"It's Not an Issue of Malice, but of Ignorance": Towards Inclusive Video Conferencing for Presenters Who are d/Deaf or Hard of Hearing

JOSH URBAN DAVIS, Dartmouth College, USA
HONGWEI WANG, University of Alberta, Canada
PARMIT K CHILANA, Simon Fraser University, Canada
XING-DONG YANG, Simon Fraser University, Canada

As video conferencing (VC) has become necessary for many professional, educational, and social tasks, people who are d/Deaf and hard of hearing (DHH) face distinct accessibility barriers. We conducted studies to understand the challenges faced by DHH people during VCs and found that they struggled to easily present or communicate effectively due to accessibility limitations of VC platforms. These limitations include the lack of tools for DHH speakers to discreetly communicate their accommodation needs to the group. Based on these findings, we prototyped a suite of tools, called Erato that enables DHH speakers to be aware of their performance while speaking and remind participants of proper etiquette. We evaluated Erato by running a mock classroom case study over VC for three sessions. All participants felt more confident in their speaking ability and paid closer attention to making the classroom more inclusive while using our tool. We share implications of these results for the design of VC interfaces and human-the-the-loop assistive systems that can support users who are DHH to communicate effectively and advocate for their accessibility needs.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing devices**.

Additional Key Words and Phrases: accessibility, video conferencing, telepresence, behavior change

ACM Reference Format:
Josh Urban Davis, Hongwei Wang, Parmit K Chilana, and Xing-Dong Yang. 2023. "It's Not an Issue of Malice, but of Ignorance": Towards Inclusive Video Conferencing for Presenters Who are d/Deaf or Hard of Hearing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 92 (September 2023), 31 pages. https://doi.org/10.1145/3610901

1 INTRODUCTION

Recent studies indicate that hearing ability correlates with employment success for the 20% of adults in the USA who are Deaf or Hard of Hearing (DHH) [9]. People who are DHH are employed at a rate 38% less than their peers who are non-DHH with comparable experience and education. Furthermore, employees who are DHH receive 34% lower salaries and wages on average compared to their non-DHH coworkers [40]. Previous investigations have illuminated a crucial element behind these inequities: effective communication in small-groups is vital to the success of employees with DHH [1, 31] and employees with DHH often struggle when communicating with their peers who are not DHH in professional environments [23]. Communication difficulties in small-groups for people with DHH extend beyond the workplace, including classrooms [44], brain-storming [12], and social engagements [32].

Authors' addresses: Josh Urban Davis, Dartmouth College, Hanover, NH, USA, josh.u.davis.gr@dartmouth.edu; Hongwei Wang, University of Alberta, Edmonton, AL, Canada, hongwei2@ualberta.ca; Parmit K Chilana, Simon Fraser University, Vancouver, BC, Canada, pchilana@cs.sfu.ca; Xing-Dong Yang, Simon Fraser University, Vancouver, BC, Canada, xingdong_yang@sfu.ca.

2474-9567/2023/9-ART92
https://doi.org/10.1145/3610901

In recent years, many of these small-group professionals and educational activities have moved online to remote video conferencing (VC) platforms such as Teams, Meet, and Zoom [22]. Prior research has investigated how to communicate to participants who are DHH on VC platforms such as improved captioning and support for signers [26, 34]. Little work, however, has investigated how systems can be designed to support speakers or presenters who are DHH and verbal. Recent commercial and research tools [1] have explored the use of automation to provide real-time support for VC presenters or feedback for presenters during rehearsal [9, 21]. However, the extent to which these tools are effective at supporting presenters who are d/DHH during real-time VC presentations or during other activities which require small-group communication with mixed-ability groups is an open question.

In this paper, we explore the design of enhancements for VC platforms to better support people who are DHH while presenting or speaking in small group discussions. We first conducted a formative fly-on-the-wall study to examine the accessibility limitations of current VC platforms to support presenters who are DHH. We found DHH participants prefer to conduct a variety of tasks on VC platforms versus in-person spaces because they allow for the use of captions, recordings, and other tools that help them understand the discussion material. However, we also discovered that supporting participants in VC spaces who are DHH and wish to speak still face various difficulties including discreetly communicating and enforcing their accommodation needs, regulating their own speech speed and volume, as well as regulating the speech speed and volume of other presenters. In addition, we found that VC meeting hosts often are unaware of the accommodation needs of participants who are DHH during small-group VC discussions.

To address these limitations, we prototyped *Erato*, a suite of tools to support inclusive video-based communication collected into a Chrome Extension. Erato employs a data-driven approach to provide real-time feedback for speakers who are DHH during VC sessions, as well as a socio-technical support system to allow participants who are DHH in VC discussions to discreetly express and remind participants of their accommodation needs. Our system also provides a subset of tools for meeting hosts to provide at-a-glance updates on the participation of VC discussion members, as well as methods to intervene and make the VC session more inclusive. We deployed Erato in a three-session longitudinal case study examining both DHH and non-DHH participant behaviors over time and validity across common tasks conducted on VC platforms. The principle contributions of this work include:

- A series of design considerations for supporting DHH speakers in VC environments informed by synthesizing the results of prior work as well as our formative surveys, interviews, and formative studies of current VC platform's accessibility limitations.
- The design and implementation of a suit of tools, called Erato, providing data-driven socio-technical support for presenters who are DHH.
- A longitudinal case study evaluation of Erato which showed that inclusive VC system etiquette could be quickly learned by participants and maintained across sessions, improving the accessibility of these systems.

We found that anonymously communicating accommodation needs were embraced by DHH and non-DHH participants alike, and that accommodation requests for etiquette behaviors were easily incorporated into the communication habits of all participants. Furthermore, these behaviors carried over across sessions with little enforcement or reminder required on behalf of the participants. In addition, speaking tools were not only used by speakers with DHH to visualize the rate and volume of their speech, but these same tools could also be used to communicate anonymously to other participants how best to speak for optimal caption performance and lip-reading efficacy. We also discovered how these tools could be applied to not only mitigate inequities of other disabilities, but also the discrimination experienced in VCs by other minorities including women, queer people, and people of color.

2 RELATED WORK

Our research intersects with many existing bodies of interdisciplinary literature including assistive technologies (2.1), video conferencing accessibility tools (2.2), and explorations of accessibility support for presenters in video conferences (2.3).

2.1 Audio-Based Assistive Communication Technologies

Significant prior research has investigated how technology can be designed to support people with disabilities in a variety of tasks, as well as how technology itself can be designed to be more inclusive and accessible [7]. A common approach of these assistive technologies is to substitute one sense for another, such as replicating the experience of sound using haptic feedback. For example, Davis et al. use haptic and audio feedback to guide people with disabilities in prototyping hardware computing circuits [18]. Ubiquitous computing and wearable technologies have also been developed to broaden types and combinations of feedback that can be supported in assistive devices [24] [14]. Other approaches explore using smartphone displays to communicate non-speech sounds because these ambient sounds provide safety, social, and critical environmental information [40] [42]. Similarly, Goodman et al. use smartwatches to evaluate ambient environmental sounds and provide visual and tactile feedback to users who are DHH [23].

Work investigating sound quality and how it can be effectively represented as captions or communicated using multimodal approaches is a vibrant, active field of continuing research. Differentiation of individual sounds from the surrounding environment to amplify the volume of crucial sounds has been explored [12] as well as communicating sound isolation using captions or tactile approaches [36] [10] [27] [31] [30] [54]. Dynamic captions that adapt to the user's field of view have been developed by analyzing eye movement and gazing patterns [10]. Similarly, gaze-adaptive captions are re-positioned based on the viewer's gaze and the objects present on the screen [31]. Information-rich subtitles have also been explored which color code voice volume and speech information to help illustrate the mood/tone of the speaker [27]. Automated speech recognition (ASR) captions used ASR technologies to generate captions for online videos in near-real time [36] [54]. Besides providing captions for English, captions for American Sign Language (ASL) have also been proposed [49]. Other projects propose improving the reliability of captions by communicating the ASR systems confidence in it's own captioning [6]. Often in VC sessions, users who are DHH will be accompanied by an ASL interpreter to help facilitate communication. Seita explored placement of the interpreters' window at the bottom of the screen in VC sessions, similar to where traditional captions would be displayed [51]. Other work has explored the social implications of incorporating interpreters into social software systems [37]. Automating the role of an ASL interpreter is also an active area of research. Elliot et al. explore the affordances of commodity smartphone technologies with their prototype app to accomplish smooth communication between a DHH ASL user and a non-DHH user unfamiliar with ASL [21]. Mixed reality (MR) and virtual reality (VR) methods have been proposed which insert visual cues (highlights on the speaker's table or augmenting the field of view with visual queues) to indicate the current speaker [22]. Head-mounted display (HMD) technologies have also been developed to project an augmented reality (AR) sign language interpreter just outside the view of the user while they are watching a television program to provide automated ASL communication [57]. AR glasses or smart glasses have also been explored to produce real-time captions for users who are DHH [29] [45]. Other approaches employ novel display interfaces to bridge accessibility communication gaps [15]. Kushalnagar et al. employed projectors to directly display real-time captions on top of the current speaker's head in a classroom environment [30]. Other approaches investigate how to use smart-glasses or projectors that insert captions beside the presenter [47] [33], or provide support for people who are not-DHH to communicate using ASL [25].

2.2 Video Conferencing Accessibility Tools

Video Conferencing has recently seen a broad expansion in everyday use-cases ranging from classroom, seminars, panels, corporate meetings, and casual social interactions [17]. This is especially prescient following the COVID-19 outbreak of early 2020, where numerous daily activities hastily moved to VC platforms [32] [20] [39]. As a result, systems expanding the capabilities of video conferencing have emerged. CLIO, for example, incorporates gesture detection and voice commands to enable a presenter to manipulate on-screen media and text [16]. In addition, many accessibility limitations of VC platforms have been documented, such as the importance of users who are DHH to see the full body and face of the presenters [32], to have fewer visual distractions during the session to mitigate cognitive load [39], the extent and limitations of incorporating automated alt-text generation [35], and to have access to reliable and rich captions [20] or interpreters when using ASL [32]. Many techniques have been proposed to address these accessibility concerns, including tools which use machine vision to optimize the furniture arrangement and body placements of users so their bodies and faces can be clearly seen [32] . To reduce the heavy cognitive workload of users who are DHH, researchers have explored consolidating points of interest in VCs to minimize visual distraction [43] [3] [11] [9]. Virtual classrooms and e-learning systems have also been proposed for their ability to bridge accessibility needs by leveraging the interaction capabilities of VR [13] [19]. Other approaches propose visually augmenting VCs, to reduce cognitive overload problem that originates from shifting attention between stimuli [44]. Other techniques explore how users with DHH can effectively insert their ASL interpreter into the meeting [50] or allow human-in-the-loop approaches to improve ASR and captions [5] [6].

VC systems have a unique set of challenges to overcome regarding cognitive overload due to the breadth of visual stimuli common to VC platforms [46]. Also, new difficulties are introduced such as people with DHH being unable to read lips when participants have their cameras turned off or the system automatically hides the person speaking [56]). Significant prior literature documents these concerns and proposes guidelines for etiquette which could make VC activities more inclusive for people who are DHH [46] [49] [8] [34] [2] [58] [56] [52] [39]. Kushalagar et al. suggest that the participants help themselves by designating a person to monitor the chatbox and read it when someone sends something or turns on video only under specific conditions [34]. After conducting a meta-analysis of current literature, Bouzid et al. suggest improving the accessibility by focusing on orienting the camera towards the face, normalizing the incorporation of ASL translators, and redesigning the visual interface [8]. This work also highlights common pain-points engaging accessibility features, such as poor internet bandwidth or the cognitive overload of text-sparse media. Vogler et al. on the other hand, suggest improving accessibility by standardizing software and hardware protocols that account for specific needs [58].

