nadine AlBkowr and Ahmad S Haider. Subtitling for people with hearing impairments in the arab world context: The case of the blue elephant 2 movie. *Online Journal of Communication and Media Technologies*, 13(4):e202347, 2023.

[4] Rahaf Alharbi, Angela D Cheong, Jaylin Herskovitz, Robin N Brewer, and Sarita Schoenebeck. "trying to piece it together": Exploring accessible error detection in emerging privacy techniques with blind people. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–20, 2025.

[5] Rahaf Alharbi, Pa Lor, Jaylin Herskovitz, Sarita Schoenebeck, and Robin N Brewer. Misfitting with ai: How blind people verify and contest ai errors. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–17, New York, NY, USA, 2024. Association for Computing Machinery.

[6] Rahaf Alharbi, John Tang, and Karl Henderson. Accessibility barriers, conflicts, and repairs: Understanding the experience of professionals with disabilities in hybrid meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, New York, NY, USA, 2023. Association for Computing Machinery.

[7] Ahmed Ali, Shammur Chowdhury, Mohamed Afify, Wassim El-Hajj, Hazem Hajj,

Mourad Abbas, Amir Hussein, Nada Ghneim, Mohammad Abushariah, and Assal Alqudah. Connecting arabs: Bridging the gap in dialectal speech recognition. *Communications of the ACM*, 64(4):124–129, 2021.

[8] Katherine Allen, Ryan Foley, Molly Glass, Abraham Glasser, Steph Jo Kent, Jeff Shaul, Ludmila Golovine, Winnie Yung-Chung Heh, Natalya Mytareva, Helene Pielmeier, and Bill Rivers. Ai interpreting solutions evaluation toolkit part a: Organization, implementation and management. `https://safeaitf.org/toolkit/`, 2025.

[9] Abdullah Mefareh Almelhi. Understanding code-switching from a sociolinguistic perspective: A meta-analysis. *International Journal of Language and Linguistics*, 8(1):34–45, 2020.

[10] 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*, pages 1–12, 2022.

[11] 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*, pages 1–14, New York, NY, USA, 2022. Association for Computing Machinery.

[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*, pages 44–53, 2023.

[13] Deborah Anderson. Digital vitality for linguistic diversity: The script encoding ini-

tiative. In *Global Language Justice*, pages 220–243. Columbia University Press, New York, NY, USA, 2023.

[14] Stefan Andrei, Lawrence Osborne, and Zanthia Smith. Designing an american sign language avatar for learning computer science concepts for deaf or hard-of-hearing students and deaf interpreters. *Journal of Educational Multimedia and Hypermedia*, 22(3):229–242, 2013.

[15] Jazz Rui Xia Ang, Ping Liu, Emma J. 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*, pages 1–16, 2022.

[16] Robin Angelini, Katta Spiel, and Maartje De Meulder. Speculating deaf tech: Reimagining technologies centering deaf people. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2025.

[17] Verónica Arnáiz-Uzquiza. *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation*. Doctoral dissertation, Universitat Autònoma de Barcelona, 2013.

[18] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan, and Ashwin Thangali. The american sign language lexicon video dataset. In *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 1–8. IEEE, 2008.

[19] Virginie Attina, Denis Beautemps, Marie-Agnès Cathiard, and Matthias Odisio. A pilot study of temporal organization in cued speech production of french syllables: rules for a cued speech synthesizer. *Speech Communication*, 44(1-4):197–214, 2004.

[20] Virginie Attina, Marie-Agnès Cathiard, and Denis Beautemps. Temporal measures of

hand and speech coordination during french cued speech production. In *International Gesture Workshop*, pages 13–24. Springer, 2005.

[21] Katherine Atwell, Danielle Bragg, and Malihe Alikhani. Studying and mitigating biases in sign language understanding models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 268–283, 2024.

[22] Jeremy N Bailenson. Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, mind, and behavior*, 2(1):1, 2021.

[23] Sharon N Barnartt and Richard K Scotch. *Disability protests: Contentious politics 1970-1999*. Gallaudet University Press, 2001.

[24] Grigoris Bastas, Maximos Kaliakatsos-Papakostas, Georgios Paraskevopoulos, Pantelis Kaplanoglou, Konstantinos Christantonis, Charalampos Tsioustas, Dimitris Mastrogiannopoulos, Depy Panga, Evita Fotinea, Athanasios Katsamanis, et al. Towards a dhh accessible theater: Real-time synchronization of subtitles and sign language videos with asr and nlp solutions. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 653–661, 2022.

[25] Sarah CE Batterbury, Paddy Ladd, and Mike Gulliver. Sign language peoples as indigenous minorities: Implications for research and policy. *Environment and planning A*, 39(12):2899–2915, 2007.

[26] H Dirksen Bauman and Joseph Murray. Reframing: From hearing loss to deaf gain. *Deaf studies digital journal*, 1(1):1–10, 2009.

[27] H-Dirksen L Bauman. Audism: Exploring the metaphysics of oppression. *Journal of deaf studies and deaf education*, 9(2):239–246, 2004.

[28] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In

Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pages 610–623, New York, NY, USA, 2021. Association for Computing Machinery.

[29] 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*, pages 161–173, 2018.

[30] Cynthia L Bennett, Amy Pavel, Aidan Bryant, Laura J McCarthy Eberly, Renee Shelby, and Shaun K Kane. "made by people, described by people": The changing work practices of audio description professionals. *ACM Transactions on Accessible Computing*, 18(4):1–28, 2025.

[31] 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. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.

[32] 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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[33] Larwan Berke, William Thies, and Danielle Bragg. Chat in the hat: A portable interpreter for sign language users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–11, 2020.

[34] Patricia Berne, Aurora Levins Morales, David Langstaff, and Sins Invalid. Ten principles of disability justice. *WSQ: Women's Studies Quarterly*, 46(1):227–230, 2018.

[35] Kaushal Santosh Bhogale, Sai Sundaresan, Abhigyan Raman, Tahir Javed, Mitesh M Khapra, and Pratyush Kumar. Vistaar: Diverse benchmarks and training sets for indian language asr. In *Interspeech*, pages 4384–4388, Dublin, Ireland, 2023. International Speech Communication Association.

[36] Manal Binkhonain, Maha Bin Najm, and Ohoud Alharbi. Understanding uses' needs in the development of indoor navigation for visually impaired users in saudi arabia. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7, New York, NY, USA, 2023. Association for Computing Machinery.

[37] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. Language (technology) is power: A critical survey of" bias" in nlp. *arXiv preprint arXiv:2005.14050*, 2020.

[38] Jan Blommaert and Ad Backus. Superdiverse repertoires and the individual. In *Multilingualism and multimodality*, pages 9–32. Brill, 2013.

[39] Matyáš Bohacek and Marek Hrúz. Learning from what is already out there: Few-shot sign language recognition with online dictionaries. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–6, 2023.

[40] Danielle Bragg, Naomi Caselli, Julie A Hochgesang, Matt Huenerfauth, Leah Katz-Hernandez, Oscar Koller, Raja Kushalnagar, Christian Vogler, and Richard E Ladner. The fate landscape of sign language ai datasets: An interdisciplinary perspective. *ACM Transactions on Accessible Computing (TACCESS)*, 14(2):1–45, 2021.

[41] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, et al. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *Proceedings of the 21st international ACM SIGACCESS conference on computers and accessibility*, pages 16–31, 2019.

[42] Danielle Bragg, Kyle Rector, and Richard E Ladner. A user-powered american sign language dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1837–1848, 2015.

[43] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[44] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, 2019.

[45] I Made Budiarsa. Language, dialect and register sociolinguistic perspective. *RETORIKA: Jurnal Ilmu Bahasa*, 1(2):379–387, 2015.

[46] Susan Burch and Alison Kafer. *Deaf and disability studies: Interdisciplinary perspectives*. Gallaudet University Press, 2010.

[47] Teresa Blankmeyer Burke. Choosing accommodations: Signed language interpreting and the absence of choice. *Kennedy Institute of Ethics Journal*, 27(2):267–299, 2017.

