Understanding and Enhancing The Role of Speechreading in Online d/DHH Communication Accessibility

Aashaka Desai
aashakad@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering
Seattle, USA

Jennifer Mankoff
jmankoff@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering
Seattle, USA

Richard E. Ladner
ladner@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering
Seattle, USA

ABSTRACT

Speechreading is the art of using visual and contextual cues in the environment to support listening. Often used by d/Deaf and Hard-of-Hearing (d/DHH) individuals, it highlights nuances of rich communication. However, lived experiences of speechreaders are underdocumented in HCI literature, and the impact of online environments and interactions of captioning with speechreading has not been explored in depth. We bridge these gaps through a three-part study consisting of formative interviews, design probes, and design sessions with 12 d/DHH individuals who speechread. Our primary contribution is to understand the lived experience of speechreading in online communication, and thus to better understand the richness and variety of techniques d/DHH individuals use to provision access. We highlight technical, environmental and sociocultural factors that impact communication accessibility, explore the design space of speechreading supports and share considerations for the design future of speechreading technology.

CCS CONCEPTS

• **Human-centered computing → Accessibility**.

KEYWORDS

Speechreading, d/Deaf and Hard-of-Hearing, Accessible Video Calls

ACM Reference Format:
Aashaka Desai, Jennifer Mankoff, and Richard E. Ladner. 2023. 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 (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3580810

1 INTRODUCTION

Any language used in communication has semantic (i.e., meaning of words), paralinguistic (i.e., pitch, volume, intonation) and nonverbal (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 [41, 47]. Sign language use is similarly low in other countries due to the deleterious impacts of oralism and audist beliefs [7]. 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 [33]. 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 cues.

Previous research has explored factors that impact speechreading ability [42, 56], visualizations to support speechreading [18, 44, 45, 66], and approaches to improve captioning experiences[9, 21, 24, 34, 36, 39, 46, 51, 57, 61–63]. 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

CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580810

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 [18, 19]. 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.

In the following sections, we describe past work in speechreading and other online communication supports (Section 2), highlighting the lack of inclusion of speechreading in that body of work. Although technological work has attempted to improve both speechreading learning and accuracy, there is a lack of formative work answering the question of what speechreaders actually want in digital tools. Section 3 describes our mixed-methods study (Section 3), which explores communication holistically, studies speechreading supports in a more realistic setting than prior work, and provides an opportunity for d/DHH-led ideation on supporting speechreading amongst other communication strategies used in online tools. Our findings (Section 4) highlight three prominent themes: *technical factors* such as balancing the use of multiple accessibility supports alongside speechreading, *environmental factors* such as online versus in-person speechreading strategies and the impact of languages, and *sociocultural factors* such as the impact of disclosure, visibility, and interdependence on access practices and success. We end by discussing the findings in a larger context (Section 5) including design considerations for speechreading technology and situating our findings in prior work.

2 BACKGROUND AND RELATED WORK

DHH people may communicate using speechreading, sign language, spoken language, written language, and tactile language, alone or in combination. Among these, speechreading has at times been used as a tool to propagate oppression of Deaf language and culture due to audism [7], the belief that those who can speak and hear or act like those who can speak and hear are superior. In 1880, the resolution at the Second International Congress on Education of the Deaf [38] 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. In no way do we condone oppression of a culture or forcing individuals to learn how to speechread. However, access is also about 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 a way 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.

2.1 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 [4]. Kushalnagar and Vogler [37] 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 [35].

However, majority of this work focuses on sign language and captioning. Research to support sign language online includes automating sign language recognition [58], animated (signing) avatars [3, 53], impact of video quality [60], and inclusive videoconferencing for signers [4]. Notably, [4] highlights innate lack of support for visual communication such as experiences of visual dispersion, and negative impact of 2D virtual space on a 3D spatial language.

When focused on captioning, research has worked to increase accuracy [24, 39], enhance caption display online [9, 36] and in augmented reality [29, 31, 51], support topic extraction [32], and improve richness by embedding emotions and punctuation in captions [21, 50, 61]. There are also explorations of group dynamics and context on caption use [46, 57, 63]. In exploring captioning for small-group conversations in in-person and remote contexts, McDonnell *et al.* [46] found participants' experiences are shaped by social, environmental and technical factors such as DHH partners' communication styles, features of videoconferencing software, and delay and accuracy of captions respectively. They argue these factors must be considered together to contextualize the use of captioning technology.

