Jod: Examining the Design and Implementation of a Videoconferencing Platform for Mixed Hearing Groups

Anant Mittal
anmittal@cs.washington.edu
University of Washington
Seattle, Washington, USA

Meghna Gupta
t-meggupta@microsoft.com
Microsoft Research
Bengaluru, India

Roshni Poddar
t-ropoddar@microsoft.com
Microsoft Research
Bengaluru, India

Tarini Naik
t-naiktarini@microsoft.com
Microsoft Research
Bengaluru, India

Seethalakshmi Kuppuraj
info@winvinayafoundation.org
WinVinaya Foundation
Bengaluru, India

James Fogarty
jfogarty@cs.washington.edu
University of Washington
Seattle, Washington, USA

Pratyush Kumar
pratyush@cse.iitm.ac.in
Microsoft Research
Bengaluru, India

Mohit Jain
mohja@microsoft.com
Microsoft Research
Bengaluru, India

Figure 1: Snapshots from user study sessions (a) a camera shot of *Jod*, (b) a Deaf or hard of hearing participant signing while using *Jod*, and (c) mixed-hearing focus group discussion with interpreter.

ABSTRACT

Videoconferencing usage has surged in recent years, but current platforms present significant accessibility barriers for the 430 million d/Deaf or hard of hearing people worldwide. Informed by prior work examining accessibility barriers in current videoconferencing platforms, we designed and developed *Jod*, a videoconferencing platform to facilitate communication in mixed hearing groups. Key features include support for customizing visual layouts and a notification system to request attention and influence behavior. Using *Jod*, we conducted six mixed hearing group sessions with 34 participants, including 18 d/Deaf or hard of hearing participants, 10 hearing participants, and 6 sign language interpreters. We found participants engaged in visual layout rearrangements based on their hearing ability and dynamically adapted to the changing group communication context, and that notifications were useful but raised a need for designs to cause fewer interruptions. We provide insights for future videoconferencing designs and conclude with recommendations for conducting mixed hearing studies.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; **Accessibility design and evaluation methods**; **Human computer interaction (HCI)**; **HCI design and evaluation methods**.

KEYWORDS

Accessibility, Deaf, Hard of Hearing, DHH, Mixed Hearing Groups, Videoconferencing, Accessible Research Methods

ACM Reference Format:
Anant Mittal, Meghna Gupta, Roshni Poddar, Tarini Naik, Seethalakshmi Kuppuraj, James Fogarty, Pratyush Kumar, and Mohit Jain. 2023. *Jod*: Examining the Design and Implementation of a Videoconferencing Platform for Mixed Hearing Groups. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23), October 22–25, 2023, New York, NY, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3597638.3608382

ASSETS '23, October 22–25, 2023, New York, NY, USA

ACM ISBN 979-8-4007-0220-4/23/10.
https://doi.org/10.1145/3597638.3608382

1 INTRODUCTION

Broad adoption of videoconferencing platforms has surged since mid-2019, primarily due to the COVID-19 pandemic. The use of popular videoconferencing platforms (e.g., Zoom, Microsoft Teams, Google Meet) was 21 times higher in the first half of 2020 compared to the first half of 2019 [8], and their usage is projected to grow in the coming decade [3]. Studies by the Pew Research Center have found that these platforms are used for many purposes, such as remote work, maintaining social connections, and telehealth, among many others [2]. With the increase in adoption and use, videoconferencing platforms also aim to provide more inclusive support through features related to accessibility needs (e.g., live captions and transcriptions, support for screen readers, and multi-pinning and multi-spotlighting to support visual layout customization [32, 38]). Particularly relevant to this paper's focus on d/Deaf or hard of hearing (DHH) individuals, Microsoft Teams and Zoom introduced sign language interpretation views [27, 35]. It prioritizes sign language users (hereafter called signers) and interpreters by maintaining a fixed view of their video tiles.

Despite these efforts, videoconferencing platforms present significant accessibility barriers for the DHH community [1, 14, 17, 26, 36], estimated to comprise over 430 million people worldwide [25]. Prior research in HCI and accessibility has examined the usage of videoconferencing platforms by DHH individuals [14, 26] and identified three main challenges. First, current videoconferencing platforms offer limited default layouts for the users to choose from and tend to automatically resize and distribute video thumbnails over multiple pages, posing obstacles for visual communication [14, 26]. Such limited layout customization capabilities hinder DHH users' ability to personalize their view of other DHH individuals, active speakers, and interpreters [13, 14, 26, 36]. Second, DHH individuals often feel uncomfortable getting other participants' attention, as they find it challenging to interject an ongoing conversation (even with the interpreter's help) [26]. Additionally, in mixed hearing videoconferencing settings, hearing and DHH individuals face challenges in remembering appropriate communication accommodations, such as hearing individuals forgetting to speak slowly or turning on their video when conversing with DHH individuals [17]. Third, videoconferencing platforms' audio-centric design cannot highlight signing individuals' video tiles [36]; instead, the interpreter who voices them gets the focus in visual layouts. Therefore, current videoconferencing platforms fail to– provide personalized visual layout arrangements, support DHH participants to interject, and enable users to remember appropriate accommodations for others.

Prior research [13, 26, 30] has primarily employed participatory design methods such as co-design workshops to explore potential design solutions to address these challenges. Design recommendations include options for resizing and reordering video frames, grouping videos, offering visual and haptic feedback to request attention, and prioritizing frames of active speakers [13, 26, 30]. However, a limited understanding remains of how these solutions would translate into action in real-world mixed hearing videoconferencing settings. To examine this, we designed and developed *Jod*[1], a videoconferencing platform to facilitate communication in mixed hearing groups. *Jod* provides users with an enhanced option to customize their visual layout, enabling them to resize, rearrange, and add/remove video tiles of participants. It also includes a notification system with preset messages to get people's attention and influence speaker behavior. *Jod* also highlights active signer(s) using a Wizard of Oz technique [9]. Furthermore, it displays accessibility indicators as part of user profiles to help gauge and identify fellow participants' needs.

To understand behaviors and perceptions when navigating mixed hearing ability conversations using *Jod*, we conducted six user study sessions with 34 participants, including 18 DHH participants, 10 hearing participants, and 6 Indian Sign Language (ISL) interpreters. Each session consisted of a tutorial, followed by task-based exploration, unstructured conversation, a game of charades, a presentation with screen share, and then concluded with a focus group discussion. To supplement our qualitative analysis, we also collected system-wide telemetry data. Participants engaged in 485 visual layout-related arrangements and sent 40 preset messages throughout the study. Our findings unveiled several novel insights, particularly a strong correlation between participants' hearing abilities and their preferred visual layout arrangements. Notably, the DHH participants made the interpreter's video tile significantly larger than the hearing participants and chose to move the closed captions closer to the interpreter's video tile. Interestingly, participants also engaged in visual layout-related rearrangements to adapt to the changing group communication context, particularly during the game of *charades* where they could prioritize the participant whose turn it was. Though such customization capabilities provided complete control over visual layouts, it also led to additional manual labor. Thus, our participants desired a balance between flexibility and system-provided automated defaults to reduce their labor. Moreover, participants reported improved communication between DHH signers, hearing users, and interpreters through preset feedback messages. While it helped in interjecting, requesting attention, and influencing speaker behavior, these features also raised a need for acknowledgments and prioritization of received messages based on the group communication context. Drawing on these findings, we synthesize key takeaways and provide guidelines for designing videoconferencing platforms to support mixed hearing communication better, focusing on visual layout customization, interactivity and reactivity of the platform, and cultural considerations. We conclude with recommendations for conducting inclusive mixed hearing studies.

In summary, our work contributes: (1) the design and development of *Jod*, a videoconferencing system integrating recommendations from prior work to facilitate communication in mixed hearing groups, (2) findings from 6 study sessions with 18 DHH participants, 10 hearing participants, and 6 ISL interpreters, examining how Jod's features interact with each other and the emergent behaviors and perceptions of participants, and (3) design guidelines for future accessible videoconferencing platforms and recommendations for conducting mixed hearing studies.

[1] *Jod*: a Hindi word, pronounced as *j-o-rr-h*, which means 'link' and emphasizes the system's ability to connect individuals.

2 RELATED WORK

Our work is informed by the communication challenges DHH individuals face in mixed hearing groups while using videoconferencing platforms. Mixed hearing groups rely on multiple communication methods, such as sign language, speech reading (also called lipreading), gestures, body language, facial expressions, captioning, pen-paper/text-based chat, and interpreters. However, most of these may not translate well into online settings, resulting in various communication challenges. We discuss some communication methods used by people with hearing disabilities and provide an overview of prior studies to understand the usage and challenges of these methods in current videoconferencing platforms.