[48] Brigitta Busch. The linguistic repertoire revisited. *Applied linguistics*, 33(5):503–523, 2012.

[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*, pages 1–15, New York, NY, USA, 2023. Association for Computing Machinery.

[50] Alison Cardinal, Laura Gonzales, and Emma J. Rose. Language as participation: Multilingual user experience design. In *Proceedings of the 38th ACM International Conference on Design of Communication*, SIGDOC '20, New York, NY, USA, 2020. Association for Computing Machinery.

[51] Alison Cardinal, Emma Rose, Laura Gonzales, Anindita Bhattacharya, Luke Byram, Kenny Coble, Perla Gamboa, Diana Parra, Faaluaina Pritchard, Elsie Rodriguez Paz, et al. From language access to language justice: Creating a participatory values statement for collective action. In *Proceedings of the 39th ACM International Conference on Design of Communication*, pages 38–45, New York, NY, USA, 2021. Association for Computing Machinery.

[52] Anna Cavender, Richard E Ladner, and Eve A Riskin. Mobileasl: Intelligibility of sign language video as constrained by mobile phone technology. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 71–78, 2006.

[53] Jasone Cenoz. Defining multilingualism. *Annual review of applied linguistics*, 33:3–18, 2013.

[54] Center for the Assessment of Sign Language Interpretation. Casli exam content outline and preparation guide, 2024. Retrieved from `https://www.casli.org/exam-preparations/for-nci-candidates/`.

[55] Roberta Cepak. *Subtitling for the deaf and the hard of hearing in Spain: a study on media accessibility, accessibility studies and the standard une 153010: 2012*. Doctoral dissertation, Universitat Pompeu Fabra, 2022.

[56] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. Talktraces: Real-time capture and visualization of verbal content in meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[57] John Lee Clark. *Where I stand: On the signing community and my DeafBlind experience*. Handtype Press, 2014.

[58] Coalition for Sign Language Equity in Technology. Deaf-safe ai: A legal foundation for ubiquitous automated interpreting. https://docs.google.com/document/d/1IqyIqVhckHVcX1gSNNUuuRh-hkPnaua9sHOk1hGy8M4/edit, 2024.

[59] Mairian Corker. Deaf and disabled, or deafness disabled?: towards a human rights perspective. *(No Title)*, 1998.

[60] R Orin Cornett. Cued speech. *American annals of the deaf*, pages 3–13, 1967.

[61] R Orin Cornett. Adapting cued speech to additional languages. *Cued Speech Journal*, 5:19–29, 1994.

[62] 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.

[63] Caluã de Lacerda Pataca, Saad Hassan, Nathan Tinker, Roshan Lalintha Peiris, and Matt Huenerfauth. Caption royale: Exploring the design space of affective captions from the perspective of deaf and hard-of-hearing individuals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024.

[64] 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*, pages 1–15, 2023.

[65] Maartje De Meulder. Is "good enough" good enough? ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22, Virtual, 2021. Association for Machine Translation in the Americas.

[66] Maartje De Meulder, Annelies Kusters, Erin Moriarty, and Joseph J Murray. Describe, don't prescribe. the practice and politics of translanguaging in the context of deaf signers. *Journal of Multilingual and Multicultural Development*, 40(10):892–906, 2019.

[67] Maya De Wit and Irma Sluis. Sign language interpreter quality: the perspective of deaf sign language users in the netherlands. 2014.

[68] Deaf Services Unlimited. Critical care, critical gap: The asl interpreter shortage in medicine. `https://deafservicesunlimited.com/critical-care-critical-gap-the-asl-interpreter-shortage-in-medicine/`, 2025.

[69] Robyn K Dean and Robert Q Pollard Jr. Application of demand-control theory to sign language interpreting: Implications for stress and interpreter training. *Journal of deaf studies and deaf education*, 6(1):1–14, 2001.

[70] Aashaka Desai, Rahaf Alharbi, Stacy Hsueh, Richard E Ladner, and Jennifer Mankoff. Toward language justice: Exploring multilingual captioning for accessibility. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2025.

[71] Aashaka Desai, Lauren Berger, Fyodor Minakov, Nessa Milano, Chinmay Singh, Kriston Pumphrey, Richard Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, and Danielle Bragg. Asl citizen: a community-sourced dataset for advancing isolated sign language recognition. *Advances in Neural Information Processing Systems*, 36:76893–76907, 2023.

[72] Aashaka Desai, Maartje De Meulder, Julie A Hochgesang, Annemarie Kocab, and Alex X Lu. Systemic biases in sign language ai research: A deaf-led call to reevaluate research agendas. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 54–65, 2024.

[73] Aashaka Desai, Jennifer Mankoff, and Richard E Ladner. Understanding and enhancing the role of speechreading in online d/dhh communication accessibility. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, New York, NY, USA, 2023. Association for Computing Machinery.

[74] Aashaka Desai, Daniela Massiceti, Richard Ladner, Hal Daumé III, Danielle Bragg, and Alex X Lu. Investigating dictionary expansion for video-based sign language dictionaries. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 22826–22841, 2025.

[75] Sandy Dietrich, Erik Hernandez, et al. Language use in the united states: 2019, 2022.

[76] Paul Duchnowski, David S Lum, Jean C Krause, Matthew G Sexton, Maroula S Bratakos, and Louis D Braida. Development of speechreading supplements based on automatic speech recognition. *IEEE transactions on biomedical engineering*, 47(4):487–496, 2000.

[77] David M. Eberhard, Gary F. Simons, and Charles D. Fennig. Ethnologue: Languages of the world. `https://www.ethnologue.com/`, 2024.

[78] Communities Creating Healthy Environments. Language justice toolkit: Multilingual strategies for community organizing. `https://static1.squarespace.com/static/5bf21032b98a7888bf3b6e21/t/5f06475412b21d0352040648/1594247018062/language_justice_toolkit.pdf`, 2012.

[79] Solène EVAIN, Benjamin LECOUTEUX, François PORTET, Isabelle ESTEVE, and Marion FABRE. Towards automatic captioning of university lectures for french students who are deaf. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, New York, NY, USA, 2020. Association for Computing Machinery.

[80] Claudio Fantinuoli. Computer-assisted interpreting: Challenges and future perspectives. *Approaches to Translation Studies*, 45, 2017.

[81] Heather A. Faucett, Matthew L. Lee, and Scott Carter. I should listen more: Real-time sensing and feedback of non-verbal communication in video telehealth. 1(CSCW), dec 2017.

[82] Heather A Faucett, Kate E Ringland, Amanda LL Cullen, and Gillian R Hayes. (in) visibility in disability and assistive technology. *ACM Transactions on Accessible Computing (TACCESS)*, 10(4):1–17, 2017.

[83] Geraldine Fauville, Mufan Luo, Anna Carolina Muller Queiroz, Jeremy N Bailenson, and Jeff Hancock. Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. *Available at SSRN 3820035*, 2021.

[84] Jane K Fernandes and Shirley Shultz Myers. Inclusive deaf studies: Barriers and pathways. *Journal of Deaf Studies and Deaf Education*, 15(1):17–29, 2010.

[85] Olivia Foulds. *Investigating How Word Clutter and Colour Impact Upon Learning*, pages 135–151. Springer International Publishing, Cham, 2022.

[86] Neil Fox, Bencie Woll, and Kearsy Cormier. Best practices for sign language technology research. *Universal Access in the Information Society*, 24(1):69–77, 2025.

[87] Gladia. Should you trust wer? `https://www.gladia.io/blog/what-is-wer`, 2024. Accessed: 2024-12-5.

[88] Laura Gonzales. Designing for intersectional, interdependent accessibility: A case study of multilingual technical content creation. *Communication Design Quarterly Review*, 6(4):35–45, 2019.

[89] Google Workspace Updates. Google meet expands language support for closed captioning and translated captions. `https://workspaceupdates.googleblog.`

com/2023/06/expanded-language-support-for-closed-captions-translated-captions-google-meet.html, 2023. Accessed: 2024-09-10.

[90] Google Workspace Updates. Closed caption support in google meet expands to an additional thirty-one languages. https://workspaceupdates.googleblog.com/2024/01/more-languages-for-google-meet-closed-captions.html, 2024. Accessed: 2024-09-10.