However for speechreading, such exploration of lived experiences is missing from literature, with the exception of [30]. There is increasing work on the use of non-verbal and social cues over video call [15, 22, 55] but does not look at d/DHH individuals specifically. While [56] 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 [29, 36, 54, 59]. Previous works discuss the visual dispersion [36] often experienced by d/DHH people; and eye tracking studies have examined eye movements while using captions in movies or prerecorded videos [54, 59]. How this plays out during video-calls with dynamic social interactions and delayed captions remains unstudied.

2.2 Approaches to Support Speechreading

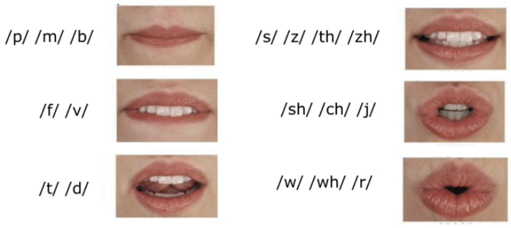


Figure 1: English Phonemes and their Corresponding Mouthshapes (Visemes)

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 [33], 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 [33]. 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 [18, 20] and factors that modulate speechreading ability [42]. Below we describe relevant research for improving speechreading accuracy.

2.2.1 *Cued Speech.*

Cued speech aims to reduce viseme ambiguity in speechreading [12]. Sounds that have the same viseme (Figure 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% [48], and has been adapted to over 50 languages [13]. The timing of cues in relation to speech [6] 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.

2.2.2 *Digital Complements to Speechreading.*

Previous work has explored a variety of visualizations to enhance speech understanding. For example, lip visualizations [45, 66] can mitigate the effects of poor camera angle on speechreading. Other work [52] 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 [26, 64, 67]. Similarly, mappings of phonemic and prosodic features to different textures and colors have been explored [65]. However, these approaches require expert knowledge [67], vary with each utterance and speaker [23], 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.* [14] created an autocueing system to support wider use of cued speech, which was further improved using hyper-realistic animation by Attina [5]. 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.* [44] 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.* [17, 18] 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 visualizations 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).

2.3 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 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 [18, 19]) 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 [46].

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.

3.1 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 communication 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 3.2.

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.

3.2 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.

3.2.1 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 [17]. 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. 2. Like PhonemeViz, our visualization shows these as overlays on videos around the speaker's head [17], which is 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 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 2) circle (as the consonant is in the **b n wh** group, as specified in Table 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. 2, top right) so both images