2.3 Accessibility Support for Video Conferencing Presenters

While prior work has explored making video conferencing more accessible for people with DHH, it focuses exclusively on making VC platforms more usable by people with DHH as audience members, and not as speakers, presenters, or session hosts. Prior work explores dynamics in mixed-ability VC enviornments [38]. Other preliminary work focuses on those fluent with ASL or with access to an interpreter [50] [53] [48] [41]. Most closely related to our interests are works which consider supporting presenters, hosts, and speakers who are DHH who might not have access to interpreters, nor be fluent in ASL. The work of Rusnák et al. suggests adding features such as “raise hands/give floor” functions in their VC prototype so that the requests from the audience could be spotted easily by speakers who are DHH [50]. Similarly, CollabAll allows participants to suggest classroom etiquette violations and politely interrupt the instructor for any needs [48]. Seita et al. remotely co-designed accessible features for VCs and found that notifications system influences speaking behaviour and suggest providing options for communication modalities and prioritizing DHH participants’ communication preferences [53]. Prior research also indicates that automated detection tools and notification mechanisms significantly regulate presenters’ behaviour. McDonnell et al. suggest that information-rich captions enabled by adding automatic tools checking speech rate and volume positively impacted the communication of both DHH and non-DHH users during VC sessions [41].

These prior works suggest that VC platforms could provide significant support for people who are DHH speaking and presenting information during a session. In this work, we survey current tools and commercial products for accessibility features, and collect relevant features into a single source [58], we consolidate visual stimuli [46] and provide data to session hosts to allow them to make decisions regrading the accessibility of their session [22] aiming for a decrease in cognitive load [32] [2] [8]. In addition, we co-design new tools to provide means to remind session participants of inclusive etiquette [56] and enable session participants to produce their own accessibility accommodation guidelines according to the needs of the group members. Finally, with these tools consolidated into a single suite we are able to produce rich qualitative insights into their effectiveness and usability by people who are DHH in a usability study.

3 FORMATIVE STUDY

To understand how current VC platforms support the needs of participants who are DHH, we designed a study consisting of surveys, semi-structured interviews, and a fly-on-the-wall observational use-case study. While prior work provided motivation for the problem space, we wanted to better understand of the nuances of the actions and behaviors around pre-existing tools and elucidate actionable problems and decided to use an observational approach [55]. Our investigation probes general pain-points and benefits of VC platforms for DHH users, compares accessibility features available in current commercial platforms, how these platforms mediate and affect communication for mixed-ability groups, and how experiences using VCs change between contexts for DHH people. We chose to include non-DHH participants and mixed-ability populations in our formative study to mimic the mixed abilities conditions common to real world VC use.

3.1 Participants

Of the 26 responses to our survey, 14 self-identified as having one or more disabilities, 7 of which self-identified as DHH. Respondents ranged in age from 18 - 44 with a median age of 26. We interviewed all 14 respondents who self-reported having a disability and experience with at least one VC platform. Of these, 7 identified as cisgender men, 5 identified as cisgender women, 1 identified as two-spirit, 1 identified as demi-male. The 4 of the 7 interviewees who disclosed they are DHH also reported having additional disabilities; ADHD, Autism Spectrum Disorders (ASD), and visual difficulties. Participants occupations were mostly students (71%) with 14% reporting they worked in education, 7% worked in other fields, and 7% reported being unemployed. Participants for our mock-classroom study were chosen based upon their experience in remote classrooms conducted via VCs. Our instructor was an experienced educator who was invited based upon their expertise in teaching as well as their experience teaching both physical and remote classrooms via VCs with mixed-ability student populations.The inclusion criteria for the interview was based on participants' previous experience using VCs with disabilities for a variety of use-cases. All DHH participants with disabilities were interviewed and DHH participants who had previous experience in both VC and physical meeting spaces were prioritized for the live study. Non-DHH participants were also included in our survey/interviews because we wanted to understand more broadly how VC systems could be made more inclusive and accessible, and many of the respondents self-reported disabilities other than DHH. In addition, we chose non-DHH participants for inclusion in our formative study who had prior experience with both VC and physical meeting environments, and we selected an event mix of participants who had prior experience in mixed-ability spaces and those who did not. A table detailing characteristics and demographics of participants included in the live study can be seen in Table 1 and a full table of characteristics for all participants included in the survey and interview study can be found in Appendix B.

Table 1. Participant table for members selected from survey and interview participants for the mock-classroom formative study. Participant noted with a * symbol indicates that this person served as the session host for the study. Fields noted as N/A indicate participant declined to answer this survey question. Full table of participant characteristics included in Survey and Interview portions of the study can be found in Appendix B.

	Self-Described Disability	Age	Gender Identity	Occupation
FP1	Progressive hearing loss	20	Cis-Man	Student
FP2	ADHD	21	Demi-Male	Student
FP2	ASD	26	Cis-Woman	Student
FP4 *	None	N/A	Cis-Woman	University Instructor
FP5	bilateral hard of hearing since birth, low vision	36	Cis-Man	Student

3.2 Methodologies and Procedure

Prior to participation in any portion of the study, participants were provided informed consent to participate as stipulated by our institutions IRB, including participants' rights to refuse participation in any portion of the study that made them uncomfortable, or quit participation in the study at any time. In addition, participants were informed that their identity would be provided anonymity in any subsequent publication resulting from the study, and any quotes would be attributed to a moniker such as P1, P2, etc. The process of informed consent was completed at each portion of the study, with an option to leave the study and remove their data from the collection at any time. We distributed a survey online through social media as well as in partnership with the Accessibility Accommodation Office at our institution. Respondents were asked 7 demographic questions including optional disclosure of any self-described disabilities, 12 short answer questions on their previous experience using VCs including contexts, use cases, and general impressions, followed by 15 Likert scale questions. From the collected responses to our survey we selected 14 participants (7 who self-identified as DHH) for follow-up semi-structured interviews to further expand on responses to the survey. Each interview lasted approximately 60 minutes. All interviews were recorded and transcribed for research purposes except for 4 participants who declined to be recorded. In these cases, hand-written notes were taken of the conversation.

We selected the use-case of a remote classroom experience as it was suggested by the majority of our participants (88%) as a real-world context in the survey feedback. The strong prevalence of this suggested use-case likely occurred as 85% of the survey participants indicated that they were students or working in education. We recruited from both online forums as well as working in coordination with our institutions Office of Accessibility, which would account for the strong student and education demographic presence. Given the deep participant familiarity with the remote classroom context, it was suitable as an environment for our observational study. 5 participants (1 non-DHH instructor, 2 DHH students, and 2 non-DHH students) were selected from our interviewees. Each participant was assigned a short reading (less than 2 pages) from a text on psychology, and asked to present a 3-5 minute summary of the paper contents followed by 3 questions to lead a group discussion on the presentation content. On the day of the study, our instructor facilitated each participant presenting their summary, followed by the presenter leading a discussion based on their presentation with the whole group. After each speaker we distributed a questionnaire asking each participant questions regarding their social experience during the presentation and discussion. Researchers turned their camera and microphones off for the duration of the presentation, observing, taking notes, and posting links to appropriate questionnaires following each presentation. After the final presentation, participants answered a final questionnaire and completed an exit interview with our research team. During exit interviews, researchers discussed questionnaire feedback and observations with the participant, clarifying and elaborating as needed. Exit interviews were scheduled individually based upon participant availability. Interviews were conducted by 3 members of the research team who also performed analysis of the qualitative data together following the conclusion of the interview process. The live VC study took 60 minutes to conduct and the exit interviews each took between 15 and 60 minutes.

3.3 Analysis

Both the survey responses and interview transcriptions were coded for common themes and recurring pain-points mentioned by participants using a grounded methodology. Interviews were conducted by 3 researchers who, after concluding all interviews, gathered to perform a cursory analysis of the data and collect a base set of codes from the qualitative results. From these base codes, intermediate codes were agglomerated from these initial codings by consensus from the 3 researchers. For example, intermediate codes included social awkwardness, feature limitations, feature preferences, and socio-technical dynamics. We then recruited interested respondents for a fly-on-the-wall use-case study to better understand the social dynamics and mental model arising from mixed-ability VC systems. We also collected accessibility support feature preferences and pain points for 3 common VC platforms through our surveys and interviews, extracting common patterns of tension reported by our participants and looked for evidence of these limitations in VC systems. We make special note of whether or not each tool is hidden behind a paywall given the financial inaccessibility findings reported by respondents. Complete findings can be seen in Appendix A.

3.4 Findings

3.4.1 Physical vs VC Spaces. Participants expressed that they preferred VC platforms to physical spaces for lecture-based classroom activities for several reasons. First, VC platforms allow the use of captions which reduce reliance on lip reading for communication. In addition, since VC platforms are video-based, they could be recorded and watched again later in case the student missed part of the lecture due to volume or lip-reading difficulties. The ability to adjust volume was also valued by participants since it provided an aural control affordance not possible in physical spaces. For students who rely on lip-reading, VC platforms encourage participants to face the camera and due to the form factor of most laptops, cameras used in VC communication are generally positioned towards the face. The positioning of the camera increases the likelihood that participants who are DHH will be able to clearly see the face and lips of the speaker, thus reducing the difficulty of lip reading that might result from physical spaces. For these reasons, all participants who are DHH prefer the use of VC platforms over physical spaces for lecture-based classroom activities.

3.4.2 Pain Points. A key pain point of current VC systems is the lack of a robust system for instantiating accessibility accommodation requests during the session. VC platforms lack a method for people who are DHH to request and have their accessibility needs met by a host or other session members. Participants reported that they have to publicly request that their accessibility needs be met, which is embarrassing and uncomfortable, often resulting in participants fore-going any accommodation at all. Our expert instructor corroborated this result, detailing how they would try to accommodate as many requests as possible, especially if a student contacted them in advance, but often these requests would get *"buried in email and I wouldn't be able to read them in time"* (FP4). Furthermore, since remote VC platforms were so rapidly adopted during 2020, common practices for making discussion sessions accessible are still not widely known or enforced as expected social etiquette.

While VC platforms were generally preferred for lecture-based classroom activities, several pain points still arose when discussion-based, brainstorming or activities requiring multiple participants to speak took place via remote VC. Participants reported that monitoring the *"text chat and captions simultaneously was exhausting"* (FP5), and resulted in students with DHH often having to *"re-watch the lecture several times to account for the parts of discussion [they] missed while reviewing chat in the text box"* (FP1). Volume and speech rate also varied from person to person in VC based discussions, and controlling the volume for each individual user is significantly taxing. When speaking themselves, participants with DHH expressed anxiety around their own volume level or speech speed while speaking, indicating a mode of communication regarding oral speech parameters was missing from current VC tool suits. Insecurities around speaking were a common theme among all participants with DHH, expressing that the usual anxieties of public speaking are compounded by difficulties hearing yourself speak.

The instructor also expressed a need for several tools which would help them better facilitate VC sessions. Data regarding who had spoken most recently, how much session participants had spoken, or who hadn't spoken at all, were indicated as data that would be valuable to the instructor in ensuring their VC classroom was inclusive and accessible. Reminders to record or turn on captions at the beginning of the session was also expressed as a valuable feature for the instructor, since these are tools which are available but often go unused. Finally, despite the presence of many accessibility tools such as captioning and video recording/transcription already being commonly included in commercial VC platforms, many of these tools are hidden behind a pay-wall. This creates a barrier of financial inaccessibility, and even when these tools are not hidden behind a pay-wall, many educators and meeting hosts are unaware that accessibility features exist, or the crucial role these tools play in making VC platforms usable by all people.

3.5 Design Goals

Our formative study results along with insights from prior work indicate that the limitations in accessibility evident in VC platforms extends beyond applicability to the classroom use-case explored in our study. Developing additional tools to complement built-in VC capabilities is necessary to support speakers who are DHH, provide necessary data to VC meeting hosts to ensure that their sessions are accessible and inclusive, as well as alter the etiquette of VC sessions to be inclusive in a variety of contexts. Based on the above findings from our formative studies, we synthesize a series of design goals to guide the development of our tool suite and subsequent research.

- **G1.** Facilitate communication of accessibility accommodation needs, and support reminders of these needs sufficient to alter behavior and instate common accessible and inclusive etiquette.
- **G2.** Support speakers in VC sessions who are DHH by providing real-time data-driven feedback.
- **G3.** Provide meeting hosts with data-driven tools to inform their instruction of the class and ensure the space remains inclusive.
- **G4.** Evaluate the accessibility and usability of speech support tools and explore how these tools can be complemented or improved to better meet the needs of end-users who are DHH.