[91] Google Workspace Updates. More languages for captions and translated captions in google meet. https://workspaceupdates.googleblog.com/2024/06/more-languages-for-captions-and-translated-captions-in-google-meet.html, 2024. Accessed: 2024-09-10.

[92] Benjamin M Gorman. Reducing viseme confusion in speech-reading. *ACM SIGACCESS Accessibility and Computing*, (114):36–43, 2016.

[93] Benjamin M. Gorman. *A Framework for Speechreading Acquisition Tools*. University of Dundee, 2018. Ph.D. Dissertation.

[94] Benjamin M. Gorman and David R. Flatla. A framework for speechreading acquisition tools. CHI '17, New York, NY, USA, 2017. Association for Computing Machinery.

[95] Benjamin M Gorman and David R Flatla. Mirrormirror: A mobile application to improve speechreading acquisition. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[96] 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*, pages 414–416, 2018.

[97] David M. Grayson and Andrew F. Monk. Are you looking at me? eye contact and desktop video conferencing. *ACM Trans. Comput.-Hum. Interact.*, 10(3):221–243, sep 2003.

[98] Beth G Greene, David B Pisoni, and Thomas D Carrell. Recognition of speech spectrograms. *The Journal of the Acoustical Society of America*, 76(1):32–43, 1984.

[99] François Grosjean. Bilingualism: A short introduction. *The psycholinguistics of bilingualism*, 2(5):5–25, 2013.

[100] François Grosjean. *Bilingual: Life and Reality*. Harvard University Press, Cambridge, MA and London, England, 2010.

[101] Shester Gueuwou, Xiaodan Du, Greg Shakhnarovich, Karen Livescu, and Alexander H Liu. Shubert: Self-supervised sign language representation learning via multi-stream cluster prediction. *arXiv preprint arXiv:2411.16765*, 2024.

[102] John J Gumperz. Linguistic repertoires, grammars and second language instruction. *Languages and Linguistics*, 18:81–90, 1965.

[103] Meng Guo, Lili Han, and Marta Teixeira Anacleto. Computer-assisted interpreting tools: Status quo and future trends. *Theory and Practice in Language Studies*, 13(1):89–99, 2023.

[104] Rinki Gupta. Expanding indian sign language recognition system using class incremental learning. In *2022 International Conference on Advances in Computing, Communication and Materials (ICACCM)*, pages 1–5, 2022.

[105] Gualberto Guzmán, Joseph Ricard, Jacqueline Serigos, Barbara E. Bullock, and Almeida Jacqueline Toribio. Metrics for modeling code-switching across corpora. In *Interspeech 2017*, pages 67–71, Stockholm, Sweden, 2017. International Speech Communication Association.

[106] Brett A Halperin, Gary Hsieh, Erin McElroy, James Pierce, and Daniela K Rosner. Probing a community-based conversational storytelling agent to document digital stories of housing insecurity. In *Proceedings of the 2023 CHI Conference on Human*

Factors in Computing Systems, pages 1–18, New York, NY, USA, 2023. Association for Computing Machinery.

[107] Brett A Halperin and Erin McElroy. Temporal tensions in digital story mapping for housing justice: Rethinking time and technology in community-based design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pages 2469–2488, New York, NY, USA, 2023. Association for Computing Machinery.

[108] Aimi Hamraie and Kelly Fritsch. Crip technoscience manifesto. *Catalyst: Feminism, theory, technoscience*, 5(1):1–33, 2019.

[109] 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*, pages 1–18, New York, NY, USA, 2023. Association for Computing Machinery.

[110] Rebecca Perkins Harrington and Gregg C Vanderheiden. Crowd caption correction (ccc). In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–2, 2013.

[111] Saad Hassan, Matyas Bohacek, Chaelin Kim, and Denise Crochet. Towards an ai-driven video-based american sign language dictionary: Exploring design and usage experience with learners. *arXiv preprint arXiv:2504.05857*, 2025.

[112] Paula Helm, Gábor Bella, Gertraud Koch, and Fausto Giunchiglia. Diversity and language technology: how language modeling bias causes epistemic injustice. *Ethics and Information Technology*, 26(1):8, 2024.

[113] 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, 2023.

[114] Jaylin Herskovitz, Andi Xu, Rahaf Alharbi, and Anhong Guo. Hacking, switching, combining: understanding and supporting diy assistive technology design by blind people. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–17, 2023.

[115] Brian Hie, Hyunghoon Cho, Benjamin DeMeo, Bryan Bryson, and Bonnie Berger. Geometric sketching compactly summarizes the single-cell transcriptomic landscape. *Cell systems*, 8(6):483–493, 2019.

[116] Adeline F Hillier, Claire E Hillier, and David A Hillier. A modified spectrogram with possible application as a visual hearing aid for the deaf. *The Journal of the Acoustical Society of America*, 144(3):1517–1520, 2018.

[117] Christopher Hoadley and Sara Vogel. Autocorrect is not: People are multilingual and computer science should be too. *Communications of the ACM*, 67(2):16–18, 2024.

[118] Megan Hofmann, Devva Kasnitz, Jennifer Mankoff, and Cynthia L Bennett. Living disability theory: Reflections on access, research, and design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[119] 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*, pages 421–430, 2010.

[120] 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, New York, NY, USA, 2023. Association for Computing Machinery.

[121] Joe Huamani-Malca and Gissella Bejarano. Comparing incremental learning approaches for a growing sign language dictionary. In *Annual International Conference on Information Management and Big Data*, pages 97–106. Springer, 2023.

[122] 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*, pages 4617–4626, 2017.

[123] Ryo Iijima, Akihisa Shitara, Sayan Sarcar, and Yoichi Ochiai. Word cloud for meeting: A visualization system for dhh people in online meetings. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21, New York, NY, USA, 2021. Association for Computing Machinery.

[124] Saki Imai, Mert Inan, Anthony B Sicilia, and Malihe Alikhani. Silverscore: Semantically-aware embeddings for sign language generation evaluation. In *Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing-Natural Language Processing in the Generative AI Era*, pages 452–461, 2025.

[125] Robert M Ingram. Sign language interpretation and general theories of language, interpretation and communication. In *Language interpretation and communication*, pages 109–118. Springer, 1978.

[126] 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*, pages 7–11, 2018.

[127] Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E Froehlich. Autoethnog-

raphy of a hard of hearing traveler. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 236–248, 2019.

[128] 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*, pages 241–250, 2015.

[129] Dhruv Jain, Angela Lin, Rose Guttman, Marcus Amalachandran, Aileen Zeng, Leah Findlater, and Jon Froehlich. Exploring sound awareness in the home for people who are deaf or hard of hearing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[130] Puneet Jain. Crip sensorama: Re-imagining xr with people with sensorimotor disabilities through criptastic hacking. In *Proceedings of the 12th Conference on Computation, Communication, Aesthetics & X (xCoAx 2024). Fabrica, Treviso, Italy*, 2024.

[131] JiWoong Jang, Sanika Moharana, Patrick Carrington, and Andrew Begel. "it's the only thing i can trust": Envisioning large language model use by autistic workers for communication assistance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, New York, NY, USA, 2024. Association for Computing Machinery.

[132] Tahir Javed, Janki Atul Nawale, Eldho Ittan George, Sakshi Joshi, Kaushal Santosh Bhogale, Deovrat Mehendale, Ishvinder Virender Sethi, Aparna Ananthanarayanan, Hafsah Faquih, Pratiti Palit, et al. Indicvoices: Towards building an inclusive multilingual speech dataset for indian languages, 2024.

[133] John Lee Clark. Against access. `https://audio.mcsweeneys.net/transcripts/against_access.html`, 2025.

[134] Boban Joksimoski, Eftim Zdravevski, Petre Lameski, Ivan Miguel Pires, Francisco José Melero, Tomás Puebla Martinez, Nuno M Garcia, Martin Mihajlov, Ivan Chorbev, and Vladimir Trajkovik. Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities. *IEEE Access*, 10:40979–40998, 2022.