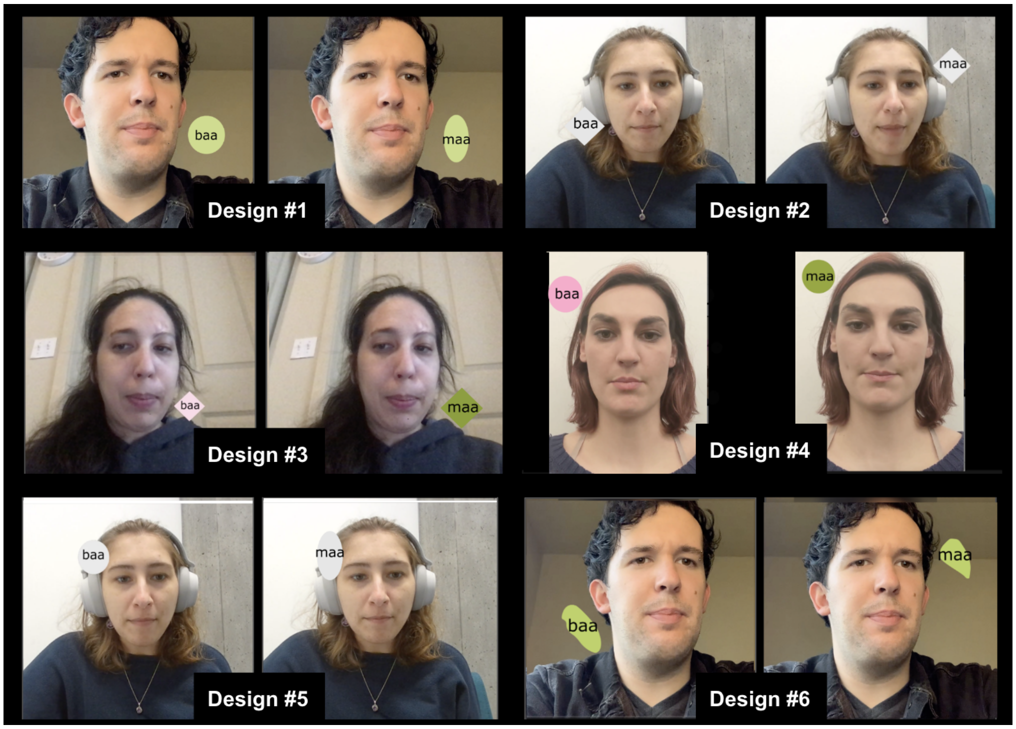


Figure 2: Screenshots of Bottom-up designs applied to syllable 'baa' vs. 'maa'. Note how mouthshape is ambiguous.: Design #1 Consonant Shape, Vowel Color; Design #2 Consonant Position, Vowel Shape; Design #3 Consonant Color, Vowel Shape; Design #4 Consonant Position, Vowel Color; Design #5 Consonant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

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 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.
(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.

3.2.2 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 and are similar to ContextCueView proposed in [18, 19]. 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.

3.2.3 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

Table 1: Group representations for Consonants. Colors, shapes, and positions are chosen to maximize differentiability, but mapping between groups and dimensions is arbitrary.

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 2: Group representations for Vowels. 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: Ordering of speakers and answers by design

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

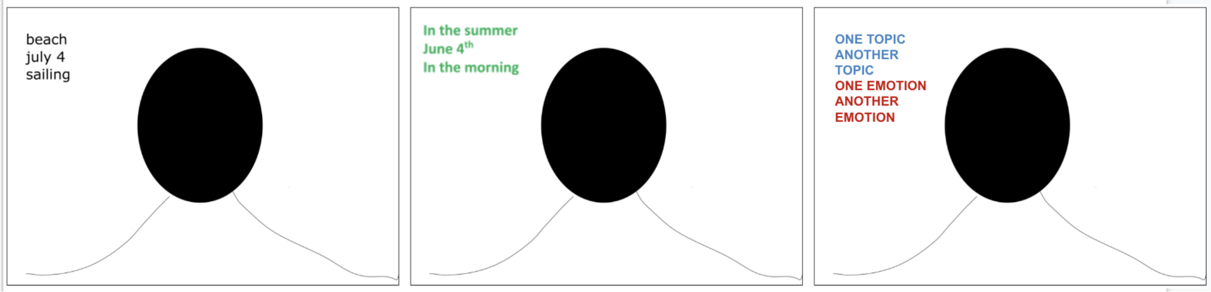


Figure 3: Mock-up Stills of Top-down Probes: (left) keyword information, (middle) temporal information, (right) transition indicators.

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 and shared their impressions of these contextual cues and other alternatives to speechreading support.

3.3 Accessibility Considerations

The study was designed with accessibility to both participants and researchers in mind [43]. 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.

3.4 Analysis Methods

We analyzed the transcripts using reflexive thematic analysis as described by Braun and Clarke [10]. 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.

3.5 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 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.

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 [1, 18, 42]. 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

Table 4: Participant Demographics

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

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)

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 dependant 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.

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.

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 backlit. 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).

4.4 Design Probe Findings

As described in Section 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 5). There are mixed, lukewarm reactions for usefulness, and all designs were rated as distracting (Figure 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 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 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 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 [19]) 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.

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.