4 ERATO

To address the above considerations, we designed and implemented Erato, a chrome extension for improving inclusiveness in VCs by connecting participants and hosts for accessibility needs of DHH participants and enabling self-improvement with automated tools. Erato is a Chrome Extension that can be installed on a Chrome Browser with version 88 and up. There are two roles when using the software: a participant role and a session host role, each with a set of fundamental features intended to expand the inclusiveness of the VC session. Features incorporated into Erato were agglomerated from multiple sources to create a complete suite of tools for evaluating effectiveness of their support for speakers who are DHH (See G4 in Section 3.4). These sources include tools present in commercial software and state-of-the-art research (which we recreated using prior literature) as well as open source tools. The remaining tools are novel and prototyped according to feedback received during our formative study.

We group Erato's features into three groups following our design goals mentioned in Section 3.4 (where appropriate, we include the source of each tool in the description below): accessibility accommodation requests and etiquette guide in Figure 1 (G1), speaker support features in Figure 2 (G2), and host exclusive features in Figure 3 (G3).

4.1 Accessibility Accommodation Requests/Etiquette Guide

Erato's **Accessibility Accommodation Requests/Etiquette Guide** establishes a channel of communication for meeting hosts and participants intended to communally practice and re-enforce accessibility accommodation requests. The etiquette guide functions on the principle of nudges and persistent reminders by bring the accommodation requests to everyone's attention [4]. Using the anonymous accommodation request interface, participants and the host are able to submit personalized accessibility accommodation requests at any time during the meeting. The requests that participants and hosts enter are anonymous to mitigate any potential hesitancy or embarrassment that may result from making accommodation requests. To ensure the submitted requests are respectful to every participant, practical for the needs of the meeting, and do not contain harmful language, they are reviewed by the host before being presented to session participants. The host can approve, delete, or archive incoming accommodation requests using the host accommodation interface. Approving a request by the session hosts adds the accommodation request to the etiquette guide, visible to all participants. Any requests deleted from the request log by the host are removed from the system. This is intended to prevent trolling or abuse of the system. Along with this presentation, all participants and hosts will receive a one-time notification at the top of the extension to bring this request to the group's attention.

To maintain adherence to the approved accessibility accommodations comprising the etiquette guide, Erato supports a communal reinforcement notification system. Whenever a participant or host want to bring attention or remind the session of a particular request, they are encouraged to use the "bell" (report) button to remind session participants of the importance of following the accessibility accommodation request. The Erato system stores the number of incident reports (i.e. report button clicks) for each accessibility accommodation, and adjusts system behavior to subtly emphasize adherence to accessibility requests with a higher rate of incident report. Once a report button has been clicked, a slide-in notification reminds all session participants of the corresponding accommodation. If multiple accommodation requites are present in the etiquette guide, Erato orders them from most reported to least. The guide is also color coded according to which etiquette items have been reported most recently. Items reported within the past 5 minutes are colored red, with opacity of the red coloring increasing with each subsequent report triggered within a 5 minute interval. This red background opacity decreases for each minute after a 5 minute interval without subsequent report button clicks, returning to white after after 15 have elapsed without report. The report feature is also anonymous as the person reporting the violation is not named in notifications, nor is their request stored by the system in any way. It is also worth noting that this notification is not sent to a specific participant, which not only underlines the requests to all participants but also reduces the pressure on any member of the session not directly adhering to the accommodation request.

All approved accommodation requests are stored in the Accessible Accommodation sections until the database is flushed. This forms a small etiquette guide for a subsequent series of meetings and relieves repetitive submission of the request whenever the meeting begins. The requests are stored in order of the report count with the consideration of the time: the request will be ranked higher if it has a large number of people reporting its violation, and the request will be ranked higher if it is newly submitted when it has the same number of report counts as its peers. The higher requests will show a darker red to emphasize their importance and catch participants' and hosts' attention. The combination of accommodation requests and reminder system is intended to encourage communal reinforcement of accessible session etiquette.

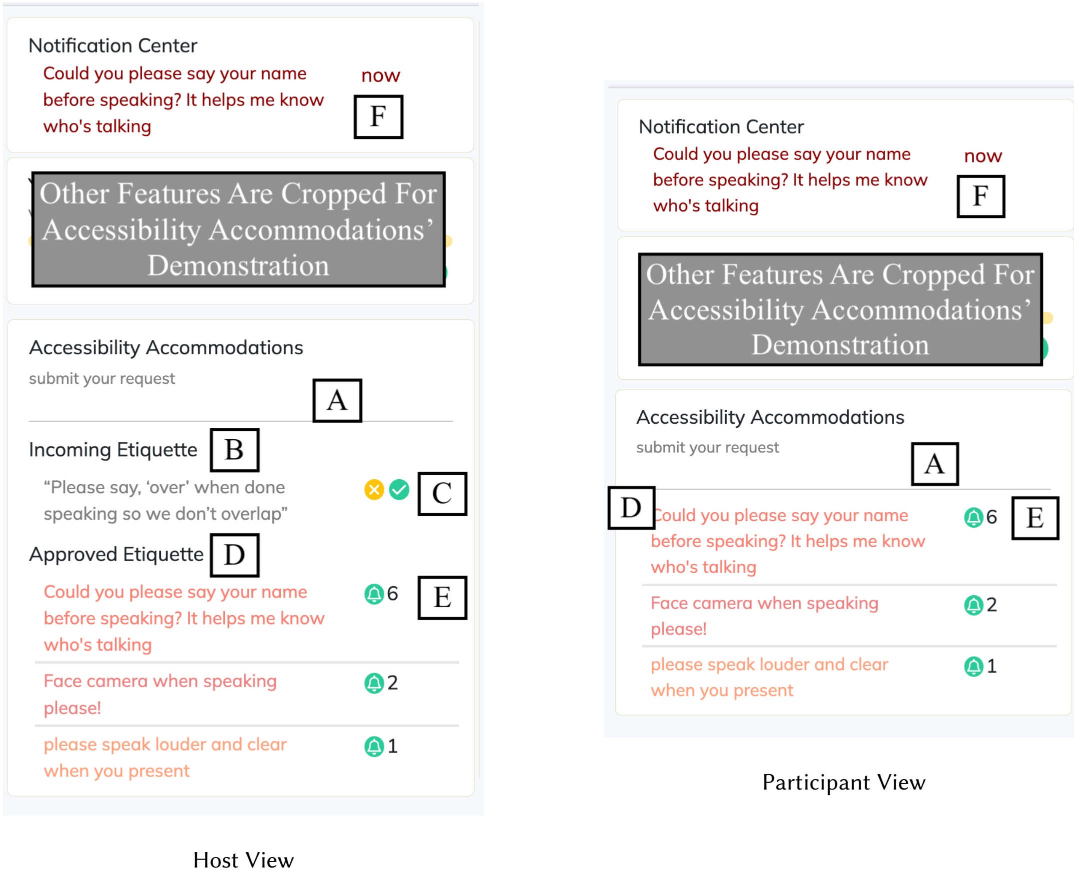


Fig. 1. Overview of Erato Accessibility Accommodation Tools: (A) Participants and hosts submit anonymous requests in the text input. (B) Submitted requests are listed for the host to approve or reject. (C) Buttons allowing hosts to approve or reject submitted requests. (D) Approved requests are listed for all participants. (E) Participants and hosts can report a violation of the request, the number next to the bell indicates the count of violations reported by participants during the session. (F) Recently approved requests and reported etiquette violations will appear as separate notifications in the Notification Center.

4.2 Speaker Support Features

In addition to the etiquette guide and behavior reinforcement system, session participants are provided a collection of tools to encourage the inclusiveness of meeting sessions and to assist with their speaking. Session participants can check their real-time spoken volume, monitor their speech speed throughout the session, and listen to their voice using direct audio playback. The direct playback feature was specified as a common need by P3 of our formative study because VC systems often make it difficult for people using hearing aids or cochlear implants to hear themselves speak. The playback feature allows participants to play their own spoken audio through the aux cable directly connected to their assistive devices.

Participants and hosts can monitor their real-time volume by glancing at the **volume meter** that has three magnitudes: the green bar in the middle, which indicates the volume is comfortable for others to hear and the two yellow bars on each side, which shows the volume is too soft or too loud. Similarly, a **speech speed meter**

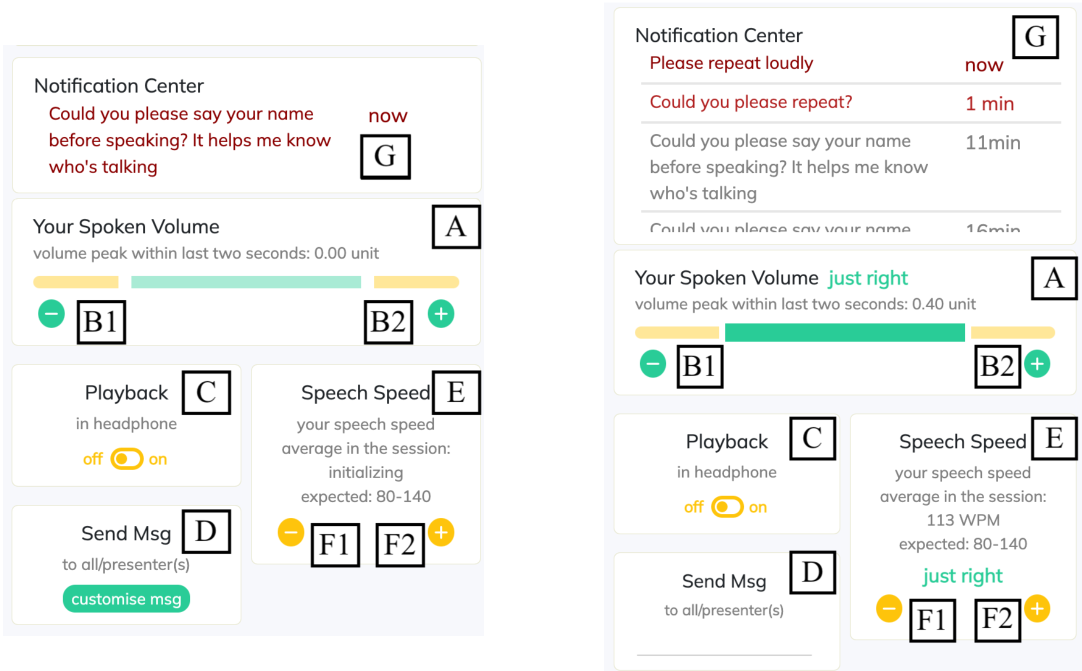


Fig. 2. Additional Participants' and Hosts' Features (A) Participants and hosts can monitor their spoken volume by glancing at the textual feedback and speech volume meter. Numerical measurement of speech volume also displayed for transparency. (B) Buttons to provide direct feedback to speaker regarding spoken volume. (C) Playback participant's or host's voice into headphones. (D) Push a one-time message into all participants and hosts' notification center. (E) Participants and hosts can monitor their speech rate by glancing at the textual feedback. Numerical measurement of speech speed also displayed for transparency. (F) Buttons to provide direct feedback to speaker regarding speech speed. (G) Notifications are displayed in the notification center.

is located adjacent to the volume meter and provides similar feedback to the volume meter. The speech volume and speed recognition systems use Mozilla speech recognition API to measure the number of words spoken in a sentence divided by the total time spoken, producing an approximate calculation of the average words spoken per minute. We filter out sentences that have few words or take a short time to speak to eliminate detection of filler words and sounds. A separate notification is sent by they system if a substantial number of filler words are detected during a given session. Default values were informed by prior literature on ideal speech speed and volume, as well as ranges employed by automated speech volume and speed feedback tools incorporated into commercial products such as Microsoft Cameo for PowerPoint [53]. Speech volume and speed magnitude ranges are adjusted when audience members use the speech speed and human volume buttons located below each respective tool. Pressing the increase or decrease volume or speed buttons shifts the magnitude ranges in the respective direction, allowing audience members to directly communicate impedance desires to the speaker. This design is intended to provide a human-in-the-loop adjustment to the automated feedback tools, allowing speakers to quickly and peripherally receive feedback on their speaking from the audience.