2.1 Sign Language and Speechreading

Sign languages are the primary mode of communication in the d/Deaf community, with over 200 global variants [24]. Unlike spoken languages, sign languages rely on spatial cognition, communicating information through hand shapes, body movements, and facial expressions [5]. Each sign language has its distinct grammar and vocabulary. For instance, Indian Sign Language (ISL), the most commonly used language by the DHH community in India [7], differs substantially from American Sign Language (ASL). Besides enabling communication, the DHH community identifies their sign language as a source of pride, thus constituting it as an essential part of their identity [15]. In the digital world, video calls enable people to interact using sign language. Prior work [14, 26, 37] has identified several challenges with it, including difficulty in reading signs due to reduced frame rates and inability to find interpreter's video tile in large groups. Access to a human interpreter is the most reliable solution for the DHH community to interact with hearing individuals [31]. However, it is often not feasible due to the scarcity and affordability of interpreters. Additionally, Kushalnagar and Vogler [14] have discussed challenges in videoconferencing platforms like limited and somewhat rigid support for organizing multiple visual elements (e.g. speaker video, interpreter video, captions, screen share). Through interviews and co-design sessions with d/Deaf signers and ASL interpreters, Ang et al. [26] reinforced that DHH signers and interpreters prefer having other signers in their view. Still, current videoconferencing platforms provide less flexibility in layout customizations. Interpreting linguistic information in sign language becomes more difficult as the size of video tiles decreases with the increasing number of participants. Further, keeping the view of the active speaker, interpreter's video, and captioning text in visual proximity to each other can be challenging [13, 36]. Mack et al. [17] used autoethnographic methods to reflect on their virtual work experience in a mixed-ability team and reported being unable to see participants who used sign language and giving more visual space to the shared screen, resulted in losing sight of the speaker or interpreter. Ang et al. [26] recommend adding the flexibility to rearrange and resize video tiles and the ability to group and pin together video tiles. To reduce the burden on DHH users when consuming information from multiple sources, the option to overlay semi-transparent video over shared workspace has also been suggested [22].

DHH individuals also use *speechreading*, a technique that relies on visual and contextual cues to observe the movements of the speaker's lips to support communication. However, prior studies have shown that DHH individuals often find speechreading challenging in videoconferencing, especially when the speaker's face is less visible, there is a lack of eye contact, or background lighting is insufficient [10, 14, 36]. A participatory design study by Kim et al. addressed these issues by providing a zoomed-in portion of the speaker close to their regular video tile, and in case of screen share, suggested that passive participants in the call be removed from the visual layout to reduce distractions [13].

2.2 Captioning in Videoconferencing

Due to speechreading challenges in video-mediated communication, DHH individuals often rely on captions, often against their preference [13]. Videoconferencing platforms use automatic speech recognition (ASR) for live captions and transcriptions, which can benefit DHH users when human interpreters and captioners are unavailable. As ASR output can be erroneous, specifically for non-native English speakers, DHH users face challenges with it [11]. Seita et al. conducted a remote study with DHH and hearing participant pairs to derive designs that let hearing people identify errors in ASR output and correct them [30]. Apart from fixing ASR-related errors, McDonnell et al. found that in small-group conversations involving mixed hearing identities, DHH participants suggested speaker identification and warnings for overlapping speakers to be built into the videoconferencing system [19], and Seita et al. found that DHH participants were more satisfied with communication wherein hearing individuals maintained neither a high nor a low speech rate [29]. While exploring future captioning designs with DHH participants, prior work discussed features like color coding speakers, having the ability to keep captions close to the active speaker, using visual or haptic means to get people's attention and notify hearing individuals to change their behavior [26, 30].

2.3 Audio-Centric Videoconferencing Designs

Given the audio-centric nature of videoconferencing designs, hearing people can gauge the listener's understanding by receiving verbal backchannel feedback [26]. *Backchannels* are verbal or non-verbal feedback given while someone is talking to show interest or attention. However, consuming backchannel feedback by DHH participants, like head nods and other non-verbal cues, is challenging and physically tiring due to the need to constantly pay attention to everyone's video tiles, as videoconferencing platforms highlight active speakers solely based on audio. This also results in DHH participants' video tiles never getting displayed or highlighted because their interpreter speaks for them [36]. Although an interpreter is essential to facilitate conversation between DHH and hearing participants, it not only created frequent conversational lags that discouraged DHH people's participation, but it also complicated efforts for the participants to identify deaf signers [37]. To address that, Kushalnagar and Vogler suggest that videoconferencing organizers should avoid making assumptions and ask DHH users about their preferred accommodations, captioning, and interpreter preferences [14].

Other prior works [14, 37] suggest having procedures and guidelines to manage turn-taking, having instructions on how to make meetings accessible, asking speakers to identify themselves, reminding participants to sit in well-lit areas, and requesting that they wear headphones with a microphone to improve audio and automated speech recognition quality.

All these prior studies use methods like participatory design, co-design, interviews, and autoethnography to identify communication challenges in mixed hearing groups and suggest design recommendations. Our work builds upon these recommendations to design and build a novel videoconferencing platform called *Jod* and to evaluate it by simulating real-world contexts where an interplay of social, environmental, and technological factors exists simultaneously.

3 *JOD*: SYSTEM DESIGN & IMPLEMENTATION

Jod's features were iteratively designed using a combination of findings from prior work (refer to Table 2 in Appendix) and feedback received from the participants in the first user study session we conducted. It additionally implements many common features of current videoconferencing platforms (e.g., chat, automated speech recognition for live captions and transcriptions, emoji-based reactions, mute, video on/off indicators, and highlighting active speaker's video tile). Figure 2 shows a screenshot of a video call on *Jod* with six active users (3 DHH, 1 interpreter, 2 hearing individuals). The top panel contains gesture and call control bars. The right panel lets users switch between People, Chat, and Transcription tabs. The remaining visual space is used for rendering video tiles and the captions box.

3.1 Features

We now describe the design of *Jod*'s key features:

Customizable Visual Layout. Multiple studies on challenges in videoconferencing for DHH users [13, 14, 26, 36] have identified that current platforms offer limited layout customization. They provide default layouts to choose from and automatically resize and rearrange video tiles. *Jod* provides customizability to users such that they can reorganize their visual layout to suit their personal preferences. All video tiles, including the participant's tile, captions box, and screen share, can be resized, added/removed, and moved anywhere in the visual layout. To resize a video tile, users click and drag the white handles on its corners (Figure 3a); to change a tile's position, they click anywhere on the video tile (except the corners) and drag it to the preferred location. Users can also fix the position and size of any video tile(s). To do so, they hover over the tile, and three buttons appear in the top-left corner (Figure 3a). The first button provides a locking feature (same as pinning) that disables resizing and fixes the particular tile's position. Users can unlock a video tile by clicking again on the same button to allow resizing and repositioning. To reduce visual clutter, users can either click on the second button to remove the video tile from their layout, or the third button to turn off the video stream. To add a removed video tile back to their screen, users need to click the "+Add" button in the People tab (Figure 2).

Preset Feedback Messages. The audio-centric nature of videoconferencing platforms makes it difficult for d/Deaf signers and interpreters to grab other signers' attention. In physical settings, they can use Deaf cultural practices, like banging on a desk or flashing lights on and off, to get attention; however, such practices do not translate well to online settings [26]. Studies [11, 29] have also discussed DHH signers' frustrations with speaking behaviors, such as speaking too fast or at a low volume. In *Jod*, hovering over a user's video tile results in six buttons to appear in the bottom-right corner of the tile (Figure 3a). These buttons can be used to send preset feedback messages, like "Please look at me", "Please keep your upper body visible", "Please turn on some lights", "Please speak slower", "Please use easier language", and "Please repeat what you said". When a message is sent, it is displayed as a toast element in the recipient's UI. To secure the receiver's attention, these notifications do not auto-dismiss. The recipient must click on them to close them.

Active Signer Identification. Videoconferencing platforms use speaking indicators to highlight the active speaker, e.g., a bright border around the video tile. This feature does not work for d/Deaf signers because the video tile of the interpreter—who is 'voicing' them—gets highlighted. We utilized a Wizard of Oz method to study this feature in *Jod*. A researcher joined the calls as "Admin," a special participant type, and used an admin panel to indicate when Deaf users started or stopped signing. For participants on the call, this appeared as if the video tiles of signing and speaking users were highlighted similarly (i.e., with a blue border around the speaker or signer's video tiles).

Accessibility Indicators. In mixed ability groups, users may need indicators to understand another user's accessibility needs. Furthermore, call participants may find it difficult to remember the appropriate accommodations in such group settings (e.g., remembering to speak slowly) [17]. *Jod* lets users quickly gauge the ability of others using explicit indicators. While joining a call on *Jod*, users select their participant type (Deaf, Hearing, or Interpreter). These abilities are indicated in the user interface through different colors, icons, and, for interpreters, an explicit "Interpreter" label.

Enhanced Transcription. Currently, transcriptions and captions in videoconferencing platforms only contain automated speech recognition output. To provide users with a holistic view of the conversation, *Jod* enhances audio transcriptions and captions to include preset feedback messages, emoji reactions, and information when a DHH user starts/stops signing (Figure 3b). Transcription text also displays the accessibility indicator of each participant.

Gesture Recognition. Hearing people can gauge if others are listening and following their conversations in online settings because of their ability to receive verbal backchannel feedback along with non-verbal cues. To increase the ways users can give feedback while being on mute, we added four emoji-based gestures: clap, hand raise, okay, and thumbs-up. Users enable this feature by clicking on "Enable Gestures" in the gesture control bar (refer to Figure 2). When a gesture is recognized, a circular progress bar gets rendered around that emoji. Once the circular progress bar is completed (in ≈1 sec), the emoji is sent to everyone on the call. (Note: Emojis can also be sent by directly clicking one of the four emoji buttons.) Zoom has a similar gesture recognition feature, however, no prior work exists on how users use them in mixed hearing conversations.