4.5.1 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

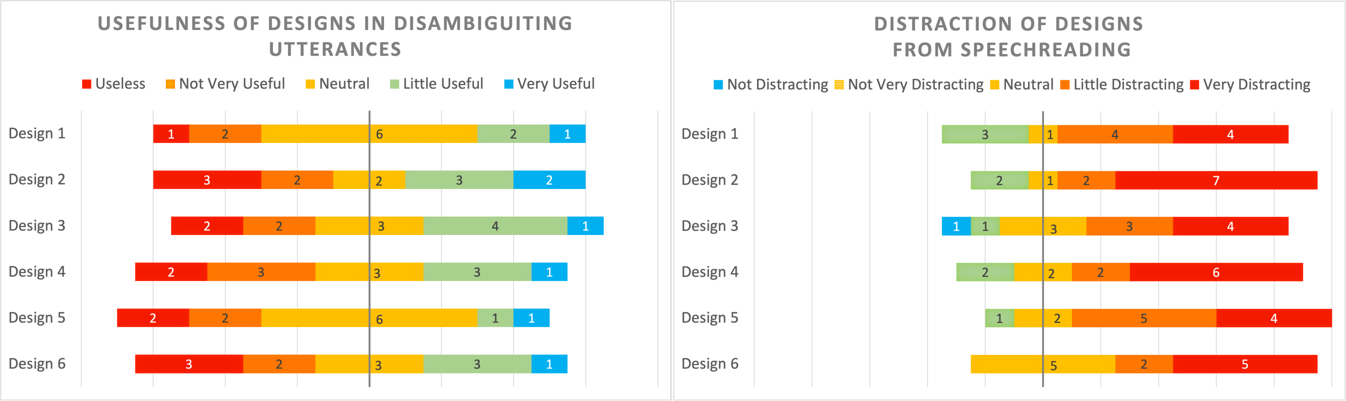


Figure 4: Results of Likert Scales on (left) Usefulness for Bottom-up Designs; the scale "Useless" to "Very Useful" is represented left to right in each bar. (right) Distraction from Speechreading for Bottom-up Design. The scale "Not Distracting" to "Very Distracting" is represented left to right in each bar. The line indicates neutral sentiment.

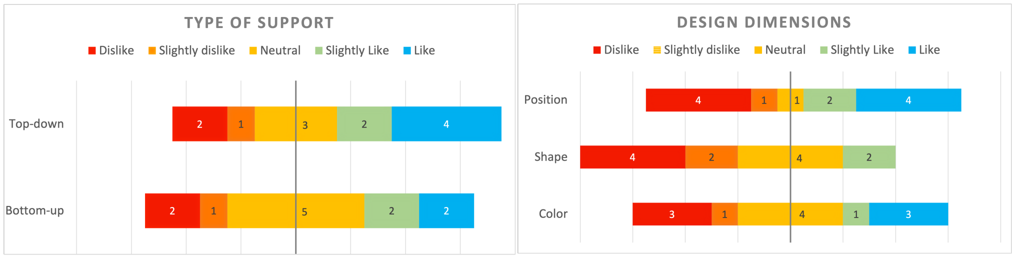


Figure 5: Results of Likert Scales on (left) Type of Speechreading support and (right) Design Dimension in Bottom-up Designs. The scale "Dislike" to "Like" is represented left to right in each bar. The line indicates neutral sentiment.

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).

4.5.2 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.

4.5.3 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. Similarly, 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)

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*.

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 [46], we reflect on technical, environmental and sociocultural aspects of technology design.

Technology's Effect 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 [2].

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 [8]. *"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 [16].

5.2 Summary and design recommendations

In this work, 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 [18, 19] . 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 [25]. 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 [11, 28], dynamic caption placement [27], caption highlight [32], AR [40, 49], emotion embedding [21, 50, 61]), but could benefit from considering speechreading needs. For example [28] 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 [32]. 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 [46, 56] and caption placement [31, 51] 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 [35], 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 [15, 22, 55] and can inform exploration of design spaces described in our work.

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 for this work 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.

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.

ACKNOWLEDGMENTS

This work was supported by the University of Washington Center for Research and Education on Accessible Technology and Experiences (CREATE), and by NSF grants CNS-2137312 and EDA-2009977. We would like to thank Shubhkarman Singh for help with piloting, study participants for sharing their experiences, and the captioners for providing access support. We would also like to thank Megan Hofmann, Kelly Mack, and Emma McDonnell for their support in performing this work.