In addition to glancing at the meter, speakers can receive at-a-glance textual feedback on the top of this section with the same colour code as the volume meter. Green indicates that they are speaking within the desired volume and speed ranges and yellow indicates that the speaker may need to adjust their volume or speed. Participants and hosts can get more specific and accurate feedback by looking at the numerical value displayed on the volume and speech meters. This numerical information is brief to avoid distracting the presenter, and is intended to provide complete data transparency that could be useful for speech preparation. The color coding and magnitude design choices are intended to reduce the cognitive load on the speaker, and provide needed information at-a-glance.

The **Playback** feature plays the audio recorded by the microphone synchronously into the participant or host's earphone. This feature was directly requested during the interviews of our formative study by a DHH participant who indicated that using an audio cable directly plugged into their hearing assistive devices gave them greater control over the ability to hear their own voice when speaking. Without such a feature, people using hearing aids often unintentionally taper their volume due to difficulty recognizing their own spoken volume levels. Participants in our formative study noted that they often used third-party software to mitigate this issue, which often was not compatible with different VC systems. We include this feature within Erato to create a seamless collection of tools for supporting speech in VCs.

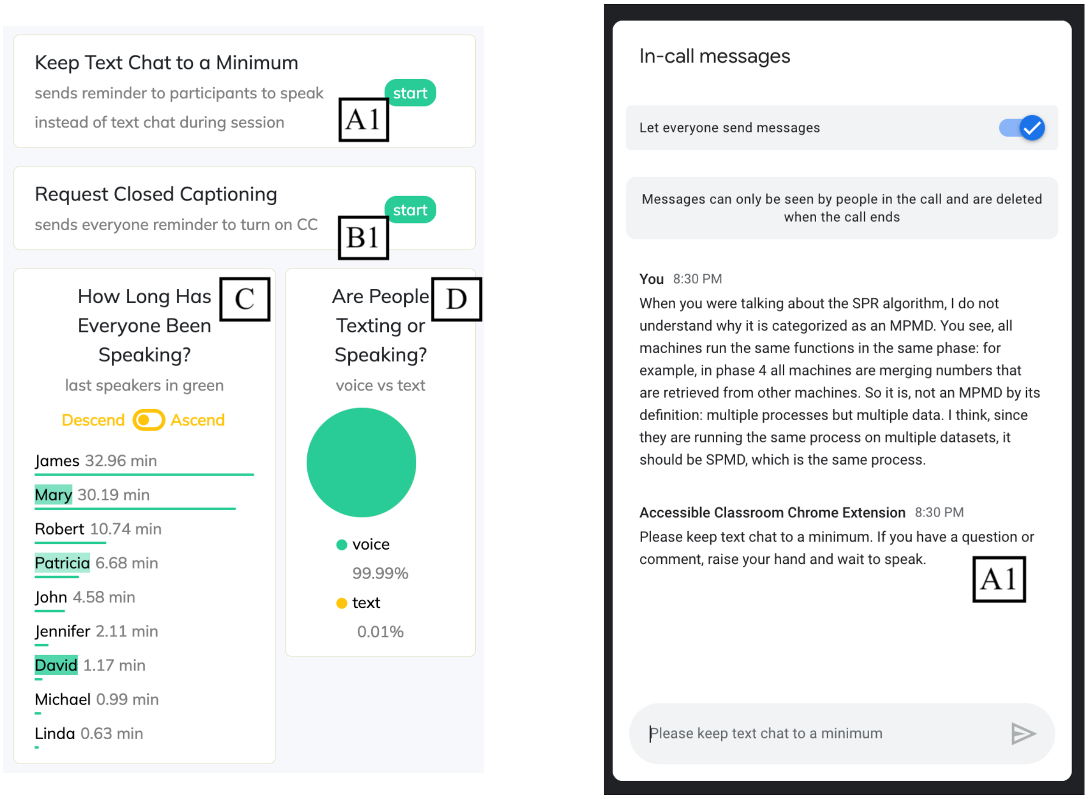
Notification Center gathers all the notifications sent by the participants and hosts into a single centralized location. Notifications are displayed in chronological order of their submission, and repeat message requests (e.g. multiple etiquette reminders generated by clicking the "bell" button, or multiple human volume button clicks) are filtered into a single notification to simplify the display and reduce cognitive load.

4.3 Host Exclusive Features

Erato also has features that are exclusive to the meeting host and provide the host with sufficient data to enable their session to be as inclusive as possible. Some tools are toggles which can be requested by session participants, but need to be triggered by the host. For example, triggering the **Keep Text Chat to a Minimum** toggle switch will automatically input into the text box automated replies should participants send text chat in the text messages field. This text message reminds the participant to keep their text chatting to a minimum and orally communicate when appropriate. The automated response will only be visible to the message sender, encouraging the sender. This feature was co-designed during our formative study by participants who were DHH and relied on speaker captions during the session to remain informed of session activities. Participants reflected that the cognitive load of reading both the captions and text messages was exhausting, and often resulted in the user becoming lost by splitting their focus between the speech captions, which often disappeared within a few moments, and the text message box. This tool is designed to mitigate cognitive load by giving session participants the opportunity to focus on captioned speech, should they desire.

To complement this, session hosts are provided data-driven tools to inform their understanding of session participant engagement and communication preferences. One panel indicates **Are People Texting or Speaking** to give an intuitive idea of how much the voice or text chat is being employed during the current session. Erato converts the text that participants and hosts typed in the In-Call Message into the number of words input, and compared this to the number of words spoken by participants during the session. This data is reflected in a pie chart to quickly provide the session host at-a-glance data regarding the communication modalities being employed in the session.

Similarly, **How Long Has Everyone Been Speaking**, provides the instructor an intuitive understanding of session participation by individual members of the session. Duration of participant speech is measured and represented using a bar chart for quick comparison of participant activity. The length of the bar chart is standardized by the participant who speaks the most, and visual adjustments are made accordingly. The dashboard also indicates to the host the last three speakers who spoke, indicated by a green highlight. Bar chart order can be ranked in ascending or descending order according to the needs and preferences of the host.



Host Features Dashboard

Text Notification

Fig. 3. Host Exclusive Features: (A1) Toggle switch to discourage participants from using the text chat to communicate. Once turned on, an automated text notification will appear in the chat display if Erato detects that participants are using the text chat. (B) Participants and hosts are reminded every 40 seconds to turn on their closed captions unless Erato detects that the participant or hast has captions turned on. (C) Hosts can monitor the amount of time each person has spoken during the session, as well as see visual indicators in green of who has spoken most recently. (D) hosts can see the ratio of communication occurring via voice or text usage in this pie chart.

5 CASE STUDY

We conducted a case study to understand the usability and accessibility of speech support tools comprising Erato (See Design Goal G4 in Section 3.4). We also wanted to examine how these tools affect social dynamics in 3 common remote VC contexts with mixed-ability groups of people. Our goal was to understand what socio-technical dynamics emerged by incorporating Erato into common VC contexts and how those differed from the behaviors observed during our formative study. We used a longitudinal approach for our study to compare how these dynamics and behaviors evolved over time and across contexts.

5.1 Participants

We recruited 12 participants for our study: 6 participants (3 DHH) participated in all 3 sessions while the other 6 participants (3 DHH) were incorporated in pairs for 1 session per pair. This way, each session comprised

Table 2. Participant table for members selected for the Erato prototype case study where P Number indicates anonymized participant number. Participant noted with a * symbol indicates that this person served as the session host for at least one session of the study.

	Self-Described Disability	Age	Gender Identity	Occupation
P1	None	23	Cis-Woman	Software Developer
P2	Deaf, Monovision	40	Cis-Man	Student
P3 *	None	32	Cis-Woman	Illustrator and Teacher
P4	None	28	Cis-Man	Student
P5	None	32	Cis-Man	Vocalist
P6 *	Progressive hearing loss	22	Cis-Man	ESL Instructor
P7 *	Bilateral hard of hearing, since birth	36	Cis-Woman	Unemployed
P8	Deaf, Low vision/vision problems	57	Trans-Man	Unemployed
P9	deaf/hard of hearing and chronic illness	24	Cis-Woman	Retail
P10	ADHD/deaf	19	Trans-Woman	Student
P11	None	24	Cis-Woman	Software Engineer
P12	None	30	Gender queer	Writer and English Instructor

8 participants (4 DHH), with 6 participating in all 3 sessions and 2 new members participating in only one session. This was designed to mimic real-world conditions since new members joining group activities are common in online VC contexts. All participants were compensated for their time. Participants were recruited by a survey distributed online and through email in partnership with the Accessibility Accommodation Office of our institution. Similar to the questionnaire of our formative study, this initial survey collected demographic information as well as short-answer and Likert data regarding respondent previous experience using VCs. We included 12 participants in a fly-on-the-wall case study based upon criteria pertaining to their experience with VC platforms and mixed-ability groups. Specifically, participants who had extensive experience performing both professional and social tasks on VC platforms were prioritized, especially those who were DHH. We intentionally included non-DHH participants with a mixture of familiarity and experience working with mixed-ability groups. Many non-DHH participants reported being completely unaware of the needs of DHH people, and were included to evaluate how inexperienced participants interact with DHH participants while using the tool. We chose to include DHH participants as well as non-DHH participants to mimic real world, mixed-abilities conditions, observe how non-DHH participants responded to accommodation requests (especially those unaware of common DHH needs and inclusive etiquette), and to evaluate any difficulties non-DHH folks may have when using the system. DHH participants were prioritized in the selection and scheduling process and non-DHH participants meeting the above criteria were included based upon their availability given the schedules of our DHH participants. A full description of participant characteristics and demographics can be seen in Table 2.

5.2 Methodology

We followed the exact same procedure for informed consent as documented in our formative study in adherence to the ethics policy provided by our institutions IRB (See Section 3). Each of the study's 3 sessions took place on a different day over the course of a week to reflect real-world conditions and provide information on how behaviors changed over time. Tasks were chosen based on their common prevalence among the use-cases reported in the qualitative feedback from our formative and case study surveys. These tasks also required participants to use various external tools common in VC contexts such as screen sharing, slide presentation, and whiteboard/collaborative brainstorming tool use. All sessions were recorded with the informed consent of the participants and transcribed for analysis. We averaged Likert scale responses to all surveys and participant speaking time and researcher notes were recorded during each session. Qualitative data was analyzed by a single member of the team for themes which appeared in the formative study, as well as novel commonalities in behavior or themes emerging from the feedback. In this study, we embrace the approach of Yin et al. and focus on a small group of participants, mimicking real world conditions, and a longitudinal period of time [26]. A smaller group of participants more closely resembles real-world VC sessions and allows researchers to adequately analyze the rich qualitative data resulting from a longer term study. The focus was on capturing richness of user interaction and their perceptions.

5.3 Task

Before each session began, we asked for a volunteer to host. The host was responsible for facilitating session operation without interference from researchers. We only received one volunteer to host per session. After the host had been chosen, our research team remained with cameras and microphones off, only appearing at the end of each session to disseminate questionnaires and schedule the next session. Each session featured a different task which lasted approximately 60 minutes. Tasks were chosen based upon their common reporting as VC tasks in our surveys, as well as those present in existing literature. Participants were given written documentation of the tools operation prior to the first session, as well as a tutorial walkthrough by our researchers at the beginning of each session.

- Day 1: Icebreaker task - participants were asked to prepare 1-3 slides introducing themselves and something they do for fun. Participants then discussed an ice-breaker question selected by the host from a list of common ice-breaker questions [19].
- Day 2: Essay Summary task - participants are given a short (1-3 page) essay to summarize (5 minutes) and lead a short (10 minute) discussion. The essays were taken from the same text used in the formative mock-classroom study discussed above.
- Day 3: Brainstorming co-design task - participants are given a design challenge and asked to work together to brainstorm solutions. This challenge was "imagine you have to explain to an alien how a bus works. How would you do this?" Whether to use external tools such as Whiteboard or Mirro were left to the discretion of the host, as well as how to delegate using these tools in collaboration with the other members of their session. Our host chose to use Whiteboard throughout the duration of the entire session.