[135] Hernisa Kacorri, Matt Huenerfauth, Sarah Ebling, Kasmira Patel, and Mackenzie Willard. Demographic and experiential factors influencing acceptance of sign language animation by deaf users. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 147–154, 2015.

[136] Sushant Kafle and Matt Huenerfauth. Evaluating the usability of automatically generated captions for people who are deaf or hard of hearing. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 165–174, 2017.

[137] 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 *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 43–55, New York, NY, USA, 2019. Association for Computing Machinery.

[138] Rie Kamikubo, Abraham Glasser, Alex X Lu, Hal Daumé III, Hernisa Kacorri, and Danielle Bragg. Exploring collaboration to center the deaf community in sign language ai. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–18, 2025.

[139] Harriet Kaplan. Speechreading. In *Seminars in Hearing*, volume 18, pages 129–138. Copyright© 1997 by Thieme Medical Publishers, Inc., 1997.

[140] Karen L. Anderson. Speechreading. `https://successforkidswithhearingloss.com/speechreading/`, 2016. Accessed: 2024-09-10.

[141] Shigeki Karita, Richard Sproat, and Haruko Ishikawa. Lenient evaluation of japanese speech recognition: Modeling naturally occurring spelling inconsistency, 2023.

[142] Naveena Karusala, Aditya Vishwanath, Aditya Vashistha, Sunita Kumar, and Neha Kumar. " only if you use english you will get to more things" using smartphones to navigate multilingualism. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, New York, NY, USA, 2018. Association for Computing Machinery.

[143] 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, 2016.

[144] Elizabeth Keating and Gene Mirus. American sign language in virtual space: Interactions between deaf users of computer-mediated video communication and the impact of technology on language practices. *Language in Society*, 32(5):693–714, 2003.

[145] Lee Kezar, Jesse Thomason, Naomi Caselli, Zed Sehyr, and Elana Pontecorvo. The sem-lex benchmark: Modeling asl signs and their phonemes. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–10, 2023.

[146] Lee Kezar, Jesse Thomason, and Zed Sehyr. Improving sign recognition with phonology. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2732–2737, 2023.

[147] Simran Khanuja, Sandipan Dandapat, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury. GLUECoS: An evaluation benchmark for code-switched NLP. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings*

of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3575–3585, Online, July 2020. Association for Computational Linguistics.

[148] Jung-Ho Kim, Mathew Huerta-Enochian, Changyong Ko, and Du Hui Lee. Signbleu: Automatic evaluation of multi-channel sign language translation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14796–14811, 2024.

[149] Amy J. Ko. Wordplay: an accessible, language-inclusive programming language for all. `https://medium.com/bits-and-behavior/wordplay-accessible-language-inclusive-interactive-typography-e4b9027eaf10`, 2023.

[150] 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, 2020.

[151] Linda Kozma-Spytek, Paula Tucker, and Christian Vogler. Audio-visual speech understanding in simulated telephony applications by individuals with hearing loss. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–8, 2013.

[152] Anjali Krishnan. Language justice and technology. `https://languagejustice.wordpress.com/2020/07/14/language-justice-and-technology/`, July 2020.

[153] Raja Kushalnagar. *Disability, Rights, and Information Technology*, chapter Who owns captioning, pages 182–198. University of Pennsylvania Press, Philadelphia, PA, USA, 2017.

[154] 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*, pages 223–230, 2015.

[155] Raja S Kushalnagar, Anna C Cavender, and Jehan-François Pâris. Multiple view perspectives: improving inclusiveness and video compression in mainstream classroom recordings. In *Proceedings of the 12th international acm sigaccess conference on computers and accessibility*, pages 123–130, 2010.

[156] Raja S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. Accessibility evaluation of classroom captions. *ACM Trans. Access. Comput.*, 5(3), jan 2014.

[157] Raja S Kushalnagar and Christian Vogler. Teleconference accessibility and guidelines for deaf and hard of hearing users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–6, 2020.

[158] 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*, pages 1–53, 2017.

[159] Annelies Kusters, Massimiliano Spotti, Ruth Swanwick, and Elina Tapio. Beyond languages, beyond modalities: Transforming the study of semiotic repertoires. *International Journal of multilingualism*, 14(3):219–232, 2017.

[160] Annelies Maria Jozef Kusters and Maartje De Meulder. Language portraits: Investigating embodied multilingual and multimodal repertoires. *Forum: Qualitative Social Research*, 20(3):10, September 2019.

[161] Paddy Ladd. *Understanding deaf culture: In search of deafhood*. Multilingual Matters, 2003.

[162] Harlan Lane. *When the Mind Hears: A History of the Deaf*. Knopf Doubleday Publishing Group, London, 1989.

[163] Harlan L Lane. *The mask of benevolence: Disabling the deaf community*. Knopf New York, 1992.

[164] Walter S Lasecki, Christopher D Miller, Raja Kushalnagar, and Jeffrey P Bigham. Legion scribe: real-time captioning by the non-experts. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 1–2, 2013.

[165] Gloshanda Lawyer. Moving with intention: Embodying language and disability justice. `https://www.youtube.com/watch?v=iJR-_UVOeTs`, 2023.

[166] Gi-Bbeum Lee, Hyuckjin Jang, Hyundeok Jeong, and Woontack Woo. Designing a multi-modal communication system for the deaf and hard-of-hearing users. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 429–434, 2021.

[167] Sirpa Leppänen and Shaila Sultana. Digital multilingualism. In *The Routledge Handbook of Multilingualism*, pages 175–190. Routledge, New York, NY, USA, 2023.

[168] Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1459–1469, 2020.

[169] W. Li. Research perspectives on bilingualism and multilingualism., 2008.

[170] Anna Lim. Translanguaging practices of a multiethnic and multilingual deaf family in a raciolinguistic world and beyond. *Languages*, 7(4):311, 2022.

[171] Frank R Lin, John K Niparko, and Luigi Ferrucci. Hearing loss prevalence in the united states. *Archives of internal medicine*, 171(20):1851–1853, 2011.

[172] Joseph Lo Bianco. The importance of language policies and multilingualism for cultural diversity. *International Social Science Journal*, 61(199):37–67, 2010.

[173] Dennis Looney and Natalia Lusin. Enrollments in languages other than english in united states institutions of higher education, summer 2016 and fall 2016. In *Modern language association*. ERIC, 2019.

[174] BjÖRn Lyxell and Jerker Rönnberg. Information-processing skill and speech-reading. *British Journal of Audiology*, 23(4):339–347, 1989.

[175] Qingsong Ma, Yvette Graham, Timothy Baldwin, and Qun Liu. Further investigation into reference bias in monolingual evaluation of machine translation. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2476–2485, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[176] 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, 2020.

[177] 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*, pages 1–18, 2021.

[178] Kelly Avery Mack, Emma J. McDonnell, Venkatesh Potluri, Maggie Xu, Jailyn Zabala, Jeffery P. Bigham, Jennifer Mankoff, and Cynthia L. Bennett. Anticipate and adjust: Cultivating access in human-centered methods. in chi conference on human factors in computing systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[179] Kelly Avery Mack, Rida Qadri, Remi Denton, Shaun K Kane, and Cynthia L Bennett. "they only care to show us the wheelchair": disability representation in text-to-image

ai models. In *Proceedings of the 2024 CHI conference on human factors in computing systems*, pages 1–23, 2024.

[180] George Major and Jemina Napier. Interpreting and knowledge mediation in the healthcare setting: What do we really mean by "accuracy"? *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, 11, 2012.

[181] Aron S Marie. Finding interpreters who can "open-their-mind": How deaf teachers select sign language interpreters in hà noi, viet nam. *Sign language ideologies in practice*, 129, 2020.

[182] Dominic W Massaro, Miguel Á Carreira-Perpiñán, David J Merrill, Cass Sterling, Stephanie Bigler, Elise Piazza, and Marcus Perlman. Iglasses: an automatic wearable speech supplementin face-to-face communication and classroom situations. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 197–198, 2008.

[183] Anna Matamala and Carla Ortiz Boix. Accesibility and multilingualism: an exploratory study on the machine translation of audio descriptions. *TRANS: revista de traductología*, 20:11–24, 2016.