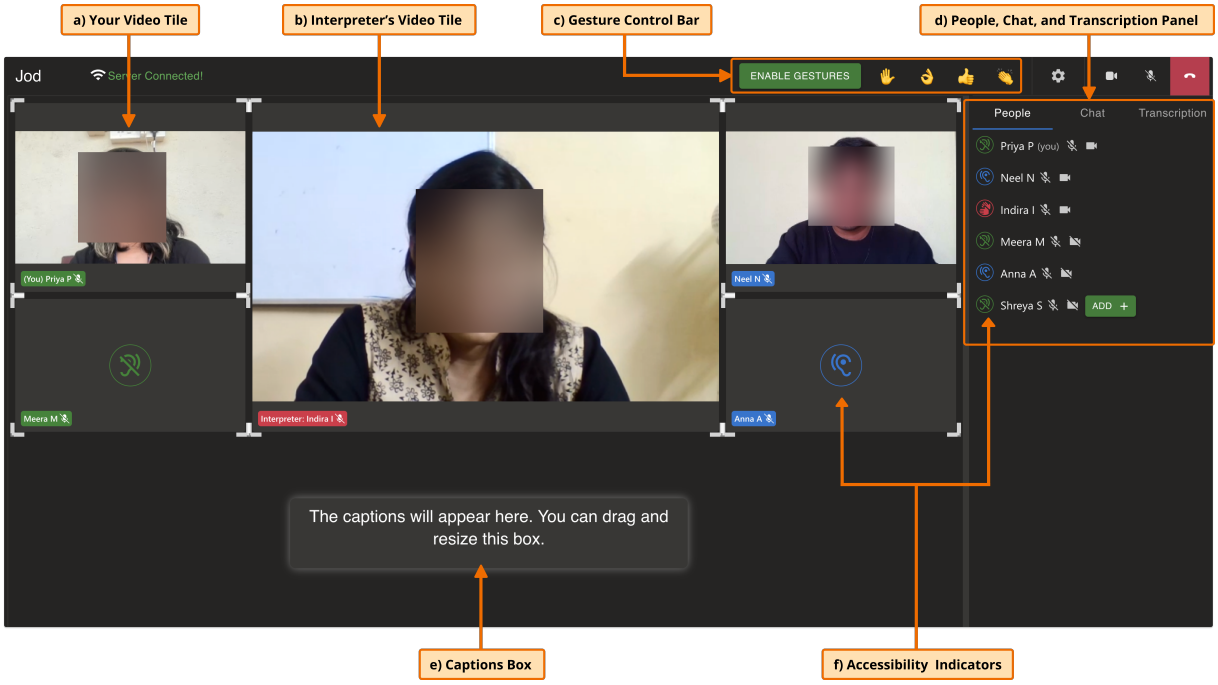
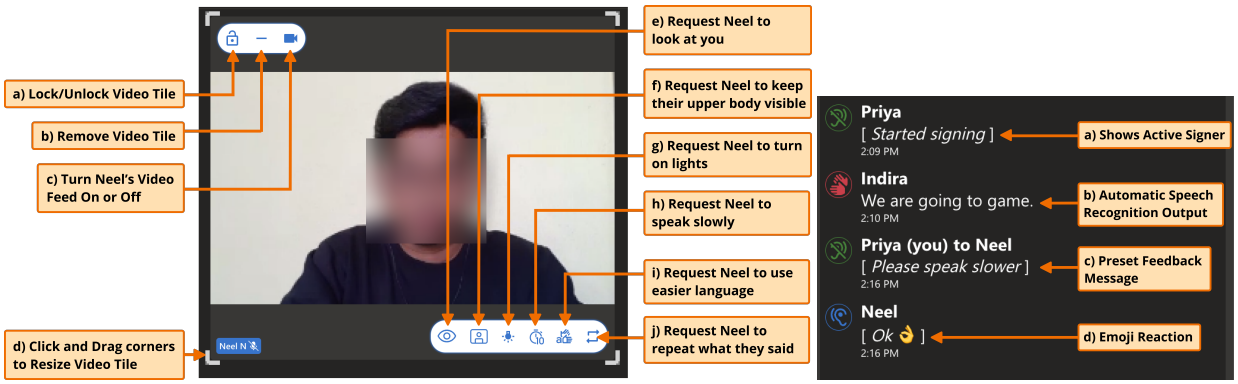


Figure 2: User Interface of *Jod* with six participants on a simulated video call. 6 participants comprise 3 DHH, 2 Hearing, and 1 Interpreter. (Participant names are pseudonyms)



(a) Video Tile Hover State **(b) Transcriptions Panel**

Figure 3: *Jod*'s System Features: Customizable Visual Layout, Preset Feedback Messages, and Enhanced Transcriptions

3.2 Implementation Details

Figure 4 provides a high-level overview of *Jod*'s architecture. *Jod* is developed to be accessible through a web browser. To enable group calling and group chat, we used Microsoft's Azure Communication Services (ACS) and Socket.IO. The user-facing component of the application, i.e., the client, was built using React, while the server was developed over Node.js and Express.js. As shown in Figure 4, to join a group video call, the participant first opens *Jod* in a browser. A request is sent to the server (1) to get a list of possible sessions the participant can join. Each session holds unique configuration identifiers that ACS needs for group calling and chat functionalities. (2) The participant is then prompted to fill in profile details (e.g. full name, session name, participant type), and this information, along with the client's unique socket identifier, gets stored in a MongoDB database. (3) Using the unique identifiers, the client sends ACS a request to join the group call and chat. Finally, ACS (4) authenticates the participant's request, adds them to the group call, and starts sending call- and chat-related information to the client. *Jod* uses Socket.IO to power its *preset feedback messages* feature for client-to-client communication. Sign detection is a complex problem [5]; due to non-existent off-the-shelf AI models that could detect signing with high accuracy, we resorted to a Wizard of Oz method for *Jod*'s *active signer identification* feature. *Jod*'s *enhanced transcription*

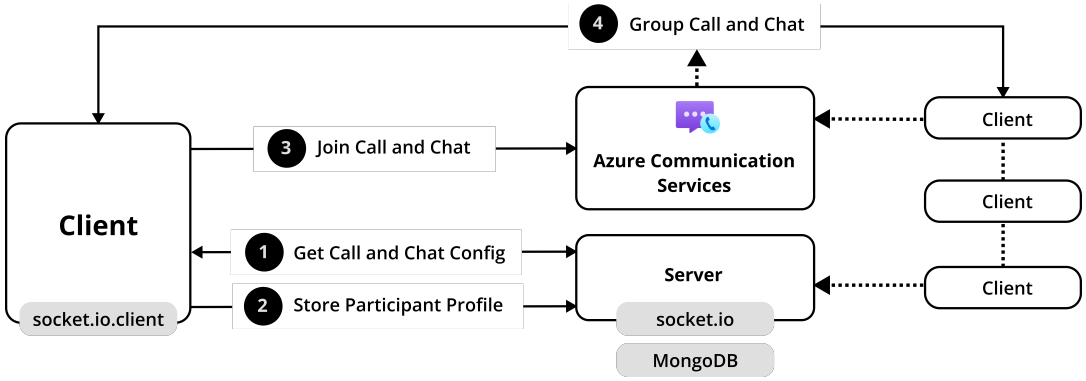


Figure 4: System Architecture of *Jod*

feature and live *captions* used ACS in-built automated speech recognition pipeline for English. The *accessibility indicator* icons for each participant type are from the Material UI library. For the *gesture recognition* feature, we built an AI pipeline to run within the client's browser. We used Google's MediaPipe Holistic model to track hands and ran a post-processing function to further classify each gesture. *Jod* uses accessible colors, adhering to web content accessibility guidelines (WCAG).

With respect to logging, *Jod* collects telemetry containing visual layout events that were logged when participants altered their layout arrangement by dragging, resizing, removing, or adding any video tile. Using this extensive log data, we could recreate a participant's visual layout including the location, arrangement, and size of each video tile. Additionally, *Jod* logs preset messages, gestures, chat messages, and click-based emoji reactions.

4 STUDY DESIGN

To investigate the behaviors and perceptions of users navigating mixed hearing conversations on *Jod*, we conducted six user study sessions (S1-S6) involving 34 participants, as detailed in Table 1. Out of the six sessions, four were conducted in person, while two were conducted remotely. Our study was approved by our Institutional Review Board (IRB) and took place between Nov-Dec 2022.

4.1 Participant Recruitment

Out of the total 34 participants (13 Female, 21 Male), 18 were Deaf or hard-of-hearing (DHH), 6 were sign language interpreters, and 10 were hearing individuals. Demographic information and session details are listed in Table 1. All hearing participants were recruited through the authors' personal and professional networks. For the remote sessions, interpreters were recruited through our professional network and DHH individuals from the National Institute of Speech & Hearing (NISH), an institute for the education and rehabilitation of individuals with speech-language and hearing impairments. For in-person sessions, DHH individuals and interpreters were recruited through our partner organization, WinVinaya Foundation, a nonprofit organization and skills training center for persons with disabilities in Bengaluru, India.

We compensated DHH participants with an INR 750 gift voucher upon completion of the study session. Interpreters were compensated with INR 2500 per session, calculated per the standard cost of interpreting services in India. All our participants had previously used video calling applications (e.g., Zoom, Microsoft Teams, Google Meet, Google Duo, WhatsApp). For 2 DHH participants and 7 hearing individuals, this was their first time video conferencing in a mixed hearing group setting. 17 out of 18 participants identified as Deaf and 1 participant was hard of hearing. ISL was the primary mode of communication for the Deaf participants. 10 out of 18 could speechread in regional languages and 3 of these 10 participants were beginner-level speechreaders in English.

4.2 Study Setup

Each study session was approximately 2.5 hours long and involved 3 DHH signers, 1 or 2 interpreters, 2 hearing individuals, and 2 hearing researchers. While one researcher moderated the call, the other acted as a wizard who was not visible to the participants on *Jod*. We began our sessions by sharing a tutorial of *Jod* followed by task-based explorations, unstructured conversations, games, and a presentation round with screen share. We concluded with a focus group discussion. The study protocol remained consistent for both remote and in-person sessions.