Following each session, participants were given a brief questionnaire containing short answer and Likert Scale questions detailing their experience with each of the tools contained in Erato, and comparing these experiences to other VC contexts in which they had previously participated. At the conclusion of the final day, participants were also given another questionnaire comprising a System Usability Scale (SUS) as well as qualitative questions detailing their overall impression of the study and Erato. Following this, each participant engaged in a 45 minute semi-structured interview with a researcher, during which time they discussed their answers to the survey and elaborated on their experience. All exit interviews were conducted by the same researcher.

6 RESULTS AND DISCUSSION

Participants found the system to be overall easy to learn, easy to use, and that they would use Erato frequently if it is available (See Figure 4). Likert ratings on a scale of 1 to 5 with 1 meaning "not at all" and 5 meaning "very much so" were comparable between participants who are DHH and non-DHH alike.

We observed that our 3 participants with DHH who participated in all 3 sessions felt more confident in their speaking ability with each subsequent session, which they attributed to using our tool. The 2 participants who joined for a single session were asked the same question, and we found that their result was within the standard deviation of our 3 participants who participated in all 3 sessions' median score reported during the first session.

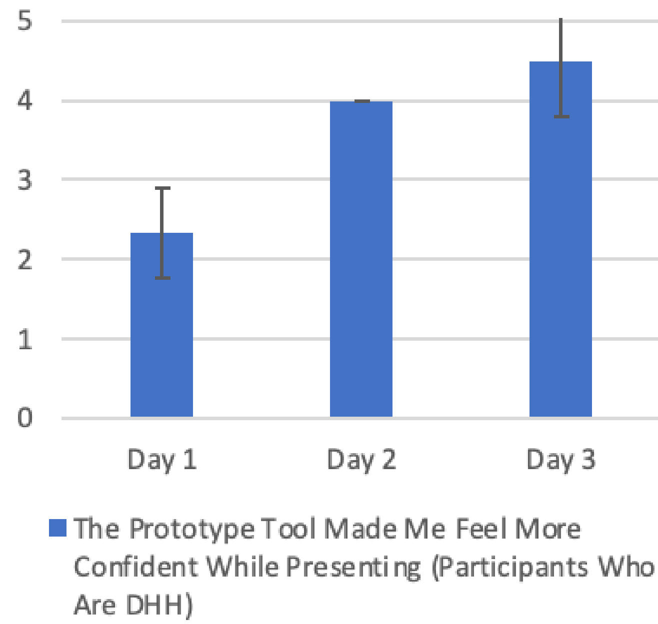


Fig. 4. 5 Point Likert Scale Response From Participants Who Are DHH to the Question "The Prototype Tool Made Me Feel More Confident While Presenting. Results Reported are on a Scale of 1 to 5 with 1 meaning "not at all" and 5 meaning "very much so".

This suggests that participant confidence may not be directly tied to specific tasks and seems to improve over time (See Figure 5). We probed which tool elements potentially resulted in speaking confidence and report these results below.

Participants suggested a total of 9 accommodation requests across all three sessions, 5 of which were approved. Approved accommodation requests were maintained across all future sessions. 3 accommodation requests were integrated into the etiquette guide during the first session, 1 during the second, and 1 during the third. All but one accommodation request was evaluated by the host, and some were omitted due to their commonality or overlap with other requests. Some accommodations which were requested were discussed orally before-hand by the group as the etiquette guideline pertained to a specific social tension arising organically during the session.

6.1 Evaluation of Individual Tools

We collected Likert scale data for each tool from all participants in all 3 sessions and found that scores were consistent and favorable across all 3 sessions during which that tool was used. This suggests that experiences and usability of individual tools were ecologically consistent regardless of task. We unpack the qualitative data and uses of these tools below. We present our results by grouping our tools according to their relevance to our design goals outlined in Section 3.

6.1.1 Etiquette Guide (G1). Both DHH and non-DHH participants indicated that this tool was the most useful of the tools supported by Erato: *"Knowing everyone's accommodations and being able to make mine known was very helpful and empowering"* (P6). One of the key features of the tool is the anonymity provided by the interface, which ameliorates the pressure point of embarrassment we documented in our formative observational study:

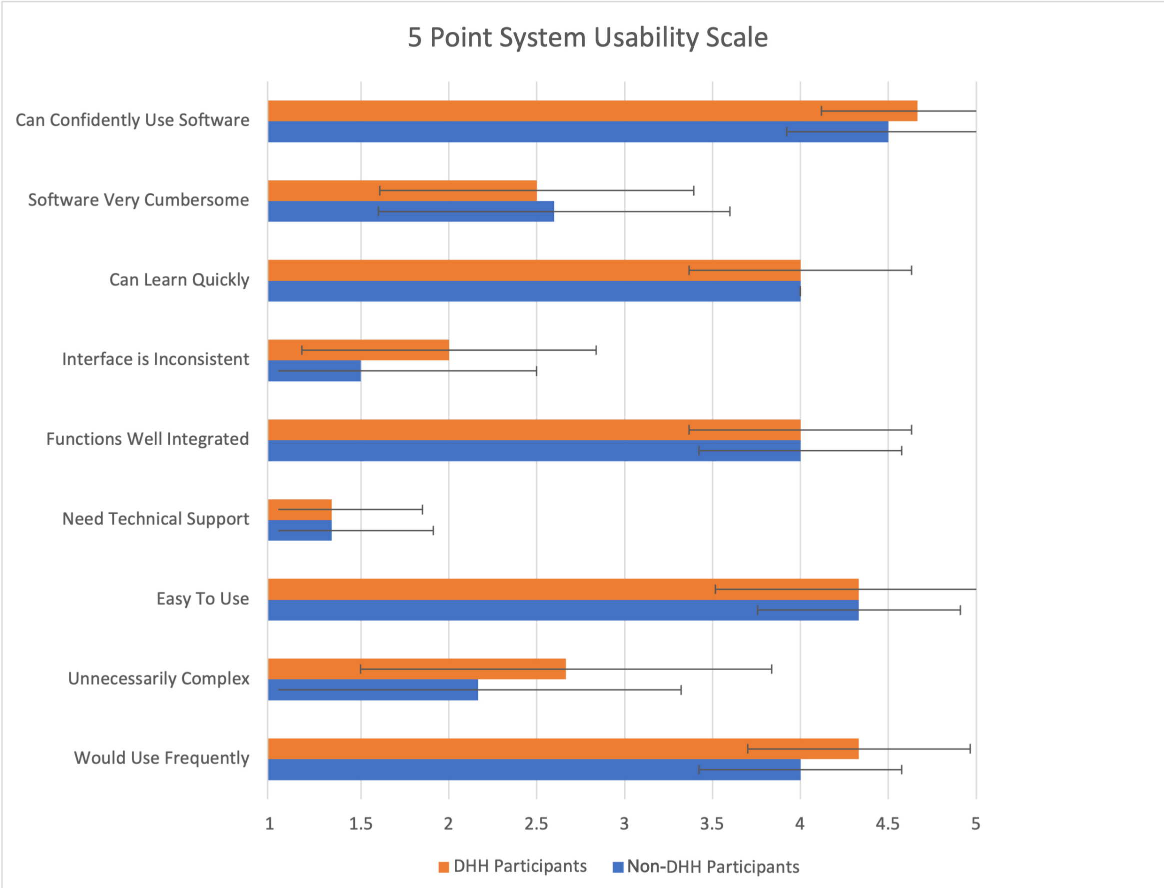


Fig. 5. Average and standard deviations for Systems Usability Scale Results Reported by Participants in our case study. Results are on a scale of 1 to 5 with 1 meaning "not at all" and 5 meaning "very much so".

"it's a great tool , especially for someone who may not be as comfortable asking for accommodations or doesn't want to interrupt or draw attention to themselves to ask for adjustments" (P8). All meeting hosts found the review accommodation request feature easy to use, and appreciated having an opportunity to familiarize themselves with accommodation etiquette prior to facilitating the session: *"It's a lead by example sort of process. I know ahead of time what to do and can make sure I do what's right"* (P6). While providing the host monitoring privileges of accommodation requests was intended to prevent potentially trolling or disruptive behavior on behalf of bad actors, none of the participants behaved this way. Interestingly, the bell button next to each approved accommodation request which sends a reminder of the associated etiquette to all participants in the session was seldom used. We unpack the social dynamics resulting from this tool below in Section 6.2. Participants differed on preferences for component location in the Erato tool window. Our initial design placed the notification center at the top of the window, assuming that this would be an intuitive place to consolidate feedback at-a-glance. However, all of our participants with DHH expressed a wish for the etiquette guide to be placed on top, citing that

Table 3. Summary of accommodation requests made during the Erato case study.

Accommodation Request	Session No.	No. Repeat Requests
"Please look directly at camera while speaking so I can read your lips"	1	3
"Please mute microphone when not speaking"	1	2
"Limit background noise in your environment if possible"	1	0
"Please say your name before speaking"	2	2
"Please say 'over' when done speaking"	3	0

this was far more useful and should be visually prioritized. *"I'd like to see the notifications area and the accessibility reminders swapped as that would be a better line of sight for those who also use captions"* (P8).

6.1.2 Speaker Assistive Tools (G2). Most non-DHH participants monitored the notification center while presenting, while DHH participants monitored individual elements, specifically the speech speed and human volume buttons. Participants reported that they did so to overcome variable audio quality in individual speakers' microphones. All participants who are DHH indicated that the speech speed button was their preferred tool for both presenting and watching other presents because adjusting speech speed had a noticeable impact on the quality of the automated captions. Furthermore, participants found the speech and volume tools to be reassuring they were presenting effectively, and felt more confident in their oral presentation quality as a result. Complications arose during more discussion-based activities where it was unclear who the volume and speech speed feedback was intended to address. Participants involved in all three sessions, however, reported that they adapted to watching the tool more closely when they were speaking, and assumed any notifications regarding speech speed and volume were directed at them if they appeared while they were speaking or within a small buffer of time before or after they spoke.

6.1.3 Host Specific Tools (G3). Hosts had access to additional tools not visible to other session participants. It is common for VC platforms to have a designated session host equipped with specific privileges and tools not afforded other session members to ensure that session activity can be monitored. Since all of our hosts were DHH, they did not require a reminder to turn on the captions since captions are a tool they would normally use. Hosts found that being able to quickly view data regarding the participation of everyone in the session was the most helpful for creating an inclusive environment for participants who are DHH. *"Ensuring students are all participating, as well as making sure people (esp of different genders, races, etc.) in meetings are all being given a chance to be heard."* (P7). All participants with prior experience teach, in particular, remarked on the usefulness and necessity of this tool when adapting their classrooms to VC platforms. *"You can't always see what everyone is doing or keep track of who hasn't spoken…seeing [this data] would keep students from slipping between the cracks"* (P3). While the "limit text chat" and "are people texting or speaking" pie chart were both used at least once, we can not conclude how useful for devising intervention strategies these tools are for session hosts.

6.2 Emergent Socio-Technical Behaviors and Dynamics

Similar to behaviors documented in prior work involving reminder systems, we also observed that participants regarded the etiquette guide once at the beginning of each session, and only returned to review the etiquette guide when someone else was speaking to monitor their own behavior regarding accommodation requests [17]. Hosts similarly regarded the incoming accommodation requests at the beginning of the session, and wouldn't integrate new accommodation requests submitted during the session task. As mentioned above, 1 accommodation request was not evaluated by the host since it was submitted close to the end of session 3, and was not seen by the session host. Interestingly, the “remind participants of etiquette” feature was only used twice across all three sessions. One explanation for this could be that new behaviors were incorporated into group dynamics quickly, and reminders were only necessary when large violations emerged, usually around single individuals. While new accommodation requests were introduced each session, participants easily incorporated these behavior requests into their social ecosystem and maintained these behavior changes across all future sessions. “It’s not an issue of malice, but of ignorance. No one is actively trying to make our lives as DHH people harder. They simply don’t know” (P7) This result suggests that creating awareness of accessibility needs in socio-technical spaces could produce substantial behavior changes, resulting in a more inclusive social ecosystem.

We also probed participants for social responsibility assessments to understand who in the group they assumed would be responsible for monitoring the etiquette guide and using the “remind participants of etiquette” feature. Two consensus emerged from this probe: one group suggested that monitoring these behaviors was the responsibility of those who submitted the request while the other group assumed it was the duty of the session host. *“It just felt potentially rude and judgemental, especially if the slip-up was tiny…I wouldn’t want someone to feel judged for messing up”* (P4) Both of these groups agreed, however, that the main reason they didn’t use the “remind participants of etiquette” feature is because behavior changes and accommodation requests were readily adhered to by the group across sessions.