[184] Roshan Mathew and Wendy Dannels. Investigating the efficacy of conference room webcams for remote group sign language interpretation sessions. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–4, 2024.

[185] Andrea Britto Mattos and Dario Augusto Borges Oliveira. Multi-view mouth renderization for assisting lip-reading. In *Proceedings of the 15th International Web for All Conference*, pages 1–10, 2018.

[186] Lloyd May, So Yeon Park, and Jonathan Berger. Enhancing non-speech information communicated in closed captioning through critical design. In *Proceedings of the 25th*

International ACM SIGACCESS Conference on Computers and Accessibility, pages 1–14, 2023.

[187] Stephen May. *Language and minority rights: Ethnicity, nationalism and the politics of language*. Routledge, New York, NY, USA, 2013.

[188] Emma McDonnell. Understanding social and environmental factors to enable collective access approaches to the design of captioning technology. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–8, New York, NY, USA, 2022. Association for Computing Machinery.

[189] 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*, pages 1–16, New York, NY, USA, 2024. Association for Computing Machinery.

[190] 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. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.

[191] 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):1–25, 2021.

[192] Rachel McKee. Quality'in interpreting: A survey of practitioner perspectives. *The Sign Language Translator & Interpreter*, 2(1):1–14, 2008.

[193] Joshua McVeigh-Schultz and Katherine Isbister. A "beyond being there" for vr meet-

ings: envisioning the future of remote work. *Human–Computer Interaction*, 37(5):433–453, 2022.

[194] Melanie Metzger. *Sign language interpreting: Deconstructing the myth of neutrality*. Gallaudet University Press, 1999.

[195] Hülya Mısır and Hale Işık Güler. Translanguaging dynamics in the digital landscape: insights from a social media corpus. *Language Awareness*, 33(3):468–487, 2024.

[196] Ross E Mitchell and Michaela Karchmer. Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the united states. *Sign language studies*, 4(2):138–163, 2004.

[197] Ross E Mitchell, Travas A Young, Bellamie Bachelda, 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.

[198] Sara Morrissey and Andy Way. Lost in translation: the problems of using mainstream mt evaluation metrics for sign language translation. In *Proceedings of the 5th SALTMIL Workshop on Minority Languages at LREC*, volume 6, pages 91–98, 2006.

[199] Amit Moryossef, Rotem Zilberman, and Ohad Langer. signwriting-evaluation: Effective sign language evaluation via signwriting. *arXiv preprint arXiv:2410.13668*, 2024.

[200] Mumtaz Begum Mustafa, Mansoor Ali Yusoof, Hasan Kahtan Khalaf, Ahmad Abdel Rahman Mahmoud Abushariah, Miss Laiha Mat Kiah, Hua Nong Ting, and Saravanan Muthaiyah. Code-switching in automatic speech recognition: The issues and future directions. *Applied Sciences*, 12(19):9541, 2022.

[201] Mahin Naderifar, Hamideh Goli, Fereshteh Ghaljaie, et al. Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in development of medical education*, 14(3):1–6, 2017.

[202] Jemina Napier and Meg J Rohan. An invitation to dance: Deaf consumers' perceptions of signed language interpreters and interpreting. In *Supporting Deaf People, 2005; Earlier discussions of preliminary data from this study were presented at the aforementioned online conferences in 2005 and 2006.* Gallaudet University Press, 2007.

[203] Jemina Napier, Robert Skinner, and Graham H Turner. "it's good for them but not so for me": Inside the sign language interpreting call centre. *Translation & Interpreting: The International Journal of Translation and Interpreting Research*, 9(2):1–23, 2017.

[204] National Association of the Deaf. What is captioning? `https://www.nad.org/resources/technology/captioning-for-access/what-is-captioning/`, 2024.

[205] National Deaf Life Museum. American sign language, a language recognized. `https://nationaldeaflifemuseum.omeka.net/exhibits/show/history-through-deaf-eyes/exhibition/awareness--access-and-change/a-language-recognized`.

[206] Gaye H Nicholls and Daniel Ling Mcgill. Cued speech and the reception of spoken language. *Journal of Speech, Language, and Hearing Research*, 25(2):262–269, 1982.

[207] Victoria Anna Sophie Nyst, M Pfau, Markus Steinbach, and Bencie Woll. Shared sign languages. *Sign language*, pages 552–574, 2012.

[208] 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*, page 1108–1120. Association for Computing Machinery, New York, NY, USA, 2020.

[209] 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 *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, New York, NY, USA, 2020. Association for Computing Machinery.

[210] Ricardo Otheguy, Ofelia García, and Wallis Reid. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review*, 6(3):281–307, 2015.

[211] Carol Padden. The deaf community and the culture of deaf people. *Sign language and the deaf community*, pages 89–103, 1980.

[212] Carol A Padden and Tom L Humphries. *Deaf in America*. Harvard University Press, 1988.

[213] Ivan Panović. 'you don't have enough letters to make this noise': Arabic speakers' creative engagements with the roman script. *Language Sciences*, 65:70–81, 2018.

[214] Fred Pelka. *What we have done: An oral history of the disability rights movement*. Univ of Massachusetts Press, 2012.

[215] 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, et al. 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*, pages 1–10, 2018.

[216] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[217] Alastair Pennycook. *Language as a local practice*. Routledge, 2010.

[218] Mary Pietrowicz and Karrie Karahalios. Sonic shapes: Visualizing vocal expression. Georgia Institute of Technology, 2013.

[219] Franz Pöchhacker. Researching interpreting quality: Models and methods. In *Interpreting in the 21st century: Challenges and opportunities*, pages 95–106. John Benjamins Publishing Company, 2008.

[220] Kerri Prinos, Neal Patwari, and Cathleen A Power. Speaking of accent: A content analysis of accent misconceptions in asr research. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1245–1254, New York, NY, USA, 2024. Association for Computing Machinery.

[221] Lorna Quandt. Teaching asl signs using signing avatars and immersive learning in virtual reality. In *The 22nd international ACM SIGACCESS conference on computers and accessibility*, pages 1–4, 2020.

[222] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518, Honolulu, Hawaii, USA, 23–29 Jul 2023. PMLR.

[223] Kevin Rathbun, Larwan Berke, Christopher Caulfield, Michael Stinson, and Matt Huenerfauth. Eye movements of deaf and hard of hearing viewers of automatic captions. *Journal on Technology and Persons with Disabilities*, 5, 2017.

[224] Timothy Reagan. Ideological barriers to american sign language: Unpacking linguistic resistance. *Sign Language Studies*, 11(4):606–636, 2011.

[225] Registry of Interpreters for the Deaf. Video relay service interpreting: Standard practice paper. `https://rid.org/resources/`, 2007.

[226] Thomas Reitmaier, Electra Wallington, Dani Kalarikalayil Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. Opportunities and chal-

lenges of automatic speech recognition systems for low-resource language speakers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17, New York, NY, USA, 2022. Association for Computing Machinery.

[227] Behnam Rezvani Sichani, Mahmoud Afrouz, and Ahmad Moinzadeh. The representation of multilingualism in dubbing and subtitling for the deaf and hard of hearing (sdh). *Multilingua*, 42(5):675–706, 2023.

[228] Silvia Rodríguez Vázquez. Exploring current accessibility challenges in the multilingual web for visually-impaired users. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, page 871–873, New York, NY, USA, 2015. Association for Computing Machinery.

[229] Cynthia B Roy. *Interpreting as a discourse process*. Oxford University Press, 2000.

[230] Dwaipayan Roy, Sumit Bhatia, and Prateek Jain. Information asymmetry in wikipedia across different languages: A statistical analysis. *Journal of the Association for Information Science and Technology*, 73(3):347–361, 2022.

[231] 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*, pages 1–16, 2022.

[232] Elena Ruiz-Williams, Meredith Burke, Vee Yee Chong, and Noppawan Chainarong. My deaf is not your deaf": Realizing intersectional realities at gallaudet university. *It'sa small world: International deaf spaces and encounters*, pages 262–274, 2015.