In-person Sessions. We conducted in-person sessions because of two reasons – (1) to ensure DHH participants were comfortable and familiar with the study space, and (2) to adjust to any unanticipated system breakdowns and quickly iterate over the study protocol if needed [18]. We conducted four sessions in-person at the nonprofit organization (S2, S4, S5, and S6). In a large open space, we positioned three tables with two chairs at each table and assigned specific seats to each participant to minimize echo and interference. To maintain the ecological validity of our study and to prevent direct communication, we ensured there was no sound or visual bleed between participants outside of *Jod*. Hearing participants were seated farther apart, and moderators were seated next to deaf participants. We provided participants with laptops and earphones. We also provided notebooks and pens to all participants for note-taking and drawing. We had one interpreter per in-person session; to avoid interpreter fatigue, we took necessary breaks based on the recommended guidelines followed at the nonprofit organization. Researchers moderating were also present in-person and helped answer any participant questions during the sessions.

Session 1 (S1)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P01	22	M	Profound	Intermediate
P02	25	M	Profound	Intermediate
P03	22	M	Profound	Intermediate
P04	26	M	Interpreter (7 years)	Expert
P05	29	F	Interpreter (7 years)	Expert
Session 3 (S3)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P12	23	F	Profound	Expert
P13	22	F	Mild	Intermediate
P14	22	F	Profound	Expert
P15	24	M	None	None
P16	25	M	None	None
P17	34	M	Interpreter (7 yrs)	Expert
P18	29	M	Interpreter (6 yrs)	Expert
Session 5 (S5)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P24	24	F	Mild	Novice
P25	25	M	Mild	Novice
P26	21	F	Profound	Novice
P27	22	M	None	None
P28	24	F	None	None
P11	35	F	Interpreter (15 yrs)	Expert

Session 2 (S2)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P06	30	M	Moderate	Intermediate
P07	28	M	Mild	Intermediate
P08	25	F	Profound	Expert
P09	22	M	None	None
P10	42	F	None	Novice
P11	35	F	Interpreter (15 yrs)	Expert
Session 4 (S4)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P19	28	M	Moderate	Expert
P20	22	F	Profound	Intermediate
P21	25	M	Moderate	Intermediate
P22	35	M	None	None
P23	24	F	None	None
P11	35	F	Interpreter (15 yrs)	Expert
Session 6 (S6)				
ID	**Age**	**Sex**	**Hearing Loss/Role**	**ISL Proficiency**
P29	24	M	Moderate	Intermediate
P30	22	M	Mild	Intermediate
P31	23	M	Profound	Expert
P32	24	M	None	None
P33	25	F	None	None
P34	26	M	Interpreter (3 yrs)	Expert

Table 1: Detailed Participant Demographics

Remote Sessions. We conducted two remote sessions (S1 and S3). All participants joined the sessions from their homes and used their personal laptops. The initial introductions, *Jod* onboarding, and focus group discussions were conducted on Zoom. The remaining study-related parts took place on *Jod*. Two interpreters took turns interpreting and switched every 20 to 30 minutes.

4.3 Procedure

Each session began with introductions and an overview of the research study. The moderators explained how user data would be collected and asked for verbal consent. Throughout the study, the interpreters and deaf participants communicated in ISL, while the moderators, hearing participants, and interpreters communicated in English. Communication between DHH participants and others was facilitated by interpreters. Sessions consisted of the following six key components, listed chronologically:

Jod Onboarding (~10 mins). To provide consistent training to all participants, we played a ~5 minute video tutorial on YouTube. The tutorial showed one of the authors using *Jod* and introducing its key features; it included a voice-over and closed captions. Additionally, an interpreter was present to facilitate communication.

Round 1: Task-based Feature Exploration (~30 minutes). After watching the video tutorial, participants were given an opportunity to ask clarifying questions. Once all participants were ready, they joined the call using *Jod*. After successfully joining the call, both researchers (moderator and wizard) also joined the call. The goal of this round was to familiarize participants with the system and let them interact with its features. To facilitate this, the moderator prompted participants by assigning 10 tasks, one after another. Participants were asked to send a "like" reaction after completing each task so that moderators knew when to proceed to the next one. Examples of tasks included "*Make participant X's video tile bigger,*" "*Inform me (the researcher) to turn ON background lights,*" and "*Perform raise hand gesture.*'. The Appendix includes the complete list of tasks. At the end of this activity, participants were given 5 minutes to freely explore the system and capture a screenshot of their preferred video-tile layout arrangement.

Round 2: Unstructured Conversation (~15-20 minutes). To encourage free-form conversations between DHH and hearing participants, the moderator initiated a casual conversation on food preferences. It further progressed to include topics like social celebrations, cities, and occupations.

Round 3: Game of Charades (~15-20 minutes). During the first in-person session (S2), we observed a lack of direct communication between DHH and hearing participants. To bridge this gap and initiate intermingling across the two groups, we added a modified version of the game of charades to the last three in-person sessions. The moderator divided participants into two teams based on hearing abilities, DHH and hearing, then provided a movie title that one team had to act out, and the other team had to guess. For example, a hearing person would enact to the DHH team, whereas a DHH person would enact to the hearing team. To ensure fair play, the participants were not allowed to sign alphabets or numbers and instead were encouraged to act out movie scenes. They used the chat tab to type their guesses.

Round 4: Screen Share Presentation (~7-8 minutes). To capture participant behaviors on customizable video tile arrangements, the moderator used screen sharing to give a 5-minute talk. She shared slides about an app for sign language users and learners. In the end, all participants were asked to capture a screenshot of their video tile arrangements while viewing the shared screen.

Focus Group Discussion (~60-90 minutes). After completing the preceding rounds, the researchers conducted a focus group discussion (FGD) with all participants to capture their general perceptions of *Jod* and gather detailed feedback on key features. In-person FGD participants gathered around in a circle. Interpreters had a dual role — as study participants and interpreters. For remote interpreters, FGDs were held on Zoom. Each FGD started with open-ended questions on the overall experience of using *Jod*. We then delved deeper into interactions and experiences with specific features, what participants liked vs disliked, and suggestions for additional features in future iterations.

We included a set of varying interaction scenarios because *Jod*'s features are intended to be general purpose, and we wanted to examine their interaction leading to emergent behaviors across scenarios. Screen sharing and non-screen sharing scenarios have been highlighted in prior work [13, 26]. We introduced charades because it requires social interaction that helps establish comfort levels among participants, similar to the Twenty Questions game used by McDonnell et al. [20].

4.4 Data Analysis

We analyzed the qualitative data, which consisted of ~7 hours of audio recordings from five focus group discussions (S2, S3, S4, S5, S6), researchers' detailed handwritten notes, participants' screenshots of *Jod*, participants' notes, and pictures clicked at the in-person study site. Audio recordings were anonymized and transcribed soon after the sessions were conducted. FGD data were analyzed using reflexive thematic analysis, as described by [6]. The field data were read several times by the first two authors to identify the initial set of codes. Multiple rounds of open coding were conducted, and codes were rigorously discussed between authors for prioritization and grouping into themes. To avoid imposing biases while analyzing the data, we refrained from using existing theoretical frameworks or lenses. Instead, we let the themes emerge bottom up. For quantitative analysis, we used telemetry data from all the rounds except task-based feature exploration round[2]. To understand the relationship between participants' ability and how they used the available screen real estate, we grouped all active participants in the call based on their ability and calculated the average video tile size. For each participant, we extracted the layouts they used for the longest duration per minute and calculated the average video tile size across round(s).

4.5 Authors' Positionality

Seven of the eight paper authors are of Indian origin and have conducted fieldwork with diverse marginalized groups in India. Four authors identify as female, and four as male. One author is a staff member of the nonprofit partner organization and has significant experience working and training deaf individuals for employment opportunities. Three authors have more than two years of research experience in studying the accessibility needs of the d/Deaf or hard of hearing (DHH) community in the Global South; one has 10 years of research experience in accessibility in North America. Our approach to this research was informed by our individual experiences working with the DHH community in India and interacting with them over video calls.

[2]S1 was a design feedback session. There was telemetry data loss during S2 and S3. We used data from S4, S5, and S6 in-person sessions for the layout-related telemetry analysis. We intend findings from the qualitative data to be our primary focus and consider telemetry data only as a valuable supplement to our qualitative analysis.

5 FINDINGS

The *Jod* system was used for ~ 10 hours across the six study sessions. Participants rearranged their visual layout 485 times, sent 40 preset feedback messages, and conveyed 30 emojis via gestures. Below, we discuss our key findings, focusing on flexibility and diverse choices of visual layouts across the different participant groups, notifications sent through preset feedback messages to influence other participants' behavior, and cultural nuances and mismatched expectations in mixed hearing settings.

5.1 Using Layout Flexibility in Videoconferencing

Jod offers users control over their visual layout, e.g., participants' video tiles, captions, etc. Our participants leveraged this flexibility to tailor the platform to their hearing ability and the continuously changing group communication context, which resulted in constant trade-offs between user labor and system efficiency.