REFERENCES

[1] 2022. CDC Speechreading. https://www.cdc.gov/ncbddd/hearingloss/parentsguide/building/speech-reading.html.

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

[3] Stefan Andrei, Lawrence Osborne, and Zanthia Smith. 2013. 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 (2013), 229–242.

[4] Jazz Rui Xia Ang, Ping Liu, Emma J. McDonnell, and Sarah Coppola. 2022. "In this online environment, we're limited": Exploring Inclusive Video Conferencing Design for Signers.. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

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

[6] Virginie Attina, Marie-Agnès Cathiard, and Denis Beautemps. 2005. Temporal measures of hand and speech coordination during French Cued Speech production. In *International Gesture Workshop*. Springer, 13–24.

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

[8] Cynthia L Bennett, Erin Brady, and Stacy M Branham. 2018. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 161–173.

[9] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. 2019. Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1â€“6. https://doi.org/10.1145/3290607.3312921

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

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

[12] R Orin Cornett. 1967. Cued speech. *American annals of the deaf* (1967), 3–13.

[13] R Orin Cornett. 1994. Adapting Cued Speech to additional languages. *Cued Speech Journal* 5 (1994), 19–29.

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

[15] Heather A. Faucett, Matthew L. Lee, and Scott Carter. 2017. I Should Listen More: Real-Time Sensing and Feedback of Non-Verbal Communication in Video Telehealth. 1, CSCW, Article 44 (dec 2017), 19 pages. https://doi.org/10.1145/3134679

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

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

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

[19] Benjamin M. Gorman and David R. Flatla. 2017. A Framework for Speechreading Acquisition Tools *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 12 pages. https://doi.org/10.1145/3025453.3025560

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

[21] Michael Gower, Brent Shiver, Charu Pandhi, and Shari Trewin. 2018. Leveraging pauses to improve video captions. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 414–416.

[22] David M. Grayson and Andrew F. Monk. 2003. Are You Looking at Me? Eye Contact and Desktop Video Conferencing. *ACM Trans. Comput.-Hum. Interact.* 10, 3 (sep 2003), 221â€“243. https://doi.org/10.1145/937549.937552

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

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

[25] Jon Henner and Octavian Robinson. 2021. Unsettling Languages, Unruly Bodyminds: Imaging a Crip Linguistics. https://doi.org/10.31234/osf.io/7bzaw

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

[27] Richang Hong, Meng Wang, Mengdi Xu, Shuicheng Yan, and Tat-Seng Chua. 2010. Dynamic captioning: video accessibility enhancement for hearing impairment. In *Proceedings of the 18th ACM international conference on Multimedia*. 421–430.

[28] Ryo Iijima, Akihisa Shitara, Sayan Sarcar, and Yoichi Ochiai. 2021. Word Cloud for Meeting: A Visualization System for DHH People in Online Meetings. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 99, 4 pages. https://doi.org/10.1145/3441852.3476547

[29] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring augmented reality approaches to real-time captioning: A preliminary autoethnographic study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. 7–11.

[30] Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E Froehlich. 2019. Autoethnography of a hard of hearing traveler. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 236–248.

[31] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E Froehlich. 2015. Head-mounted display visualizations to support sound awareness for the deaf and hard of hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 241–250.

[32] Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2019. Evaluating the Benefit of Highlighting Key Words in Captions for People Who Are Deaf or Hard of Hearing. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 43â€“55. https://doi.org/10.1145/3308561.3353781

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

[34] Saba Kawas, George Karalis, Tzu Wen, and Richard E Ladner. 2016. Improving real-time captioning experiences for deaf and hard of hearing students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 15–23.

[35] Linda Kozma-Spytek, Paula Tucker, and Christian Vogler. 2013. 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*. 1–8.

[36] Raja S Kushalnagar, Gary W Behm, Aaron W Kelstone, and Shareef Ali. 2015. Tracked speech-to-text display: Enhancing accessibility and readability of real-time speech-to-text. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 223–230.