[233] Debra Russell and Risa Shaw. Power and privilege: An exploration of decision-making of interpreters. *Journal of Interpretation*, 25(1):7, 2016.

[234] Allison Sauppé and Bilge Mutlu. How social cues shape task coordination and communication. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, page 97–108, New York, NY, USA, 2014. Association for Computing Machinery.

[235] Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9:845–874, 2021.

[236] Britta Schneider. Multilingualism and ai: The regimentation of language in the age of digital capitalism. *Signs and Society*, 10(3):362–387, 2022.

[237] Richard K Scotch. " nothing about us without us": Disability rights in america. *Magazine of History*, 23(3):17–22, 2009.

[238] Zed Sevcikova Sehyr, Naomi Caselli, Ariel M Cohen-Goldberg, and Karen Emmorey. The asl-lex 2.0 project: A database of lexical and phonological properties for 2,723 signs in american sign language. *The Journal of Deaf Studies and Deaf Education*, 26(2):263–277, 2021.

[239] 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*, W4A '21, New York, NY, USA, 2021. Association for Computing Machinery.

[240] 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*, pages 1–8, 2020.

[241] Prem Selvaraj, Gokul Nc, Pratyush Kumar, and Mitesh M Khapra. Openhands: Making sign language recognition accessible with pose-based pretrained models across lan-

guages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2114–2133, 2022.

[242] Gary F Simons. Two centuries of spreading language loss. *Proceedings of the Linguistic Society of America*, 4:27–1, 2019.

[243] 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, 2023.

[244] Sins Invalid. Language justice is disability justice. `https://www.sinsinvalid.org/news-1/2021/6/8/la-justicia-de-lenguaje-es-justicia-para-personas-con-discapacidadeslanguage-justice-is-disability-justice`, 2021.

[245] Robert Skinner, Jemina Napier, and Sabine Braun. Interpreting via video link: Mapping of the field. *Here or there: Research on interpreting via video link*, pages 11–35, 2018.

[246] J Logan Smilges. *Crip negativity*. U of Minnesota Press, Minneapolis, MN, USA, 2023.

[247] Adele Smolansky, Miranda Yang, and Shiri Azenkot. Towards designing digital learning tools for students with cortical/cerebral visual impairments: Leveraging insights from teachers of the visually impaired. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '24, New York, NY, USA, 2024. Association for Computing Machinery.

[248] Kristin Snoddon. Brokering understanding: Canadian deaf interpreters' role and practice. *Perspectives*, 34(1):93–108, 2026.

[249] Paul Spence. Disrupting digital monolingualism: A report on multilingualism in digital theory and practice, 2021.

[250] Tereza Spilioti. From transliteration to trans-scripting: Creativity and multilingual writing on the internet. *Discourse, Context & Media*, 29:100294, 2019.

[251] Advaith Sridhar, Roshni Poddar, Mohit Jain, and Pratyush Kumar. Challenges faced by the employed indian dhh community. In *IFIP Conference on Human-Computer Interaction*, pages 201–223. Springer, 2023.

[252] Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, et al. Popsign asl v1. 0: An isolated american sign language dataset collected via smartphones. *Advances in Neural Information Processing Systems*, 36:184–196, 2023.

[253] Kalin Stefanov and Mayumi Bono. Towards digitally-mediated sign language communication. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pages 286–288, 2019.

[254] Cella M Sum, Franchesca Spektor, Rahaf Alharbi, Leya Breanna Baltaxe-Admony, Erika Devine, Hazel Anneke Dixon, Jared Duval, Tessa Eagle, Frank Elavsky, Kim Fernandes, et al. Challenging ableism: A critical turn toward disability justice in hci. *XRDS: Crossroads, The ACM Magazine for Students*, 30(4):50–55, 2024.

[255] Ippei Suzuki, Kenta Yamamoto, Akihisa Shitara, Ryosuke Hyakuta, Ryo Iijima, and Yoichi Ochiai. See-through captions in a museum guided tour: exploring museum guided tour for deaf and hard-of-hearing people with real-time captioning on transparent display. In *International Conference on Computers Helping People with Special Needs*, pages 542–552. Springer, 2022.

[256] 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, 2011.

[257] Agnieszka Szarkowska, Jagoda Żbikowska, and Izabela Krejtz. Strategies for rendering multilingualism in subtitling for the deaf and hard of hearing. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, 13:274–291, 2014.

[258] 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, New York, NY, USA, 2015. Association for Computing Machinery.

[259] Christopher Tester. How american sign language-english interpreters who can hear determine need for a deaf interpreter for court proceedings. *Journal of Interpretation*, 26(1):3, 2018.

[260] The Caviart Group. Job/task analysis for national interpreter certification (nic). `https://www.casli.org/wp-content/uploads/2017/07/NIC-JTA-Report.pdf`, 2017.

[261] Kerstin Monika Tönsing and Gloria Soto. Multilingualism and augmentative and alternative communication: Examining language ideology and resulting practices. *Augmentative and Alternative Communication*, 36(3):190–201, 2020.

[262] Jessica J Tran, Eve A Riskin, Richard E Ladner, and Jacob O Wobbrock. Evaluating intelligibility and battery drain of mobile sign language video transmitted at low frame rates and bit rates. *ACM Transactions on Accessible Computing (TACCESS)*, 7(3):1–26, 2015.

[263] Nina Tran, Paige S DeVries, Matthew Seita, Raja Kushalnagar, Abraham Glasser, and Christian Vogler. Assessment of sign language-based versus touch-based input for deaf users interacting with intelligent personal assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2024.

[264] Nina Tran, Richard E Ladner, and Danielle Bragg. Us deaf community perspectives on automatic sign language translation. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–7, 2023.

[265] Máté Akos Tündik, György Szaszák, Gábor Gosztolya, and András Beke. User-centric evaluation of automatic punctuation in asr closed captioning. 2018.

[266] Graham H Turner. How is deaf culture?: Another perspective on a fundamental concept. *Sign Language Studies*, 83(1):103–126, 1994.

[267] Toon Vandendriessche, Mathieu De Coster, Annelies Lejon, and Joni Dambre. Representing signs as signs: One-shot islr to facilitate functional sign language technologies. *arXiv preprint arXiv:2502.20171*, 2025.

[268] Eva Vanmassenhove, Dimitar Shterionov, and Matthew Gwilliam. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty, editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Online, April 2021. Association for Computational Linguistics.

[269] Aditya Vashistha and Richard Anderson. Technology use and non-use by low-income blind people in india. *ACM SIGACCESS Accessibility and Computing*, 116:10–21, December 2016.

[270] 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*, pages 1–5, 2013.

[271] Vrana, Adele and Sengupta, Anasuya and Pozo, Claudia and Bouterse,Siko. Decolo-

nizing the internetś languages. https://whoseknowledge.org/wp-content/uploads/2020/02/DTIL-Report.pdf, 2020.

[272] Cecilia Wadensjo. *Interpreting as interaction*. Routledge, 2014.

[273] Mike Wald. Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time. In *International Conference on Computers for Handicapped Persons*, pages 683–690. Springer, 2006.

[274] 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):1–25, 2018.

[275] Xu Wang, Li-Fang Xue, and Dan Yang. Speech visualization based on wavelet transform for the hearing impaired. In *2007 International Conference on Wavelet Analysis and Pattern Recognition*, volume 4, pages 1827–1830. IEEE, 2007.

[276] Camilla Warnicke and Sarah Granberg. Interpreter-mediated interactions between people using a signed respective spoken language across distances in real time: a scoping review. *BMC Health Services Research*, 22(1):387, 2022.

[277] Akira Watanabe, Shingo Tomishige, and Masahiro Nakatake. Speech visualization by integrating features for the hearing impaired. *IEEE transactions on speech and audio processing*, 8(4):454–466, 2000.

[278] Sherman Wilcox and Barbara Shaffer. Towards a cognitive model of interpreting. *Topics in Signed Language Interpreting: theory and practice, Amsterdam, Philadelphia: John Benjamins Publishing Company*, 2005.

[279] Rua M Williams and LouAnne E Boyd. Prefigurative politics and passionate witnessing. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 262–266, New York, NY, USA, 2019. Association for Computing Machinery.