While the accommodation request tool alleviated experiences of social awkwardness by facilitating direct dialogue between participants with DHH, the host, and other session members, hosts experienced dilemmas around effectively intervening based on the data reflecting individual session participation. Each of our hosts mentioned that they noticed certain individuals who are DHH hadn’t participated in the discussion, but were not comfortable directly calling on them to participate. We probed all participants on their comfortable being called-on and found that most actively wanted to participate but felt they *“couldn’t get in… the conversation was incredibly active and I didn’t want to be rude and interrupt”* (P10). This highlights a tension between the mental model of host and participant desires for intervention provided data around individual participation.

6.3 Design Insights

While the results of our preliminary study suggest some benefit to all the tools comprising Erato, 3 tools in particular produced the most insightful findings. We thus present these three features below as design suggestions to inform the future development of VC platforms to be more inclusive and accessible.

6.3.1 Anonymous Accommodation Request and Communication. We demonstrated that anonymous accommodation requests not only alter behavior by bringing awareness of participants to accessible VC etiquette, but that these behavior changes remain consistent between sessions with different tasks and require little additional effort to maintain on behalf of hosts and participants with disabilities. It is our recommendation that these tools for anonymous accommodation request be integrated into existing VC platforms to further streamline the accessible and inclusive social interactions they afford, as well as address other accessibility and discriminatory concerns.

As noted in prior work, the embarrassment of making accommodation requests is mitigated in larger groups where anonymity of the person making the request is shielded by the size of the group [56]. However this may not remain true in groups with only a single person with accessibility needs. While making accommodation requests is an anonymous process, the session participants anonymity is removed if they are the only person with accessibility needs in the group. While our groups were large enough to provide some anonymity, we noted strategies emerging that could protect the identity of the person with accessibility needs if they were the only participant with accessibility needs in the session. We noted in our case study that both DHH and non-DHH participants made accommodation requests. On day 2 and day 3, non-DHH participants made repeat accommodation requests the noted on the first day. In addition, "please mute your microphone" was an accommodation request made by both

DHH and non-DHH participants simultaneously. The instructor as well made accommodation requests on behalf of students.

6.3.2 Gender Equality and Accessibility Beyond DHH. We initially designed our tool to focus on ameliorating the accessibility limitations of speakers with DHH with a long-term vision of adapting these tools to assist and support people with other disabilities. Surprisingly, p3 brought to our attention that one benefit of visualizing individual participation was the potential for attenuating disparity and discrimination based on gender. *"I wish this was required on all platforms because anytime women speak for more than 30% of any meeting discussion, men think that women are dominating the conversation"* (P3). We probed this idea with other participants who identified as women and found that all had similar experiences of feeling discriminated against when speaking during VC sessions. *"It happens all the time [in the physical spaces] and you just get used to it or don't notice it…having cold data makes it hard to dismiss and inexcusable."* (P3). For this reason, we suggest incorporating this tool into future design of VC platforms to highlight and potentially address the discrimination experienced by minorities when participating in VC spaces. Such a tool is helpful beyond accessibility for people who are DHH. *"For accessibility, diversity, and feminism, we must adopt this tool into widespread use"* (P3).

6.3.3 Human to Human Feedback. While our Speech Speed Meter and Volume Meter used automated feedback evident in other speech feedback tools, the frequent attenuation to the Human Volume Button and Speak Faster/Slower Buttons indicates that fully automated feedback is not enough. *"I would use this immediately. I really wish I had this for church sermons. I attend church online and I would love to be able to discreetly tell the speaker to slow down so I can read their lips. They talk so fast I get lost. Even the captions get scrambled."* (P6). While automated tools for providing real-time feedback on participant volume and speed are a helpful baseline for producing more confident speakers, the specific needs of individuals observing the speaker vary. Providing a means of direct human-to-human feedback such as the Human Volume Button and Speak Faster/Slower Button ensure that the individual and subjective needs of the audience are attenuated. Supporting the individual needs of participants could broaden the inclusivity of VC platforms.

7 LIMITATIONS AND FUTURE WORK

This work explored the usability and effectiveness of a suite of VC tools to support speakers who are DHH, provide valuable data to session hosts regarding participation, and broaden the inclusivity of VC platforms by encouraging inclusive social etiquette. While our initial results show the promise of how incorporating such tools can improve the accessibility and inclusivity of VC sessions across 3 common tasks, we uncovered several opportunities for future work and investigation into how these platforms can be made more equitable.

7.1 Mitigating Bad Actors and Potential System Conflicts

Our study made the assumption that all participants would act in good faith when interacting with the system, and not abuse the tools or intentionally stress the system. While some design choices were made to directly mitigate potential issues (e.g. requiring the session host to approve accommodation requests) other opportunities for abuse are evident in the system. The human volume and speech speed buttons, for example, could be repeatedly pressed by a bad actor to intentionally frustrate the speaker. One way to prevent these potential issues would be to limit how frequently a user could press these buttons, possibly locking these features if persistent use is deemed abusive. Another approach could be to give additional controls to the session host for locking the tools of specific users, similar to a session hosts ability to mute microphones for participants. Still, even if acting in good faith, disagreements on ideal volume levels or speech speed may emerge from interacting with the system, and these struggles may prove disruptive for the participants. One potential solution to mitigating this problem is devise heuristics that detect potential disagreements, and provide individual suggestions for users based on their system state such as "please increase your volume" if the user's volume is turned down low, or "please follow along with this transcript" if the disagreement involves speech speed. While these disagreements did not emerge during our study, examining solutions to these conflicts would be an ample avenue for future work.

7.2 Embarrassment and Plurality

We also noted in our discussion above to the limitations of anonymity afforded by the accommodation request feature. In groups where only one person has a known disability, the systems ability to minimize the potential embarrassment or discomfort evoked by making accommodation requests might be diminished. While we noted several emergent strategies that may mitigate these concerns, future work will explore how a greater plurality of representative abilities may affect the experience of users. While Erato focused on supporting participants who are DHH, future work will explore how other abilities can be supported using a similar system. Supporting a wider set of system supported abilities could help further mitigate the embarrassment or discomfort evoked by making accommodation requests by increasing the likelihoods of a plurality of represented abilities, and thus decreasing the likelihood that a single individual would be making requests. Additional work into eliminating embarrassment altogether remains the subject of future investigations.

7.3 Meeting Session Size and Scope

While we found compelling evidence that integrating tools comprising Erato into existing VC systems increases the confidence of speakers who are DHH as well as broadens the inclusivity and accessibility of VC platforms, these findings are confined to small discussion based groups. Many education and professional use-cases require large VC group meetings, and how the tools present in Erato would scale to such sessions is outside the scope of this initial investigation. Our formative study, for example, documented tensions regarding the cognitive load incurred by a participant with DHH switching from reading captions to reading chat occurring in the text field. This phenomenon was not evident in our case study, and we do not have sufficient evidence to suggest that our tool in its current form resolved this tension. Future work will investigate how the tools comprising Erato scale to large group VC contexts, and how the scale of these gatherings affect the behaviors documented in our initial investigations.

7.4 Extension to ASL Signers

This work focused on users who are DHH that may not know ASL or have access to an interpreter. While considerable prior research exists exploring how to incorporate ASL signing into VC presentations, it remains an open problem. The method proposed in this work could be extended to incorporate the needs of ASL signers, as they could benefit from our approach. One avenue for future research in this domain could be to extend our experiment to incorporate signers and determine their usability by employing the ASL-SUS [28]. The intricacies and needs of this user-group are significant and nuanced, thus necessitating their own study, which we intend as future work.

7.5 Session Host Tools

Our tool suite was designed primarily to support DHH speakers, and we prototyped several new features for session hosts contributing to this goal. Session host tools were informed by prior literature as well as qualitative feedback form an experienced VC instructor, and additional feedback was provided by 3 session participants using Erato's session host tools in our case study. While the initial feedback on these tools is promising, the results are limited by the small number of participants contributing feedback to the tools design. Future work will explicitly focus on improving tools for session hosts who are DHH to support their capacity to effectively facilitate VC session tasks.

7.6 Personalizing Feedback for Discussions

As demonstrated in our user study, individuating the intended recipient of specific real-time feedback becomes difficult during discussions involving quick exchanges of speakers. This produces difficulties in not only understanding who individual pieces of feedback are the intended audience, but also which feedback is meant to be attenuated by the current speaker. One possibility would be to leverage the captions transcript as a source of individuation since current captioning systems are already able to distinguish between individual participants. Clicking a button on one of Erato's tools and dragging to the name of a specific user intended to receive the feedback triggered by that button could ensure that the right feedback is attenuated by the right person, even in rapid exchanges. Examining methods for individuating this sort of feedback is the topic of future work.

7.7 Communal Agglomeration of Accessibility Etiquette Taxonomy

One benefit of Erato's accommodation request system is that it maintains a library of previously approved accommodation requests, and preserves them across sessions. Similar to the approach proposed in prior work [17], sufficient dissemination of Erato could lead to two beneficiary outcomes. First, broader knowledge and awareness of accessible etiquette guidelines could be brought to the attention of diverse target end-users of VC platforms. This alone could greatly improve the accessibility of VC sessions and support long-term behavior change to make socio-technical dynamics more inclusive. Second, prolonged use of Erato by diverse users could over-time agglomerate a taxonomy of accessible etiquette guidelines, formalizing and communally aggregating behaviors which make VC platforms more accessible. While prior literature has documented some etiquette that could improve VC communication, our proposed approach could over-time create a large scale collection of accessible etiquette [38]. Monitoring and creating a consensus on these guidelines presents an interesting set of challenges which would be an ample topic for further discovery.

8 CONCLUSION

While current VC systems are more accessible for various use-cases than physical environments for DHH users, they still mimic several forms of oppressive and discriminatory behavior evident in physical spaces. Addressing these concerns requires systems which can agglomerate and disseminate information regarding inclusive etiquette when participating in social activities mediated by socio-technical systems. We presented Erato, a tool suite comprised several existing tools evident in VC platforms as well as several new tools developed specifically to address the needs expressed by participants with DHH in our formative study. While all of these tools provide a modicum of assistance in improving the accessibility of VC platforms and supporting people with DHH when speaking during VC activities, 3 tools in particular show great promise. 1) A system enabling direct and anonymous communication of accessibility accommodation requests, monitored by the session host, and disseminated to all session participants, 2) A data-driven tool to allow hosts to monitor individual participation of session members as well as quickly understand who most recently spoke, and 3) A method for complementing automated speech speed and spoken volume that allows human-to-human communication of subject speech speed and volume needs. Our formative and case studies demonstrate that these three tools broaden the inclusivity and accessibility of VC sessions for 3 common tasks conducted on VC platforms. Furthermore, qualitative feedback from participants suggests that the applicability of ameliorating discrimination experienced by minorities in VC social spaces extends beyond people who are DHH to people with other disabilities as well women and gender minorities. For these reasons and others mentioned in our study results, we encourage the consideration and adoption of these tools into the design of future VC platforms.

ACKNOWLEDGMENTS

We would like to thank Lessley Hernandez, Kay Frei, and Seyed Hussaini for their help and enthusiasm with this work, as well as Alison May and the Dartmouth Student Accessibility Services for their support and encouragement. In addition, we would like to thank all the study participants for their perseverance and trust.