[37] Raja S. Kushalnagar and Christian Vogler. 2020. Teleconference Accessibility and Guidelines for Deaf and Hard of Hearing Users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 9, 6 pages. https://doi.org/10.1145/3373625.3417299

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

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

[40] Gi-Bbeum Lee, Hyuckjin Jang, Hyundeok Jeong, and Woontack Woo. 2021. 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)*. 429–434. https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00097

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

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

[43] Kelly Mack, Emma J. McDonnell, Venkatesh Potluri, Maggie Xu, Jailyn Zabala, Jeffery P. Bigham, Jennifer Mankoff, and Cynthia L. Bennett. 2022. 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*. 1–18.

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

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

[46] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 434 (oct 2021), 25 pages. https://doi.org/10.1145/3479578

[47] Ross E Mitchell, Travas A Young, Bellamie Bachelda, and Michael A Karchmer. 2006. How many people use ASL in the United States? Why estimates need updating. *Sign Language Studies* 6, 3 (2006), 306–335.

[48] Gaye H Nicholls and Daniel Ling Mcgill. 1982. Cued Speech and the reception of spoken language. *Journal of Speech, Language, and Hearing Research* 25, 2 (1982), 262–269.

[49] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. *Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-Day Captioning.* Association for Computing Machinery, New York, NY, USA, 1108â€“1120. https://doi.org/10.1145/3379337.3415817

[50] Kotaro Oomori, Akihisa Shitara, Tatsuya Minagawa, Sayan Sarcar, and Yoichi Ochiai. 2020. 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* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 54, 4 pages. https://doi.org/10.1145/3373625.3418032

[51] 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. 2018. Speechbubbles: Enhancing captioning experiences for deaf and hard-of-hearing people in group conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[52] Mary Pietrowicz and Karrie Karahalios. 2013. Sonic shapes: Visualizing vocal expression. Georgia Institute of Technology.

[53] Lorna Quandt. 2020. Teaching ASL signs using signing avatars and immersive learning in virtual reality. In *The 22nd international ACM SIGACCESS conference on computers and accessibility*. 1–4.

[54] Kevin Rathbun, Larwan Berke, Christopher Caulfield, Michael Stinson, and Matt Huenerfauth. 2017. Eye movements of deaf and hard of hearing viewers of automatic captions. *Journal on Technology and Persons with Disabilities* 5 (2017).

[55] Allison Sauppé and Bilge Mutlu. 2014. How Social Cues Shape Task Coordination and Communication. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work &amp; Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 97â€“108. https://doi.org/10.1145/2531602.2531610

[56] Matthew Seita, Sarah Andrew, and Matt Huenerfauth. 2021. Deaf and Hard-of-Hearing Users' Preferences for Hearing Speakers' Behavior during Technology-Mediated in-Person and Remote Conversations. In *Proceedings of the 18th International Web for All Conference* (Ljubljana, Slovenia) *(W4A '21)*. Association for Computing Machinery, New York, NY, USA, Article 25, 12 pages. https://doi.org/10.1145/3430263.3452430

[57] Matthew Seita and Matt Huenerfauth. 2020. Deaf Individuals' Views on Speaking Behaviors of Hearing Peers when Using an Automatic Captioning App. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[58] Kalin Stefanov and Mayumi Bono. 2019. Towards Digitally-Mediated Sign Language Communication. In *Proceedings of the 7th International Conference on Human-Agent Interaction*. 286–288.

[59] Agnieszka Szarkowska, Izabela Krejtz, Zuzanna Klyszejko, and Anna Wieczorek. 2011. 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 (2011), 363–378.

[60] Jessica J. Tran, Ben Flowers, Eve A. Risken, Richard E. Ladner, and Jacob O. Wobbrock. 2014. Analyzing the Intelligibility of Real-Time Mobile Sign Language Video Transmitted below Recommended Standards. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers &amp; Accessibility* (Rochester, New York, USA) *(ASSETS '14)*. Association for Computing Machinery, New York, NY, USA, 177â€“184. https://doi.org/10.1145/2661334.2661358

[61] Máté Akos Tündik, György Szaszák, Gábor Gosztolya, and András Beke. 2018. User-centric evaluation of automatic punctuation in ASR closed captioning. (2018).