[280] Rua Mae Williams and Chorong Park. Cyborg assemblages: How autistic adults construct sociotechnical networks to support cognitive function. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, New York, NY, USA, 2023. Association for Computing Machinery.

[281] Christine T Wolf and Kathryn E Ringland. Designing accessible, explainable ai (xai) experiences. *ACM SIGACCESS Accessibility and Computing*, (125):1–1, 2020.

[282] Ryan Wong, Necati Cihan Camgoz, and Richard Bowden. Signrep: Enhancing self-supervised sign representations. *arXiv preprint arXiv:2503.08529*, 2025.

[283] Marisol Wong-Villacres, Adriana Alvarado Garcia, and Javier Tibau. Reflections from the classroom and beyond: Imagining a decolonized hci education. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[284] World Federation of the Deaf. Statement on use of signing avatars. `https://wfdeaf.org/resources/statement-on-use-of-signing-avatars/`, 2018.

[285] World Federation of the Deaf. Presentation at the sign x ai summit. `https://wfdeaf.org/presentation-at-the-sign-x-ai-summit/`, 2026.

[286] Lei Xie, Yi Wang, and Zhi-Qiang Liu. Lip assistant: Visualize speech for hearing impaired people in multimedia services. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 4331–4336. IEEE, 2006.

[287] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[288] Shakib Yazdani, Yasser Hamidullah, Cristina España-Bonet, Eleftherios Avramidis, and Josef van Genabith. A critical study of automatic evaluation in sign language translation. *arXiv preprint arXiv:2510.25434*, 2025.

[289] Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. Including signed languages in natural language processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360, 2021.

[290] Kayo Yin, Chinmay Singh, Fyodor O Minakov, Vanessa Milan, Hal Daumé III, Cyril Zhang, Alex Xijie Lu, and Danielle Bragg. Asl stem wiki: Dataset and benchmark for interpreting stem articles. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14474–14490, 2024.

[291] Soledad Zárate. *Captioning and subtitling for d/deaf and hard of hearing audiences*. UCL Press, London, UK, 2021.

[292] Isabelle A Zaugg. Language justice in the digital sphere. In *Global Language Justice*, pages 244–274. Columbia University Press, New York, NY, USA, 2023.

[293] Isabelle A Zaugg, Anushah Hossain, and Brendan Molloy. Digitally-disadvantaged languages. *Internet Policy Review*, 11(2):1–11, 2022.

[294] Han Zhang, Rotem Shalev-Arkushin, Vasileios Baltatzis, Connor Gillis, Gierad Laput, Raja Kushalnagar, Lorna C Quandt, Leah Findlater, Abdelkareem Bedri, and Colin Lea. Towards ai-driven sign language generation with non-manual markers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–26, 2025.

[295] Annuska Zolyomi and Jaime Snyder. Social-emotional-sensory design map for affective computing informed by neurodivergent experiences. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–37, 2021.

[296] Annuska Zolyomi and Jaime Snyder. Designing for common ground: Visually rep-

resenting conversation dynamics of neurodiverse dyads. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–33, 2023.

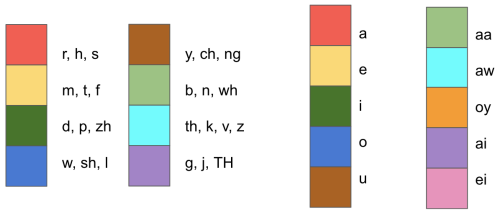
[297] Zoom Support. Enabling automated captions. `https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0058810`, 2024.

[298] Victor Zue and Ronald Cole. Experiments on spectrogram reading. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 116–119. IEEE, 1979.

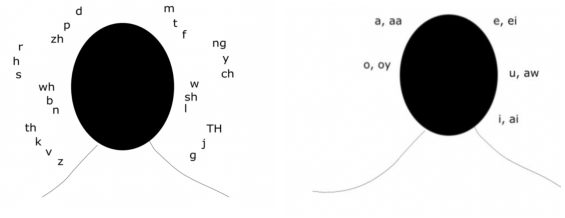
Appendix A

SUPPLEMENTARY FIGURES AND TABLES

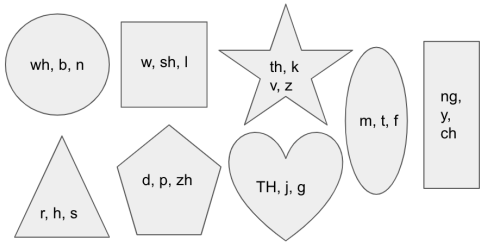
A.1 Design Dimension Representations and Design Probe Ratings



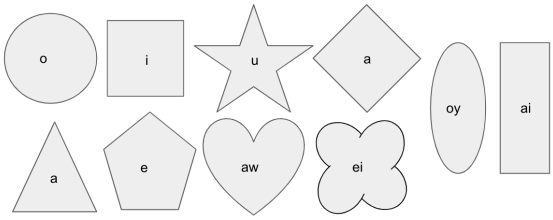
(a) Consonant (left) and Vowel (right) Representations using Color Design Dimensions



(b) Consonant (left) and Vowel (right) Representations using Position Design Dimensions



(c) Consonant Representations using Shape Design Dimensions



(d) Vowel Representations using Shape Design Dimensions

Figure A.1: Consonant and Vowel Representation Using Design Dimensions

Table A.1: Participants' Usefulness Rating and Task Accuracy

Participant	**Average Usefulness Rating**	**Standard Dev**	**Accuracy**
P1	2.3	0.51	77.27
P2	1	0	100
P3	2.5	0.83	100
P4	3	0	95.4
P5	2.6	1.03	100
P6	3.5	0.83	100
P7	4.8	0.4	100
P8	3	0.63	100
P9	1.6	1.03	95.4
P10	3.8	0.44	95.4
P11	3.8	0.98	90.9
P12	2.4	1.67	100

A.2 Core Dictionary Performance

Table A.2 corresponds to section 5.4.1, where we establish a baseline of the classifier performance and our feature-representation approach using the test split of our core vocabulary.

Setup	DCG	Rec@1	Rec@10
Classifier	0.7787 ± 0.0051	0.6221 ± 0.0101	0.8860 ± 0.0020
All Features	0.7719 ± 0.0016	0.6111 ± 0.0052	0.8815 ± 0.0021
Centroid	0.8050 ± 0.0013	0.6604 ± 0.0040	0.9048 ± 0.0010

Table A.2: Establishing baseline performance using classifier and our feature-based approach. The 'all features' setting corresponds to naive use of learned feature representations, and centroid corresponds to calculating centroid for each sign first.

A.3 Adding a Single Sign

Table A.3 summarized results of our experiments using different contributors when simulating our 'long-term contributor' scenario in section 5.4.3.

Setup	DCG	Rec@1	Rec@10
P52 (seed signer)	0.6881 ± 0.0025	0.4955 ± 0.0006	0.8121 ± 0.0070
P33	0.5556 ± 0.0017	0.3275 ± 0.0035	0.6781 ± 0.0014
P11	0.4715 ± 0.0031	0.2482 ± 0.0024	0.5615 ± 0.0043
P50	0.6163 ± 0.0035	0.4004 ± 0.0031	0.7445 ± 0.0056
P27	0.6346 ± 0.0037	0.4215 ± 0.0038	0.7678 ± 0.0038

Table A.3: Dictionary expansion performance when adding a single sign to the dictionary with a single example and using a specific contributor's videos to represent all signs. We see performance varies significantly across chosen contributor.

A.4 Adding Many Signs with Many Examples

Table A.4 corresponds to section 5.4.4, where we experiment with adding many signs to the dictionary using all available samples – it summarizes performance of both added signs and core signs with each stage of expansion (adding 100, 500, 750, 1000 signs). We report averages across each the three randomly sampled sets for each stage of expansion.