REFERENCES

[1] Karan Ahuja, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan, and Yuvraj Agarwal. 2019. EduSense: Practical Classroom Sensing at Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 71 (sep 2019), 26 pages. https://doi.org/10.1145/3351229

[2] Elham Alsadoon and Maryam Turkestani. 2020. Virtual Classrooms for hearing-impaired students during the coronavirus COVID-19 pandemic. *Revista Romaneasca pentru Educatie Multidimensionala* 12, 1Sup2 (2020), 01–08. https://doi.org/10.18662/rrem/12.1sup2/240

[3] Paul Ayres and John Sweller. 2005. *The Split-Attention Principle in Multimedia Learning*. Cambridge University Press, 135–146. https://doi.org/10.1017/CBO9780511816819.009

[4] Kristoffer Bergram, Marija Djokovic, Valéry Bezençon, and Adrian Holzer. 2022. The Digital Landscape of Nudging: A Systematic Literature Review of Empirical Research on Digital Nudges. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 62, 16 pages. https://doi.org/10.1145/3491102.3517638

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

[6] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 155–164. https://doi.org/10.1145/3132525.3132541

[7] Johnna Blair and Saeed Abdullah. 2020. It Didn't Sound Good with My Cochlear Implants: Understanding the Challenges of Using Smart Assistants for Deaf and Hard of Hearing Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 118 (dec 2020), 27 pages. https://doi.org/10.1145/3432194

[8] Yosra Bouzid and Mohamed Jemni. 2021. The Deaf Experience in Remote Learning during COVID-19. In *2021 8th International Conference on ICT and Accessibility (ICTA)*. 1–3. https://doi.org/10.1109/ICTA54582.2021.9809424

[9] Alessandra Brandão, Hugo Nicolau, Shreya Tadas, and Vicki L. Hanson. 2016. SlidePacer: A Presentation Delivery Tool for Instructors of Deaf and Hard of Hearing Students. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 25–32. https://doi.org/10.1145/2982142.2982177

[10] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Brussels, Belgium) *(TVX '15)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2745197.2745204

[11] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 2009. ClassInFocus: Enabling Improved Visual Attention Strategies for Deaf and Hard of Hearing Students. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, Pennsylvania, USA) *(Assets '09)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/1639642.1639656

[12] Teresa Y. Ching and Harvey Dillon. 2013. A brief overview of factors affecting speech intelligibility of people with hearing loss: Implications for amplification. *American Journal of Audiology* 22, 2 (2013), 306–309. https://doi.org/10.1044/1059-0889(2013/12-0075)

[13] Debabrata Chowdhuri, Narendra Parel, and Amrita Maity. 2012. Virtual classroom for deaf people. In *2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA)*. 1–3. https://doi.org/10.1109/AICERA.2012.6306730

[14] J. Davis. 2019. IllumiWear: A Fiber-Optic eTextile for MultiMedia Interactions. In *NIME*.

[15] Josh Davis. 2022. *PokerFace Mask: Exploring Augmenting Masks with Captions through an Interactive, Mixed-Reality Prototype*. https://doi.org/10.24251/HICSS.2022.398 Accepted: 2021-12-24T17:47:44Z.

[16] Josh Urban Davis, Paul Asente, and Xing-Dong Yang. 2023. Multimodal Direct Manipulation in Video Conferencing: Challenges and Opportunities. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1174–1193. https://doi.org/10.1145/3563657.3596099

[17] Josh Urban Davis, Jun Gong, Yunxin Sun, Parmit Chilana, and Xing-Dong Yang. 2019. CircuitStyle: A System for Peripherally Reinforcing Best Practices in Hardware Computing. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software*

and Technology (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 109–120. https://doi.org/10.1145/3332165.3347920

[18] Josh Urban Davis, Te-Yen Wu, Bo Shi, Hanyi Lu, Athina Panotopoulou, Emily Whiting, and Xing-Dong Yang. 2020. TangibleCircuits: An Interactive 3D Printed Circuit Education Tool for People with Visual Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376513

[19] A.S. Drigas, D. Kouremenos, S. Kouremenos, and J. Vrettaros. 2005. An e-learning system for the deaf people. In *2005 6th International Conference on Information Technology Based Higher Education and Training*. T2C/17–T2C/21. https://doi.org/10.1109/ITHET.2005.1560236

[20] Lisa Elliot, Michael Stinson, James Mallory, Donna Easton, and Matt Huenerfauth. 2016. Deaf and Hard of Hearing Individuals' Perceptions of Communication with Hearing Colleagues in Small Groups. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 271–272. https://doi.org/10.1145/2982142.2982198

[21] Lisa B. Elliot, Michael Stinson, Syed Ahmed, and Donna Easton. 2017. User Experiences When Testing a Messaging App for Communication Between Individuals Who Are Hearing and Deaf or Hard of Hearing. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 405–406. https://doi.org/10.1145/3132525.3134798

[22] Abraham Glasser, Edward Mason Riley, Kaitlyn Weeks, and Raja Kushalnagar. 2019. Mixed Reality Speaker Identification as an Accessibility Tool for Deaf and Hard of Hearing Users. In *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 80, 3 pages. https://doi.org/10.1145/3359996.3364720

[23] Steven Goodman, Susanne Kirchner, Rose Guttman, Dhruv Jain, Jon Froehlich, and Leah Findlater. 2020. Evaluating Smartwatch-Based Sound Feedback for Deaf and Hard-of-Hearing Users Across Contexts. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376406

[24] Steven M. Goodman, Ping Liu, Dhruv Jain, Emma J. McDonnell, Jon E. Froehlich, and Leah Findlater. 2021. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 63 (jun 2021), 23 pages. https://doi.org/10.1145/3463501

[25] Jan Gugenheimer, Katrin Plaumann, Florian Schaub, Patrizia Di Campli San Vito, Saskia Duck, Melanie Rabus, and Enrico Rukzio. 2017. The Impact of Assistive Technology on Communication Quality Between Deaf and Hearing Individuals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 669–682. https://doi.org/10.1145/2998181.2998203

[26] Trista Hollweck. 2016. Robert K. Yin. (2014). Case Study Research Design and Methods (5th ed.). Thousand Oaks, CA: Sage. 282 pages. *The Canadian Journal of Program Evaluation* 30 (03 2016). https://doi.org/10.3138/cjpe.30.1.108

[27] Richang Hong, Meng Wang, Xiao-Tong Yuan, Mengdi Xu, Jianguo Jiang, Shuicheng Yan, and Tat-Seng Chua. 2011. Video Accessibility Enhancement for Hearing-Impaired Users. *ACM Trans. Multimedia Comput. Commun. Appl.* 7S, 1, Article 24 (nov 2011), 19 pages. https://doi.org/10.1145/2037676.2037681

[28] Matt Huenerfauth, Kasmira Patel, and Larwan Berke. 2017. Design and Psychometric Evaluation of an American Sign Language Translation of the System Usability Scale. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 175–184. https://doi.org/10.1145/3132525.3132540

[29] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People Who Are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[30] 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* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 15–23. https://doi.org/10.1145/2982142.2982164

[31] Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376266

[32] Raja Kushalnagar. 2019. A Classroom Accessibility Analysis App for Deaf Students. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 569–571. https://doi.org/10.1145/3308561.3354640

[33] Raja S. Kushalnagar, Gary W. Behm, Aaron W. Kelstone, and Shareef Ali. 2015. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Real-Time Speech-To-Text. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and*

Accessibility (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 223–230. https://doi.org/10.1145/2700648.2809843

[34] 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

[35] Makayla Lewis, Miriam Sturdee, John Miers, Josh Urban Davis, and Thuong Hoang. 2022. Exploring AltNarrative in HCI Imagery and Comics. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 7, 13 pages. https://doi.org/10.1145/3491101.3516394

[36] Franklin Mingzhe Li, Cheng Lu, Zhicong Lu, Patrick Carrington, and Khai N. Truong. 2022. An Exploration of Captioning Practices and Challenges of Individual Content Creators on YouTube for People with Hearing Impairments. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 75 (apr 2022), 26 pages. https://doi.org/10.1145/3512922

[37] Sheng Lihua and Xu Jiacheng. 2010. Using social software to improve learning performance of deaf university learner. In *2010 2nd IEEE International Conference on Information Management and Engineering*. 703–706. https://doi.org/10.1109/ICIME.2010.5478205

[38] Kelly Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, John Tang, Andrew Begel, Erin Beneteau, Josh Urban Davis, Abraham Glasser, Joon Sung Park, and Venkatesh Potluri. 2021. Mixed Abilities and Varied Experiences: A Group Autoethnography of a Virtual Summer Internship. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 21, 13 pages. https://doi.org/10.1145/3441852.3471199

[39] Susan M. Mather and M. Diane Clark. [n. d.]. an issue of learning the effect of visual split attention in classes for. https://files.eric.ed.gov/fulltext/EJ976477.pdf

[40] Tara Matthews, Janette Fong, and Jennifer Mankoff. 2005. Visualizing Non-Speech Sounds for the Deaf. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, MD, USA) *(Assets '05)*. Association for Computing Machinery, New York, NY, USA, 52–59. https://doi.org/10.1145/1090785.1090797

[41] 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

[42] Matthias Mielke and Rainer Brueck. 2015. Design and evaluation of a smartphone application for non-speech sound awareness for people with hearing loss. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5008–5011. https://doi.org/10.1109/EMBC.2015.7319516

[43] Ashley Miller, Joan Malasig, Brenda Castro, Vicki L. Hanson, Hugo Nicolau, and Alessandra Brandão. 2017. The Use of Smart Glasses for Lecture Comprehension by Deaf and Hard of Hearing Students. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1909–1915. https://doi.org/10.1145/3027063.3053117

[44] Dorian Miller, Karl Gyllstrom, David Stotts, and James Culp. 2007. Semi-Transparent Video Interfaces to Assist Deaf Persons in Meetings. In *Proceedings of the 45th Annual Southeast Regional Conference* (Winston-Salem, North Carolina) *(ACM-SE 45)*. Association for Computing Machinery, New York, NY, USA, 501–506. https://doi.org/10.1145/1233341.1233431

[45] 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. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1108–1120. https://doi.org/10.1145/3379337.3415817

[46] Marios A. Pappas, Eleftheria Demertzi, Yannis Papagerasimou, Lefteris Koukianakis, Dimitris Kouremenos, Ioannis Loukidis, and Athanasios S. Drigas. 2018. E-Learning for Deaf Adults from a User-Centered Perspective. *Education Sciences* 8, 4 (2018). https://doi.org/10.3390/educsci8040206

[47] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[48] Anthony Peruma and Yasmine N. El-Glaly. 2017. CollabAll: Inclusive Discussion Support System For Deaf and Hearing Students. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 315–316. https://doi.org/10.1145/3132525.3134800

[49] Jazz Rui Xia Ang, Ping Liu, Emma McDonnell, and Sarah Coppola. 2022. "In This Online Environment, We're Limited": Exploring Inclusive Video Conferencing Design for Signers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 609, 16 pages. https://doi.org/10.1145/3491102.3517488

[50] Vít Rusňák, Pavel Troubil, Desana Daxnerová, Pavel Kajaba, Matej Minárik, Svatoslav Ondra, Tomás Sklenák, and Eva Hladká. 2016. CoUnSiL: A video conferencing environment for interpretation of sign language in higher education. In *2016 15th International Conference*

on Information Technology Based Higher Education and Training (ITHET). 1–8. https://doi.org/10.1109/ITHET.2016.7760711

[51] Matthew Seita. 2016. Closed ASL Interpreting for Online Videos. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 337–338. https://doi.org/10.1145/2982142.2982147

[52] 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

[53] Matthew Seita, Sooyeon Lee, Sarah Andrew, Kristen Shinohara, and Matt Huenerfauth. 2022. Remotely Co-Designing Features for Communication Applications Using Automatic Captioning with Deaf and Hearing Pairs. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 460, 13 pages. https://doi.org/10.1145/3491102.3501843

[54] Brent N. Shiver and Rosalee J. Wolfe. 2015. Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 231–238. https://doi.org/10.1145/2700648.2809857

[55] Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication.* Cambridge University Press, USA.

[56] John Tang. 2021. Understanding the Telework Experience of People with Disabilities. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 30 (apr 2021), 27 pages. https://doi.org/10.1145/3449104

[57] Vinoba Vinayagamoorthy, Maxine Glancy, Christoph Ziegler, and Richard Schäffer. 2019. Personalising the TV Experience Using Augmented Reality: An Exploratory Study on Delivering Synchronised Sign Language Interpretation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300762

[58] Christian Vogler, Paula Tucker, and Norman Williams. 2013. Mixed Local and Remote Participation in Teleconferences from a Deaf and Hard of Hearing Perspective. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) *(ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, Article 30, 5 pages. https://doi.org/10.1145/2513383.2517035

A APPENDIX: QUESTIONNAIRES FROM USER STUDIES

To help with replication and future comparisons, we have included our surveys and questionnaires from our studies below.