[62] Mike Wald. 2006. Captioning for deaf and hard of hearing people by editing automatic speech recognition in real time. In *International Conference on Computers for Handicapped Persons*. Springer, 683–690.

[63] Emily Q Wang and Anne Marie Piper. 2018. Accessibility in action: Co-located collaboration among deaf and hearing professionals. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

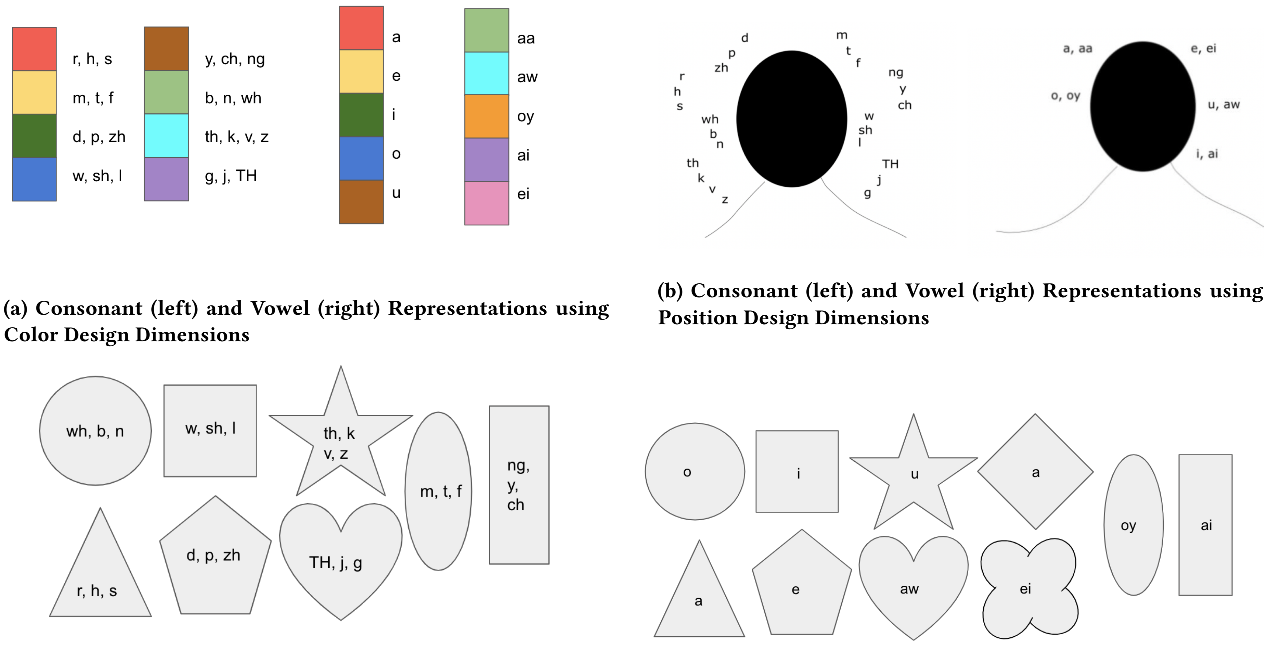
[64] Xu Wang, Li-Fang Xue, and Dan Yang. 2007. Speech visualization based on wavelet transform for the hearing impaired. In *2007 International Conference on Wavelet Analysis and Pattern Recognition*, Vol. 4. IEEE, 1827–1830.

[65] Akira Watanabe, Shingo Tomishige, and Masahiro Nakatake. 2000. Speech visualization by integrating features for the hearing impaired. *IEEE transactions on speech and audio processing* 8, 4 (2000), 454–466.

[66] Lei Xie, Yi Wang, and Zhi-Qiang Liu. 2006. Lip assistant: Visualize speech for hearing impaired people in multimedia services. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 5. IEEE, 4331–4336.

[67] Victor Zue and Ronald Cole. 1979. Experiments on spectrogram reading. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4. IEEE, 116–119.

A 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 6: Consonant and Vowel Representation Using Design Dimensions

Table 5: 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