Setup	**DCG**	**Rec@1**	**Rec@10**
Adding 1	0.8042 ± 0.0037	0.6538 ± 0.0047	0.9105 ± 0.0040
Adding 100	0.7984 ± 0.0090	0.6433 ± 0.0107	0.9078 ± 0.0084
Adding 500	0.7644 ± 0.0031	0.5949 ± 0.0030	0.8815 ± 0.0040
Adding 750	0.7494 ± 0.0080	0.5730 ± 0.0100	0.8689 ± 0.0093
Adding 1000	0.7397 ± 0.0076	0.5603 ± 0.0101	0.8600 ± 0.0072

Setup	**DCG**	**Rec@1**	**Rec@10**
Original	0.8050 ± 0.0013	0.6604 ± 0.0040	0.9048 ± 0.0010
Adding 100	0.7968 ± 0.0015	0.6484 ± 0.0038	0.8989 ± 0.0018
Adding 500	0.7694 ± 0.0025	0.6090 ± 0.0052	0.8774 ± 0.0022
Adding 750	0.7548 ± 0.0034	0.5882 ± 0.0067	0.8657 ± 0.0030
Adding 1000	0.7436 ± 0.0033	0.5727 ± 0.0057	0.8567 ± 0.0037

Table A.4: Dictionary Expansion performance when adding multiple signs with multiple examples. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. Top rows in each table are baselines from previous settings.

A.5 Adding a Many Signs with Single Example

Tables A.5 and A.6 corresponds to section 5.4.5, where we experiment with adding many signs to the dictionary using a single only one expansion sample. They summarize results with two different setups: using a random contributor's video for each sign vs. using the seed signer's video across the board. We report performance of both added signs and core signs with each stage of expansion (adding 100, 500, 750, 1000 signs). We report averages across each the three randomly sampled sets for each stage of expansion.

Setup	**DCG**	**Rec@1**	**Rec@10**
Adding 100	0.4527 ± 0.0062	0.2203 ± 0.0016	0.5457 ± 0.0163
Adding 500	0.4332 ± 0.0043	0.2016 ± 0.0023	0.5200 ± 0.0094
Adding 750	0.4263 ± 0.0066	0.1971 ± 0.0040	0.5091 ± 0.0114
Adding 1000	0.4192 ± 0.0024	0.1910 ± 0.0013	0.4990 ± 0.0051

Setup	**DCG**	**Rec@1**	**Rec@10**
Adding 100	0.8009 ± 0.0021	0.6543 ± 0.0059	0.9027 ± 0.0009
Adding 500	0.7909 ± 0.0037	0.6402 ± 0.0078	0.8946 ± 0.0006
Adding 750	0.7852 ± 0.0035	0.6327 ± 0.0078	0.8895 ± 0.0010
Adding 1000	0.7792 ± 0.0048	0.6251 ± 0.0085	0.8845 ± 0.0008

Table A.5: Dictionary Expansion performance when adding multiple signs with a single example, and using a random contributor's video as a sample. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. We see large disparities between added and core sign performance at each expansion stage.

Setup	DCG	Rec@1	Rec@10
Adding 100	0.6813 ± 0.0062	0.4877 ± 0.0118	0.8038 ± 0.0057
Adding 500	0.6504 ± 0.0035	0.4477 ± 0.0037	0.7743 ± 0.0086
Adding 750	0.6308 ± 0.0074	0.4231 ± 0.0091	0.7561 ± 0.0088
Adding 1000	0.6164 ± 0.0034	0.4047 ± 0.0055	0.7401 ± 0.0075

Setup	DCG	Rec@1	Rec@10
Adding 100	0.7070 ± 0.0049	0.5205 ± 0.0043	0.8284 ± 0.0064
Adding 500	0.6764 ± 0.0056	0.4808 ± 0.0069	0.8005 ± 0.0092
Adding 750	0.6613 ± 0.0052	0.4614 ± 0.0065	0.7855 ± 0.0091
Adding 1000	0.6477 ± 0.0056	0.4443 ± 0.0077	0.7728 ± 0.0087

Table A.6: Dictionary Expansion performance when adding multiple signs with a single example and using the seed signer's videos to represent all signs. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. We see improvement in added sign performance compared to the random contributor setup, but core sign performance lags behind.

A.6 Small Dictionaries

Tables A.8 and A.7 and A.9 corresponds to section 5.4.6, where we experiment with simulating dictionary expansion for small dictionaries. We report results on three different core-unseen splits: randomly generated (Table A.7), semantically distributed (Table A.8), and specialized (language learning lexicon) (Table A.9). To sample the semantically distributed glosses, we extracted GloVE embeddings [216] for all glosses in ASL Citizen dataset (2731) and used geometric sketching [115] to sample 100 representative glosses. In the following tables, we report performance of both added signs and core signs with each stage of expansion (adding 100, 500, 750, 1000 signs). We report averages across each the three

randomly sampled sets for each stage of expansion.

Setup	**DCG**	**Rec@1**	**Rec@10**
Adding 1	0.8314	0.6722	0.9580
Adding 10	0.7877	0.5821	0.9522
Adding 50	0.7654	0.5735	0.9180
Adding 75	0.7572	0.5714	0.8987
Adding 100	0.7328	0.5285	0.8937

Setup	**DCG**	**Rec@1**	**Rec@10**
Original	0.8302	0.6953	0.9290
Adding 10	0.8311	0.6967	0.9396
Adding 50	0.7925	0.6430	0.9071
Adding 75	0.7783	0.6244	0.8915
Adding 100	0.7625	0.6054	0.8754

Table A.7: Performance with small randomly sampled dictionary, adding multiple signs with multiple examples. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. Top rows in each table correspond to establishing a baseline and adding single signs.

Setup	DCG	Rec@1	Rec@10
Adding 1	0.8230	0.6549	0.9551
Adding 10	0.8300	0.6695	0.9491
Adding 50	0.7544	0.5383	0.9236
Adding 75	0.7481	0.5511	0.8959
Adding 100	0.7511	0.5555	0.9015

Setup	DCG	Rec@1	Rec@10
Original	0.8427	0.7073	0.9498
Adding 10	0.8304	0.6965	0.9409
Adding 50	0.7962	0.6530	0.9086
Adding 75	0.7818	0.6305	0.8980
Adding 100	0.7658	0.6056	0.8841

Table A.8: Performance with small semantically diverse dictionary, adding multiple signs with multiple examples. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. Top rows in each table correspond to establishing a baseline and adding single signs.

A.7 Visual Similarity of Common Confusions

We present some common confusions that occurred for an experimental split under the Base Setup scenario (Section 5.4.2). We present ground truth and predicted glosses alongside phonological labels to describe visual similarity. These phonological labels were derived from ASL-Lex [238] (a database of phonological properties of signs which all ASL Citizen glosses are cross-referenced against).

Setup	DCG	Rec@1	Rec@10
Adding 1	0.8528	0.7083	0.9689
Adding 10	0.8261	0.6601	0.9568
Adding 50	0.8112	0.6446	0.9468
Adding 75	0.7835	0.6016	0.9307
Adding 100	0.7692	0.5818	0.9150

Setup	DCG	Rec@1	Rec@10
Original	0.8255	0.6694	0.9507
Adding 10	0.8291	0.6801	0.9481
Adding 50	0.8007	0.6423	0.9233
Adding 75	0.7863	0.6184	0.9139
Adding 100	0.7749	0.6053	0.8964

Table A.9: Performance with small specialized lexicon dictionary (for language learners), adding multiple signs with multiple examples. Top table is performance of added signs and bottom table is corresponding performance of core signs after dictionary expansion. Top rows in each table correspond to establishing a baseline and adding single signs.

Type	Gloss	Handshape	Movement	Location
Truth	SHOVEL1	s	Curved	Neutral
Predicted	DIGUP	a	Curved	Neutral
Truth	ASSEMBLY	a	Straight	Body
Predicted	ATLANTA	a	Curved	Body
Truth	RING	g	Straight	Hand
Predicted	FILTER	5	Straight	Neutral
Truth	VISUALIZE	s	Straight	Head
Predicted	IMAGINE2	s	Straight	Head
Truth	CALM	closed-b	Curved	Neutral
Predicted	QUIET	closed-b	Curved	Head

Table A.10: Common Confusions from Base Setup Scenario (Adding One Sign With Many Examples)