A.1 Accessible Remote Video Conferencing Survey

This survey was used for collecting general responses to experiences using remote video conferencing software.

1. What is your age?
2. What is your gender identity?
3. What is your occupation?
4. Are you Deaf or Hard of Hearing?
5. How would you describe any disabilities you may have?
6. Do you have experience with remote video conferencing software such as Zoom or Teams?
7. What are some examples of times you have used video conferencing systems in the past? Please list as many as you can.
8. Describe some of the barriers or pressure points you experience during remote video conferencing sessions?
9. What are the best features of remote video conferencing environments? (e.g. Taking meetings from anywhere, less background noise, etc.) Do you use any accessibility tools during in-person lectures, meetings, or presentations?
10. If so, what are they and are they provided by you, your employer, or someone else?
11. Where do you typically take your remote video conferences?
12. What sort of difficulties do you face when giving presentations in remote conferencing environments?

13. Please compare your experience giving presentations or talking in remote environments versus physical environments.
14. I am able to customize remote video conferencing environments to suit my needs.
15. How well did the remote environment meet your accessibility needs without further customization?
16. I feel confident presenting or giving a lecture in remote video conferencing environments
17. I feel confident talking or socializing in remote video conferencing environments
18. I feel confident brainstorming in remote video conferences
19. I feel confident presenting in physical environments
20. I find it easier to present in remote video conferencing environments than physical environments
21. Please explain your answer to the previous question in detail I am able to communicate effectively in remote video conferencing environments.
22. I am able to communicate effectively in in-person presentation environments.
23. I am able to focus effectively during remote video conferencing sessions.
24. I am able to focus effectively during in-person presentation sessions.
25. I am comfortable communicating with meeting hosts to request accommodations for remote video conferences.
26. I am comfortable communicating with meeting hosts to request accommodations for in-person meetings or conferences.
27. Would you be willing to participate in a follow-up interview to this questionnaire? All follow-up interview participants will be compensated for their participation in the interviews scheduled at their convenience.

A.2 Mock Classroom Study Questionnaire

This questionnaire was given after each presenter during the mock classroom formative study.

1. Please enter your participant ID
2. Please enter the discussion section number (1-6) meaning which speaker just finished speaking?
3. What elements of the classroom interface do you find helpful for supporting and sustaining your engagement?
4. What elements of the classroom interface create tension or pressure points which makes remaining engaged in the classroom discussion or presentation difficult?
5. Please indicate from 1 to 5 with 1 meaning "not at all" and 5 meaning "very much so" how engaged you felt during this section
6. Please indicate from 1 to 5 with 1 meaning "not at all" and 5 meaning "very much so" how much of the material from the presentation and discussion you feel you retained.

A.3 Use-Case Study Questionairre

The following questionnaire was given during the use-case study described in Section 5.

1. Did you present or participate in today's session?
2. How do you feel you performed overall during your presentation?
3. Could you comment on your overall experience using the prototype tool?
4. Did you use the prototype tool while presenting? Which elements of the prototype tool did you use while presenting? (If you did not use the tool, just put "NA")
5. Which elements of the prototype tool did you use while other participants were presenting? (If you did not use the tool, just put "NA")
6. Could you please compare your experience using the prototype tool today versus previous experiences without the tool?

6. Please rank the following elements of the prototype in order from "most useful" to "least useful". If you didn't use a specific element of the prototype, select "NA". [Accessibility Accommodation Request] [Etiquette Guide] [Report Etiquette Violation] [Speach Volume Meter] [Speak Louder/Quieter Buttons] [Speech Speed Meter] [Speak Quicker/Slower Buttons] [Headphone Playback] [Ensure Captions Toggle] [Limit Text Chat Toggle] [Who Spoke Last Dashboard] [Text vs Spoken Chat Pie Chart]

7. Please explain your above answer

8. Please rate the following features of the prototype on a scale from 1 to 5 where 1 indicates "not at all useful" and 5 indicates "very useful"

9. Accessibility Accommodation Request - at the bottom of the chrome extension where requests for accommodation could be entered to become part of the etiquette guide [Rating]

10. Please explain your rankings in the above question.

11. Etiquette Guide - The list of accommodation requests submitted by members of the session [Rating]

12. Please explain your rankings in the above question.

13. Etiquette Reminder Alert - The ability for session participants to remind members of the session to be mindful of accommodation requests listed in the etiquette guide by pressing the bell button. [Rating]

14. Please explain your rankings in the above question.

15. Speech Volume Meter - Volume meter at top of chrome extension that provides feedback to the speaker regarding their speech volume [Rating]

16. Please explain your rankings in the above question.

17. Speech Rate Meter - Meter at top of chrome extension that provides feedback to the speaker regarding their speech speed [Rating]

18. Please explain your rankings in the above question.

19. Speak Louder/ Quieter Buttons - Buttons beneath volume meter to allow participants to suggest that the person speaking should speak louder [Rating]

20. Please explain your rankings in the above question.

21. Speak Quicker/ Slower Buttons - Buttons beneath volume meter to allow participants to suggest that the person speaking should speak slower or quicker [Rating]

22. Please explain your rankings in the above question.

23. Headphone Playback - Toggle switch which plays audio back through headphones [Rating]

24. Please explain your rankings in the above question.

25. Ensure Captions Toggle - Toggle switch to remind the host and session participants to turn on captions [Rating]

26. Please explain your rankings in the above question.

27. Limit Text Chat Toggle - Toggle switch that reminds session participants to keep text chat to a minimum [Rating]

28. Please explain your rankings in the above question.

29. (Host) Who Spoke Last Dashboard - Lists all session participants indicating who spoke last and how much time each participant has spoken during the session [Rating]

30. Please explain your rankings in the above question.

31. (Host) Text vs Spoke Chat Pie Chart - Pie chart demonstrating how much chatting has occurred via text in the chat window versus spoken chat in the meeting room [Rating] Please explain your above answer I am confident I presented well today [Answer]

32. The prototype tool was helpful while I was speaking [Answer]

33. The prototype tool helped me be aware of the needs of other people in my session [Answer]

34. Other people presenting during my session spoke confidently, clearly and well [Answer]

35. I used the "report" feature to encourage other participants in the session to adhere to the etiquette guide [Answer]
36. The prototype tool was difficult to use [Answer]
37. The prototype tool was overly complicated [Answer]
38. I had difficulty learning to use the prototype tool [Answer]
39. I was overwhelmed by the number of notifications provided by the prototype tool [Answer]
40. The prototype tool was distracting [Answer]
41. The prototype tool made me feel more confident while presenting [Answer]
42. The prototype tool helped make the session more inclusive for everyone [Answer]
43. I would use the prototype tool in future video conferencing sessions. [Answer]
44. Is there anything else you would like to tell us about your experience using the tool today?

B APPENDIX: COMMERCIAL VC ANALYSIS

Here we collect results of the commercial VC platform analysis we describe in Section 3.

Table 4. Benefits and pain points of Microsoft Teams as described by DHH/HoH survey and interview participants in Formative Study (See Section 3).

Pros	Cons
Make a written record of your meeting	Captioning is a little obscure to find
Captioning seemed more accurate than Meet and Zoom	Accessibility tools not on main menu bar (requries searching)
Can add interpreters to meetings	Mobile apps for Teams have limited capabilites
Captions can distinguish between speakers very well	Person sharing screen cannot see captions
Captions can catch when rapidly switching between speakers	No visual representation of speakers volume
Caption size is bigger Zoom	Unless you are a paying (subscription) Teams member, you do not get captions (financial inaccessibility)
Name of person listed in transcript and captions	Caption speed not adjustable
Caption size adjustable	Captions not visible in breakout rooms

Table 5. Additional benefits and pain points of Zoom compared to Teams and Meet as described by DHH/HoH survey and interview participants in Formative Study (See Section 3).

Pros	Cons
Has captioning and captioning is a free service	Breakout rooms do not have captions
Pop up message notifies speaker that mic is muted	Transcripts not available for breakout rooms
Allows third-party captioning (which can be better than the built-in standard captioning) but these services often require additional fees	For HoH participants, it is difficult to tell whether the headphones aren't working or people are speaking low
Supports having a live transcriber who types what is being said	Captions are very inaccurate (worse than both Meet and Teams)
Captions displayed while sharing screen (this feature is not evident in Meet nor Teams)	Captions disappear when using direct message feature on mobile devices
Create a live transcription	Live transcripts are slow to produce and very inaccurate
Visually displays volume of person speaking	Captions often confuse who is speaking and are unable to detect when the speaker switches rapidly
Direct messaging supported if enabled by meeting host	Captions display filler words
Font size of captions is adjustable	Captions do not adjust to screen size
Captions can detect and display a person speaking while sharing a video via screen sharing	When speaker shares a video via screen sharing, their audio is significantly reduced while other meeting participants can still speak at full volume

Table 6. Benefits and pain points of Google Meet as described by DHH/HoH survey and interview participants in Formative Study (See Section 3).

Pros	Cons
Has captioning and switching between captioning languages is a free service	Non-English captions are slower and far more inaccurate than English captions
Text is bigger than Zoom	Captions are slightly inaccurate
Camera quality is better than Teams and Zoom	Lacks transcript capabilities
Can add Chrome extensions to extend capabilities of Meet	Caption speed cannot be adjusted
Lists the name of person speaking in captions	Participants cannot be directly messaged
Captions can detect quick switching between speakers	Unless you are a paying (subscription) Teams member, you do not get captions (financial inaccessibility)
Adjusts screen to fit the captions onscreen	The person sharing the screen can't see the captions
Captions also available in breakout rooms	No visual indication of spoken volume
Captions can detect and display when two participants speak at the same time	Captions are only displayed on the tab of the Meet session and are not displayed on other tabs nor while sharing screen
Captions are robust to background noise and music	Captions not displayed when using whiteboard function or other tools and chrome extensions
Captions can detect and display audio from video shared via screen sharing	Does not caption laughter nor ambient sounds

C APPENDIX: FULL FORMATIVE STUDY PARTICIPANT TABLE

Table 7. Description of participants in formative studies. Fields noted as N/A indicate participant declined to answer this survey question. An abbreviated version of this table containing only participants included in the live study was described in Section 3.

Participant Number	Self-Described Disability	Age	Gender Identity	Occupation	Study Participation Portion
FP1	Progressive hearing loss	20	Cis-Man	Student	Survey, Interview, Live Study
FP2	ADHD	21	Demi-Male	Student	Survey, Interview, Live Study
FP2	ASD	26	Cis-Woman	Student	Survey, Interview, Live Study
FP4 *	None	N/A	Cis-Woman	University Instructor	Survey, Interview, Live Study
FP5	bilateral hard of hearing since birth, low vision	36	Cis-Man	Student	Survey, Interview, Live Study
FP6	None	19	Cis-Man	Student	Survey
FP7	None	30	Trans-Man	Graduate Student	Survey
FP8	partial vision, low vision	28	Cis-Man	Unemployed	Survey
FP9	None	19	Cis-Woman	Student	Survey
FP10	bilateral hearing loss	28	Cis-Woman	Administration	Survey, Interview
FP11	None	32	Cis-Man	Vocalist	Survey
FP12	None	20	Cis-Man	Student	Survey
FP13	None	29	Trans-Woman	Lecturer	Survey
FP14	None	28	Cis-Woman	Teacher	Survey
FP15	legally blind	31	Cis-Man	Unemployed	Survey, Interview
FP16	ADHD	20	Cis-Man	Student	Survey, Interview
FP17	Hard of Hearing and ADHD	36	Cis-Man	Administration	Survey, Interview
FP18	ADHD and Dyslexia	26	Cis-Woman	Student	Survey, Interview
FP19	None	19	Cis-Woman	Student	Survey
FP20	ASD	27	Cis-Man	Analyst	Survey, Interview
FP21	None	19	Cis-Woman	Student	Survey
FP22	None	22	Cis-Man	Food Service	Survey
FP23	None	26	Cis-Man	UX Design	Survey
FP24	Hard of Hearing and ASD	44	Cis-Woman	Unemployed	Survey, Interview
FP25	deaf	23	Two Spirit	Student	Survey, Interview
FP26	deaf and immunocompromised by chronic heart condition	18	Cis-Woman	Student	Survey, Interview