5.1.1 Agency to Customize Layout. In the task-based feature exploration round, we asked participants to explore the system and organize their visual layouts per their preference. In response, they actively interacted with the customizable elements and rearranged the video tiles of everyone, including their own. For DHH participants, we observed that the interpreter was a priority and essential to their communication on the platform; as $P19_{DHH}$ mentioned, *"I really like the option that I can resize the interpreter and see it clearly,"* and $P29_{DHH}$ described his layout choice:

> "I first chose the interpreter and made their tile bigger because the speaking people are not my priority...the interpreter is my priority. Being deaf, I want the interpreter screen to be big." – $P29_{DHH}$

Personal priorities were reflected in the layout arrangements (Figure 5) across participants. On comparing the sizes of the video tiles, DHH participants accorded to other individuals on the conference call; we found that they allocated maximal visual space to the interpreter (Figures 5c - 5f and Figure 6a). For DHH participants, a Kruskal-Wallis test revealed a significant effect of *participant ability* on average video tile size ($\chi^2{}_3 = 24.99, p < 0.0001$). A pairwise comparison using Wilcoxon rank sum test with Bonferroni correction showed significant differences between the interpreter's video tile size and (1) DHH participants ($Z = -3.64, p < 0.01$), (2) hearing individuals ($Z = -3.68, p < 0.01$), and (3) their self-video tile ($Z = 3.74, p < 0.01$). Similarly, in the screen share presentation round, the interpreter's video tile remained significantly different except relative to the screen share video tile (Figures 5i - 5k

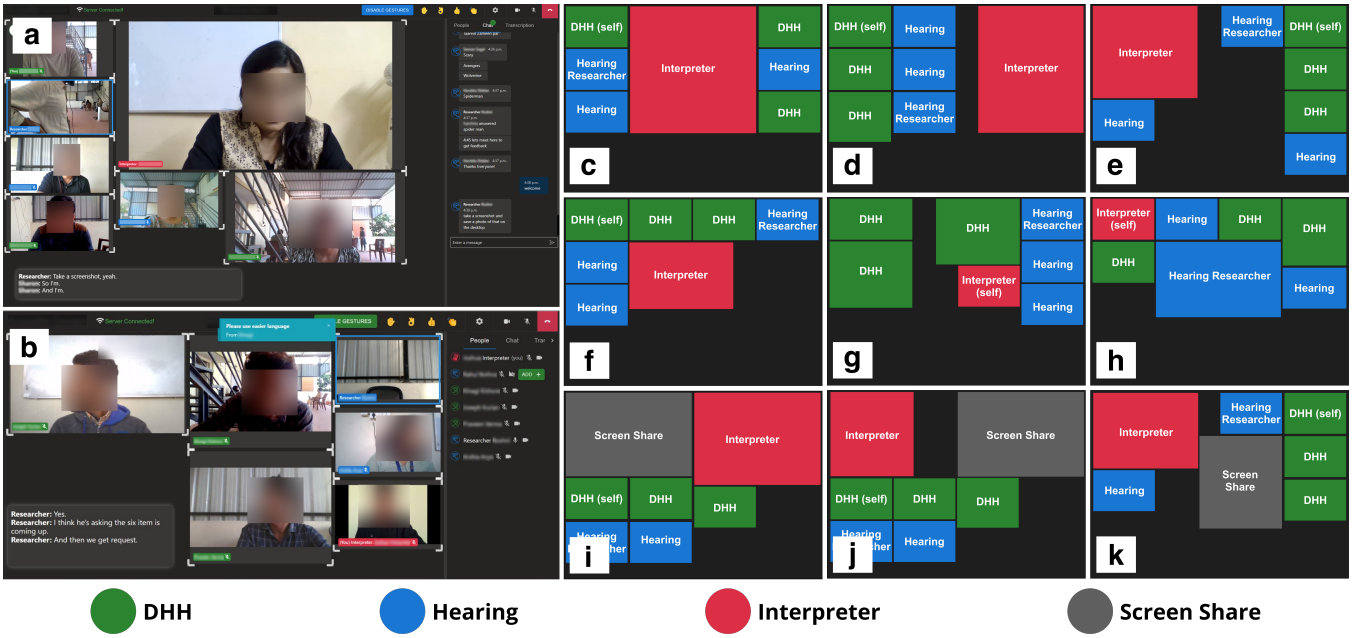


Figure 5: Screenshots of *Jod* from study sessions and visual layout abstractions generated using telemetry: (a) DHH participant's layout where the interpreter's video tile is the largest, (b) interpreter's layout with DHH participants' video tiles larger than others and one hearing participant's video tile removed, (c) to (f) are examples of other arrangements DHH participants created keeping the interpreter's video tile largest, (g) interpreter's layout where DHH participants' video tiles were enlarged, (h) interpreter's layout where hearing researcher's video tile was enlarged, and (i) to (k) are examples of arrangements DHH participants created when screen sharing was active, with interpreter's video tile and screen share competing for visual space.

and Figure 6b). This suggests that DHH participants gave equal importance to the interpreter and the screen share. The participant's ability had a significant impact on the interpreter's video tile area ($F_{1,13}$ = 5.1473, p ≈ 0.04). Comparing the size of the interpreter's video tile between DHH and hearing participants, we found that the interpreter tile in the DHH participants' visual layouts (59.7±22.2) significantly exceeded the size the hearing participants gave to the interpreter (33.7±20.9), with t=−2.3, p≈0.04. This was also true in the screen share presentation round.

In addition to allotting prominent visual space to the interpreter, DHH participants discussed their layout choices for organizing other DHH and hearing participants. While some preferred to keep all participants on the screen, with hearing participants occupying minimal visual space, others chose to remove hearing participants entirely. For instance,

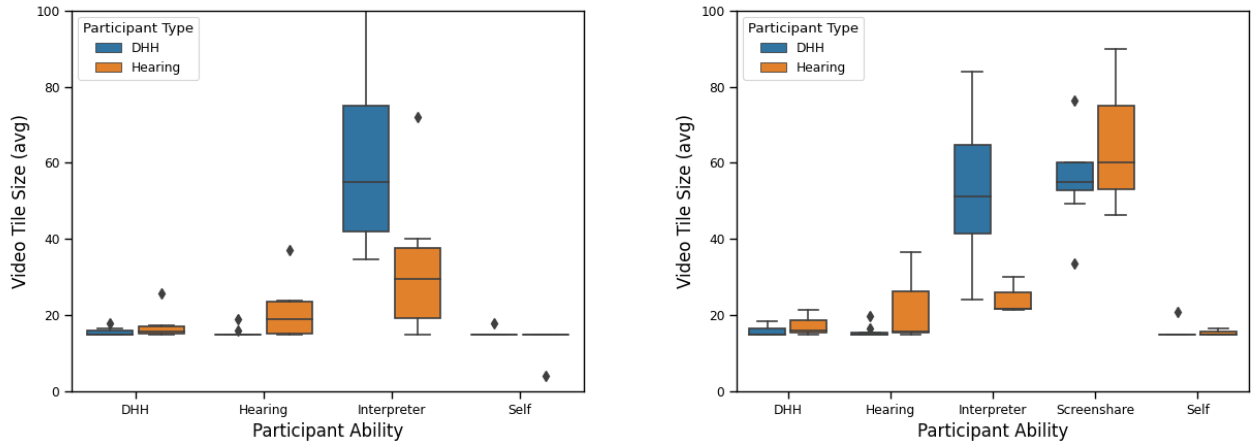
> "I would only want to see the deaf participants... so I can have all the deaf participants and the speaker (interpreter) on the screen. This allows me to manage the screen so the interpreter and the deaf participant are side-by-side." – $P13_{DHH}$.

Similarly, we found that interpreters resized the video tiles of DHH participants and the researcher conducting the study session, making them bigger than the other video tiles (Figure 5g, 5h). This behavior was motivated by the need to follow the DHH participants' signing and facial expressions, as $P34_I$ noted, *"My main priority was to see the deaf candidates clearly and understand what they are signing... if their tile is very small, then I would not be able to understand their signs properly."* All but one interpreter kept all the hearing participants on the screen; $P18_I$ surrounded their video tile with DHH participants and removed all other hearing participants except the hearing researcher. Another interpreter, $P11_I$, kept the "*deaf participants on the top... to see all their reactions.*"

In addition to rearranging participants' video tiles, participants actively interacted with other customizable visual elements, such as closed captions and the screen share tile. DHH and hearing participants interacted and reorganized the captions (Figure 5a). For instance, $P33_H$, a hearing participant, described her arrangement of the captions and the interpreter's video tile to grasp the ongoing interpretation better:

> "I arranged it like... I had all the hearing people (on the left side), deaf people (in the center), and the interpreter (on the right side), and the captions below that. I made the captions and interpreter larger so that I can keep up with the interpreting and make sense of how the words are being interpreted." – $P33_H$

This flexibility to reorganize multiple visual elements augmented the participants' communication abilities and facilitated comprehension. Most participants felt agency and control to align the *Jod* platform to their personal preferences. As $P12_{DHH}$ shared, *"It was very independent. I could resize whoever I want. Like the hearing people, I could move them aside... put them below the deaf people. It was very good overall."*



(a) No Screen Share **(b) With Screen Share**

Figure 6: Video tile sizes (average) in DHH and hearing participants' visual layout

5.1.2 Adapting to Dynamic Group Communication Context. Besides aligning *Jod* according to their hearing abilities, participants did on-the-fly visual rearrangements to keep pace with the continuously changing group communication context. Dynamic rearrangements also supported participants in keeping their video layout organized and helped them prioritize the active speaker/signer. For instance, the ease of dynamic rearrangement helped a hearing participant prioritize different players responsible for acting during the game of charades:

> "My usual goal was to keep as few tiles as possible on the screen. I would usually just have the researcher's tile who was speaking... on the right side. On the left side, I would have the interpreter's tile just out of curiosity to see how the interpretation was going on. And closed captions running at the bottom. While playing charades, whoever was doing sign gestures, I would just add their tile." – $P32_H$

In addition, *Jod* conveyed someone's signing by highlighting the active signer's video tile and adding the "*started signing*" message in both the closed caption and transcription. $P11_I$, who had been interpreting for 15 years, shared that it helped her track the signer(s) since it is challenging to keep track of who is signing on video calls. It also helped her envision future notification modalities to help with attending to the active signer.

> "Let's say in a group of 30 hearing and 2 DHH folks, it is hard to keep track when someone starts signing... but as it appears [in the] captions box, I can keep track. There should be a way to notify the interpreter that someone started signing to focus on their video tile." – $P11_I$

$P32_H$, a hearing participant, recalled that he preferred a minimum number of video tiles in his layout, he had the researcher (who was the active speaker), the interpreter's video tile, and the close captions running in the bottom. However, during charades, the "started signing" message helped him identify whose video tile to bring back to the visual layout. Other participants also described dynamically adding and increasing the visual space of the active speaker's video tile. For instance, $P25_{DHH}$, who would usually remove the hearing participants from her video layout, said,

> "If they were speaking or asking some question, I would bring them to the screen – otherwise, I would just remove them from my screen." – $P25_{DHH}$

While having the active hearing participant on screen was not a necessity for most DHH participants, they engaged in such usage patterns when provided with an easy option to do so.

5.1.3 Offering Flexibility through Multiplicity. *Jod* also offered flexibility to its users through the multiplicity of various videoconferencing features. Users can understand the context of ongoing conversations by following the interpreter and reading the automated speech recognition output, either in closed captions or transcriptions. We found that while conversing with hearing individuals, DHH participants (like $P29_{DHH}$) simultaneously referred to the interpreter's video tile and the transcriptions/captions. Transcriptions were preferred to catch up on conversations, while real-time captions were used to verify if anything was missed by the interpreter or lost in translation. As $P19_{DHH}$ explained, *"Both are useful and [I] used both. Because cc (closed captions) happens in real time... if I have forgotten something, I could go up and see it in the transcript"*. Captions also served as a fallback mechanism for DHH participants to continue conversations when the interpreter was unavailable; as $P13_{DHH}$ noted, *"If there was an internet lag and the interpreter froze, I could look at captions"*. Interestingly, a hearing participant, $P32_H$ also referred to the transcriptions when he "*missed something in the captions*", e.g., when someone used a reaction that he missed because it went away too fast.

In addition, participants could communicate emoji reactions (like thumbs-up) either via signing (through *Jod*'s gesture recognition

feature) or by clicking on the emoji icons. In the study sessions with telemetry data for emoji reactions (i.e., S3–S6), people sent a total of 72 emoji reactions, of which 30 were through AI gesture recognition and 42 were click-based. DHH and hearing participants felt that gesture recognition took too long to send a reaction: "*It wasn't super useful for me, partly because it took like 4s to detect. So, keeping my hand raised in the air for 4 seconds? It's easier for me to just click that button.*" – $P32_{DHH}$. Participants also brought up instances when there were false positives; as $P28_H$ mentioned, "*when I was holding my pen up... it recognized it as a thumbs-up gesture.*"

5.2 Connections through Notifications in Mixed Hearing Settings

Jod introduces a novel way to disseminate notifications, i.e., through Preset Feedback Messages, in mixed hearing conversations. We detail how these notifications helped connect the DHH, hearing, and interpreters. Though they generally enhanced communication among participants, they were occasionally found to be obtrusive and to leave the sender in limbo due to a lack of acknowledgment of receipt mechanisms.

5.2.1 Connecting DHH, Hearing, and Interpreter. When asked about using feedback messages, participants recounted being able to connect with others of different hearing abilities without interrupting the ongoing conversation. Prior studies highlighted that DHH participants seek minimal clarifications to avoid interrupting the conversation [26]. However, *Jod* let DHH participants overcome this. As $P12_{DHH}$ said, *"This is much better because it does not distract other people – I could just directly send them the feedback – can you please repeat – so that was really good, actually, very different."* Moreover, these notification mechanisms helped DHH participants connect with interpreters in multiple ways, from flagging their attention to requesting better background lighting. For instance, $P19_{DHH}$ noted, *"One issue we always have is the issue of getting the interpreter's attention or getting another deaf person's attention in sessions when there are deaf on the call."* To this, $P11_I$ added,

> "...therefore, they (DHH) always flash on the camera. If some person is talking and they want that person's attention, they do <flashing>. However, if I specifically want $P19_{DHH}$'s attention when there are 50 people, I would repeatedly do <flashing> and his sign name. If he sees me, he'll say $P11_I$. So that's how we would get each other's attention." – $P11_I$

However, while using *Jod*, $P29_{DHH}$ instead chose to use the *"Please look at me"* feedback message to capture the interpreter's attention.

In addition to augmenting the communication between DHH participants and interpreters, these feedback messages also helped the DHH and hearing participants converse directly. $P22_H$, who had never previously conversed with DHH individuals, shared how preset messages helped him directly converse with a DHH participant:

> "Another interesting feature I realized initially – the way I arranged my screen, I removed my tile. I was like why should I see my own tile. Instead, I will make everyone else bigger. I think someone sent me a message saying to be more visible. I realized that I should put my video tile back so that I can reorient." – $P22_H$

He further described it as a *"new kind of experience"* and agreed with others that it made things easier by removing interpreters from the loop for these conversations.

5.2.2 Capturing attention and acknowledging notification messages. While these notification mechanisms aided in capturing attention, several participants highlighted two major shortcomings: its obtrusive design and lack of acknowledgment of receipt mechanisms. When asked about their experience with feedback messages, participants reported the notification messages often overlapped with the video tiles, which felt visually *"distracting."* For instance, $P33_H$ shared,

> "The way it was coming, it was actually coming in the middle of the screen, and a lot of notifications were coming together. So that was a bit distracting from what was going on. So maybe if it comes on the side or in the chat, then that would be better… because I was missing what people were signing/speaking on the screen. There were a lot of notifications, and unless I went and clicked on them, it did not disappear." – $P33_H$

Participants must click on notification messages to dismiss them and clear their screens. Additionally, such visual distractions were particularly challenging for interpreters, requiring them to pause their signing, possibly resulting in information gaps momentarily. As $P11_I$ mentioned, *"I think [$P25_{DHH}$] sent 'turn your lights on' to me thrice by mistake, and that remained on my screen... So I had to put my sign down to disable all those three notifications. I had to manually click on the notifications to disable it."* As a result, multiple notifications hampered the ongoing interpretation, causing the interpreter to miss the signing.

Participants also highlighted the lack of acknowledgment or receipt mechanisms for these notifications. This resulted in participants being unsure about whether the receiver received their sent feedback message, as $P12_{DHH}$ mentioned

> "When we click – can you please repeat – to send it to the interpreter, there is no feedback feature to know if the interpreter has actually received that message... The message has been sent to the interpreter, but how does the sender know that the interpreter has received that message?" – $P12_{DHH}$

To overcome this, one DHH participant manually clarified his confusion and *"asked [$P11_I$] if she got a notification, and she said yes."*

In addition to manual interventions seeking acknowledgment of notification receipt, participants suggested their desire for automated ways to acknowledge notifications. For instance, $P28_H$ suggested, *"I would want it to be acknowledged. If I am on the receiver's end, then I would want to acknowledge it – am I in a position to do that? Have I made that change? Can I not make the change? Will I make it later?"* This demonstrates that acknowledgment extends beyond mere confirmation; it is equally important for the receiver to inform the sender if, how, and when they will respond to the

request. However, $P11_I$, an interpreter, expressed reservations about this suggestion: *"That might be challenging from an interpreter's perspective because while they are interpreting, they might not be able to provide an acknowledgment by clicking – so we might just have to do a sign and say yes or okay."* As a result, it might be useful to explore non-click-based acknowledgment mechanisms.

5.2.3 Supporting additional preset messages. Our participants were inspired by preset messages and made creative suggestions to support mixed-ability conversations. For example, $P07_{DHH}$ requested "please mute/unmute yourself" because, *"for the hard of hearing, some of them rely on voice. Also, if there's a lot of background noise, they can request to disable it."* Other participants discussed the utility of feedback messages during communication breakdowns, e.g., preset messages could inform users that their *"internet is bad"* or *"screen is frozen."* For internet issues at the interpreter's end, such feedback messages could be beneficial in alerting everyone and preventing information gaps. $P08_{DHH}$ added

> "If, for example, the internet is slow and someone is signing – you are in a very odd position. (laughs). Is there a way to message, "Sorry I'll join back" or something like that instead of just freezing their video." – $P08_{DHH}$

Apart from suggestions about different preset messages, one participant commented on the design of feedback notifications, reflecting on their use during the charades round. She noted that most participants would speak/sign "repeat" while guessing during charades instead of using the "please repeat yourself" preset message. She added,

> "Why is it easier to say it and have it interpreted than to just use that button? The point of the button was to reduce the labor of that action...you need a much larger or bolder notification that does not look like other notifications to ask you to repeat yourselves." – $P28_H$

This indicates the need to consider communicating different preset messages using different form factors. For example, a bolder notification might be helpful if the message is essential and requires urgent attention.

5.3 Flexibility with Automated Support to Reduce Labor

Study participants appreciated the flexibility offered by *Jod*. However, our participants realized they had to labor extra to align the platform with their communication needs. We now detail how participants envisioned complementing flexibility with automated support from the platform to enhance their experience.

5.3.1 Customized templates to reduce labor. We observed that the flexibility to reorganize the visual layout per specific needs enhanced participants' communication experience. Yet, a few participants found it challenging to navigate through this flexibility to create the best layout for themselves. For instance, $P28_H$ complained, *"The chat is one thing, the on-screen captions is another, and the interpreter's video is another. So currently, it's like... it's the labor of the hard-of-hearing participant, that they have to maneuver everything out – how do I see everything together? It should be on the part of the technologists to provide all these together easily."* Other participants also noted the labor required, especially for repetitive tasks. For instance, $P13_{DHH}$ mentioned, *"I shouldn't have to go and remove individual participants... we can have one option where I can click to show only deaf participants... we could just click that."* Automating such repeated tasks, including adding/removing participants based on hearing ability, would help to reduce user labor.

In addition, $P23_H$ suggested adding a feature to revert the rearranged layout to the default one automatically:

> "With respect to the resizing that we do in the starting – if there was an option to revert to the original layout, like default mode because what happened with me, accidentally, I think, I increased someone's screen, I mean, someone's window and the button for the window disappeared somewhere, and I just couldn't go back – the resizing, the white corner, yeah, it accidentally went to someplace. So it would be really nice to have that kind of an option." – $P23_H$

Participants suggested providing custom layout templates to reduce their initial effort in reorganizing the default layout. $P27_H$ said, *"With respect to the maneuvering, maybe you can have a bunch of templates instead of leaving everything to the user? They can pick, they don't have to do everything, but they can if they want to."* To decide on the custom templates, a hearing participant, $P28_H$, suggested basing it on focus group discussions and usage patterns of *Jod*:

> "...hearing all of these conversations, it would be so nice to have an optional template for the interpreter, an optional template for DHH participants that takes into account all these different perspectives and comes up with the best possible layout. For example, now we know that the interpreter needs to see the hard of hearing participants – there should be a template that reduces the labor of the interpreter. Similarly for DHH participants, if you constantly keep hearing that there is no point in seeing the hearing participants – then there could be a template that could cater to that." – $P28_H$

Such templates could replicate the most frequent layout of each participant group, and platforms could offer flexibility as an additional feature. In addition to providing custom video tile arrangements, it is essential to consider for each custom template the placement and size of widgets (such as chat windows, closed captions, transcripts, and reaction buttons) and how the system should react (e.g., update the size and position of other video tiles when a participant is customizing their visual layout). Participants had to manually resize those 'other' tiles to use their visual layout in *Jod* optimally. For example, $P19_{DHH}$ mentioned,

> "So I just had 4 participants (on my screen), and I resized one of the video tiles... the others should get automatically resized to fit that grid. I shouldn't have to manually increase the size of the others... it should automatically maximize others' video tiles to reduce the blank space on the grid." – $P19_{DHH}$

Manually resizing was time-consuming and redundant labor, creating challenges for participants to use their visual space optimally. Overall, we find that flexibility comes with added costs, which could be reduced by offering automated support and customized templates to participants based on their hearing abilities.

5.3.2 Automated support for grabbing attention. When designing *Jod*, we gave the shared screen a slightly larger tile size than the participants' video tiles. Still, we did not make it as prominent as current video conferencing platforms do. Though most participants navigated their way and reorganized it (Figure 5i - 5k), we observed a strong desire for automated ways to prioritize the shared screen. $P07_{DHH}$ commented, *"When someone else is sharing their screen, it doesn't pop up on my screen... It comes as a small window. That person had to inform me that he had shared the screen, and I zoomed in on that screen."* This caused information gaps and additional labor on the participant's end. Instead, participants wanted the shared screen to be larger than other tiles when it loaded to capture attention and then have added flexibility to resize if required.

In addition to the shared screen, some participants also expected smart behaviors from *Jod* to grab attention, especially while interrupting or asking questions. For example, $P21_{DHH}$ said,

> "If someone raises their hand – automatically they should come to the main grid. If they have a question or they have a doubt, then they can ask, so I know who is exactly asking the question or doubt." – $P21_{DHH}$

While common videoconferencing platforms such as Zoom offer these capabilities, participants complained about the constant video tile switching in these platforms, which makes it particularly challenging for DHH participants. A hearing participant, $P23_H$, suggested that *"the interpreter should stay static, and maybe the others – we could have some priority order. If there is a crosstalk kind of a thing – it shouldn't switch that much."* Therefore, it might be beneficial to design automated mechanisms to capture attention yet avoid unnecessary switching and enable the ability to set priorities for certain participants.

Participants also suggested providing automatic focus toward other widget elements, such as the chat window, in case of new messages. Particularly, DHH participants and the interpreter complained about missing new messages unless someone explicitly informed them; as $P11_I$ stated, *"when they were chatting, I did not realize that they had typed in the chat unless they told me."* This is perhaps because DHH participants and interpreters are constantly engaged in signing, making it hard for them to look away to stay updated with the chat. To mitigate this, $P22_H$, a hearing participant, suggested,

> "The DHH participants were doing the actions, but we were guessing in the chat. They were also pausing and looking in the chat. At some stage, these interactions have to grab your attention. The chat has to be bang in the middle. So it has to be like you know you overlay the text on the entire screen because when we were signing, they were looking at their screen… not looking at the corner. So overlay the text over the video – especially for games like charades, not always." – $P22_H$

In general, then, we find that *Jod*'s flexibility lets participants customize layouts to meet their preferences; there is an inherent need for intelligent support to achieve optimal layouts and visual spacing and to flag user attention.

5.4 Beyond Communication: Norms and Mismatched Expectations

We now discuss the varying cultural and communication norms among DHH and hearing groups we observed in our study and how that can result in mismatched expectations in mixed hearing communication contexts.

5.4.1 Cultural and communication norms in mixed hearing settings. During the user study, we discovered that some participants—specifically, DHH participants and interpreters—relied on various cultural practices to ensure efficient communication. For instance, to capture people's attention in a group conversation, $P11_I$ shared,

> "If the person is talking and they (DHH) want that person's attention, they always flash on the camera... they keep blocking [and unblocking] the camera, you notice something going black and white, they do that. However, if I specifically want [DHH person name]'s attention when there are 50 people (on the call), I would repeatedly do this [sign their name]. If they see me, they'll say [sign back my name]. So that's how we would get each other's attention." – $P11_I$

These workarounds make communication between DHH participants and interpreters more efficient. Similarly, using 'sign names'[3] when communicating with one another is common practice in deaf communication. However, participants (both DHH and interpreters) were not familiar with other participants' sign names.

To navigate this, we found that interpreters relied on alternatives, such as "fingerspelling their name" or saying "S hearing person or M hearing person" to provide contextual speaker information while interpreting. However, $P24_{DHH}$ talked about the time-consuming nature of such strategies: *"Say a person is asking a question, I don't know their sign name, and spelling their entire name is time-consuming... if we could have a number along with the names of the participants – like 1, 2, 3, 4, 5, and the name... I could just say number 1, like an ID, to save time."* Since hearing participants often do not have sign names, designing such suggested solutions could save time and enhance everyone's user experience.

Another key characteristic of deaf communication is the extensive use of visual cues, such as facial expressions and backchanneling gestures. As $P25_{DHH}$ mentioned, *"The deaf like to respond a lot while people are talking...They are very expressive, that's the deaf culture. So they might give a thumbs-up while someone is speaking."* This was not the case for hearing participants, who primarily relied on audio cues to establish conversational connections. A few DHH participants even wanted hearing participants to be more expressive, as $P30_{DHH}$ reveals:

[3] In deaf culture and sign language, a sign name (or a name sign) is a special sign used to identify a person (a name).

> "I want hearing people to use their expressions so that I can connect with their captions – what they are feeling and what they are trying to say. So that it can help me to understand better." – $P30_{DHH}$

5.4.2 Inherent lags and mismatched expectations. Despite *Jod*'s assistive capabilities, several communication gaps persist within mixed hearing group conversations. These gaps often resulted from the mismatched communication norms and expectations of different participant groups. For example, DHH participants' reliance on visual cues and expressions vs hearing participants' reliance on audio cues produced communication gaps: *"As a hearing person I rely on audio cues when someone starts speaking to me. I am not necessarily always looking at everyone's video tile. So, say, when a deaf person wants my attention or when they have started signing in a charades game, I don't realize it until unless the interpreter tells me that this person is speaking to me."* – $P28_H$. Moreover, several participants expressed frustration about the inherent gaps due to interpretation delays. Multiple delays were witnessed during the informal conversation round when hearing participants (including the researcher) or DHH participants told a joke. And the other participant group needed to wait for it to be interpreted. For example:

> "Anytime you (hearing user) make a joke, we (other hearing users) will always laugh first, whereas half of the participant group (DHH users) has not yet had the joke interpreted for them. There is a lag, which kinda puts the hearing participants on the upper hand of the power dynamic because we are almost able to have different levels of conversation that might not be inclusive." – $P28_H$

We find similar communication lag in conversation dynamics during the game of charades, especially when the movie name was guessed first by hearing participants. Moreover, during the FGDs, we observed that except for a few DHH participants who wanted hearing users to be more expressive, most DHH participants were content with communicating through the interpreter. Interestingly, a few hearing participants sought a deeper connection with DHH participants, extending beyond the interpreter's verbal communication. $P33_H$, a hearing participant, even expressed uncertainty about whether what she said was being understood by DHH participants:

> "I am not very confident if my words have been reached, if a deaf person has identified that 'oh, $P33_H$ is speaking', have they registered that? Do they feel that particular connection with me? Or not? Or they're just thinking it to be a part of the talk... or just a grand continuation of what was going on." – $P33_H$

Overall, we find that the communication norms used by DHH and hearing participants differed significantly, resulting in communication gaps, uncertainties, and misaligned expectations. Furthermore, these gaps were a barrier to deeper connections sought by some participants in mixed hearing videoconferencing.

6 DISCUSSION

This paper examined the usage of *Jod* by mixed hearing groups. We find that the flexibility and multiplicity that *Jod* offered enabled users to customize their interface to meet their personal preferences and continuously changing group communication context. Notifications tailored to mixed hearing ability conversations helped different participant groups to better communicate with each other. Observing participants use *Jod* showcases the need (1) for balance, to provide customization with automated support, (2) to overlay context-aware notifications with means for acknowledgment, and (3) to further explore features adhering to cultural practices. Below we discuss them in detail.

Flexibility vs System-Provided Defaults: Prior work has identified layout-related challenges faced by DHH users in videoconferencing platforms [13, 14, 26, 36], including the inability to keep other signers in view, difficulty in consuming information when the signers' video tile is small, and the inability to reduce visual clutter while consuming information from multiple sources. In our sessions, we observed participants actively customize *Jod*'s visual layout to create diverse layout arrangements, e.g., enlarging the interpreter's video tile, removing hearing participants, and rearranging DHH user tiles closer to each other. They updated their layout preferences multiple times as the study sessions progressed and the group communication context changed. Though such customization provides users control of their visual environment, it can increase user labor; many participants therefore wanted responsive layouts that would automatically fill up empty screen space or a way to transition back to the default layout. Some DHH participants felt the burden of individually removing/resizing each participant's video tile. We witnessed this constant tension between the need for complete flexibility versus the support they expected from the platform.

Design Recommendations: To reduce user labor and increase platform support, we recommend adding options for quick layout modifications (e.g., one-click actions to add/remove video tiles based on hearing ability, a back button to revert any layout changes, etc.), similar to hiding non-video participants option that Zoom offers [34]. Additionally, we recommend having optional video layout templates to choose from based on group context, substantiating Ang et al.'s suggestion for customizable layout templates [26]. These predefined templates need to be dynamic and should account for several attributes of the ongoing mixed hearing group conversation (e.g., group and individual accessibility needs, number of signers with active videos, and presence/absence of interpreters) to suggest layouts that are contextual and useful. Though the interpreter was available in our study, we observed DHH participants relying on captions and transcriptions for multiple use cases, such as to verify interpreters' voicing or when the interpreter's video got stuck due to low internet bandwidth. Thus, these layout templates must also accommodate appropriate placement for captions and transcriptions. Finally, future research should study this amalgamation of flexibility with templates, particularly automated ways to optimize screen real estate while supporting users in creating their preferred layout.

Context-aware Notifications: In *Jod*, participants used preset messages to influence others' behaviors. Prior work have studied

the designs of notification systems for DHH individuals to grab other's attention and communicate feedback [26, 30]. Our findings offer novel insights into various notifications' design and delivery mechanisms. We observed the disruptive nature of alerts that participants speculated in a prior work [20]. We implemented a click-to-dismiss interaction to ensure that notifications were dismissed only after the receiver had seen them. However, while signing, interpreters are usually slightly away from their videoconferencing setup to ensure their upper body, hands, and head are visible in the video. This made it difficult for them to dismiss notifications quickly, thus cluttering their visual layout with messages. Similarly, hearing participants felt that notifications were distracting and they felt interrupted. Furthermore, *Jod*'s design did not inform the sender if, how, and when the recipient of their message will respond to their request, leading them to send more notifications and further causing interruptions for the recipient.

Design Recommendations: In future iterations, researchers could explore making notifications less distracting. Further designs can be explored on how recipients could acknowledge them and how this information gets relayed to the sender. Our findings suggest that notifications are not equally urgent and may have an underlying priority based on the group communication context. For example, requesting active speakers to repeat what they said is more critical than asking passive participants to adjust their upper body. The priority of a message can be represented through visual design concepts like high-contrast colors and larger font sizes. The system could filter the repeats to not overwhelm the recipient with the same notification. Besides user notifications, we should have system notifications to support mixed hearing groups. For example, intermittently losing an interpreter's audio or video introduces information gaps in a mixed hearing ability conversation. Thus, similar to network connection notifications like poor connectivity, the interpreter's absence can be communicated at a system level. Similarly, informing users that they are out-of-frame can also be the system's responsibility. For example, using vision algorithms to detect if someone's upper body and hands are not visible or if they are sitting in poor lighting. On the recipient's end, there should be multiple ways to acknowledge the received message (e.g., "I will do it", "I cannot do it"). To enable users to interact with notifications while they are signing or interpreting, additional modalities (like swipe right/down gestures) can be studied further.

Integrating Deaf Cultural Norms: A sign name (or a name sign) is a unique sign used to identify a person, and it's an integral part of Deaf culture [21]. As the hearing participants and researchers did not have sign names, DHH participants and the interpreter shared their struggle in referring them using fingerspelling [5], leading to increased labor and further information gaps. Furthermore, not knowing each other's sign names could also lead to a disconnect with the DHH individuals on the call. A DHH participant suggested adding numeric identifiers for each hearing individual in the platform to ease the action of referring them.

Design Recommendations: To be more inclusive towards the LGBTQ community, videoconferencing platforms added an option for users to add and share their pronouns as part of their identity [33]. Similarly, videoconferencing platforms could allow adding sign names to user profiles through short self-recorded videos. Future explorations would need to distill how this integration works for hearing users because typically, sign names are given to hearing individuals by another person from the Deaf community [4, 23]. We believe this could be a small step towards introducing a rich part of Deaf culture to videoconferencing platforms. Moreover, the user profiles on the videoconferencing platform could also ask users to add their accessibility needs and preferred communication methods. As discussed previously, these details could help the system increase its awareness and provide contextual support.

6.1 Towards Conducting Inclusive Mixed Hearing Studies

With the emergence of research surrounding video-mediated communication within mixed hearing groups [12, 16, 17, 20, 29], several studies have outlined considerations for designing and facilitating inclusive studies [18, 26, 30]. Some recommendations proposed by studies employing participatory design to explore the future of videoconferencing include DHH representation within research team [19, 26], developing communication norms [18, 26], and use of appropriate phrasings [26]. Mack et al. discussed that academic papers often omit access accommodations and the labor put into making research methods accessible in accessibility studies [17]. Based on our experience, we now reflect and highlight several considerations and discuss implications for future research.

While conducting our study, we realized the "messiness" of our method and the importance of iterating over the study protocol. In the initial sessions (S2 and S3), we primarily relied on a researcher-facilitated informal conversation to encourage interactions among DHH and hearing participants. Though our participants were engaged, the conversations remained organized, researcher-driven, and lacked intermingling between the two groups. In the fourth study session, we introduced a *Charades* play round to improve this. In addition to facilitating cross-communication, *Charades* enhanced the overall experience and made the study much more enjoyable for our participants. Based on our learnings, we encourage accessibility researchers to be more flexible, open, and adaptable to quick iterations. Future studies could also explore novel, creative methods similar to *Charades* that could facilitate better intermingling and comfort and create a playful experience in mixed hearing studies. Such methods could particularly benefit studies involving system exploration, as they would facilitate closer to the natural, real-world interactions among both DHH and hearing groups.

Prior studies in HCI and Accessibility have also highlighted the need to consider the accessibility of the full-method pipeline, from selecting a research method to analyzing the data [18]. In our study, the in-person sessions were conducted in the workplace of our DHH participants. We opted to conduct focus group discussions (FGDs) instead of semi-structured interviews for two main reasons: (1) to encourage participants, both DHH and hearing, to express their individual and collective viewpoints and engage in group discussions, and (2) to mitigate the burden of interpretation and minimize transcription expenses. We observed a clear distinction between the remote and in-person FGDs. The FGDs conducted in person, where participants and researchers were in close

physical proximity to each other, proved to be more engaging and interactive, as opposed to the FGDs conducted over Zoom. However, given the structured nature of remote FGDs and the advanced capabilities of Zoom, the transcription was straightforward, unlike that for the in-person FGDs, which posed difficulties due to lack of established communication norms. E.g., speaker identification posed a significant challenge during the transcription of in-person FGDs, as the interpreter failed to indicate the corresponding DHH participant while interpreting, leading to information gaps in our audio recordings. To address this issue, we relied on our handwritten notes to map the participant quotes with the respective speakers. We argue in-depth discussions are necessary to establish effective communication protocols, specifically around *when*, *how*, and *where* to lead focus groups in mixed hearing settings.

Lastly, as most of our hearing participants had limited experience interacting with DHH individuals, they were unsure of how to communicate with the DHH participants through the interpreter. For example, one hearing participant asked whether to direct her gaze toward the DHH signer or the interpreter. In alignment with prior recommendations [26], we encourage establishing clear communication protocols for both DHH and hearing participants.

6.2 Limitations

The *Jod* system and study design have several limitations. First, our findings focused on medium-sized mixed hearing groups and may not generalize to large group settings. Second, some of *Jod*'s design choices may not scale well to large groups of people. For instance, the participants anticipated the effort it would take to manually resize and remove/add video tiles if more people were on the call. Third, a critical use case for videoconferencing platforms is to present information through screen sharing, and the type of shared content varies. Though we explored a screen sharing experience during the study session, it was limited since DHH users did not experience the complexities that arise with sharing multimedia presentations. Fourth, as the DHH participants and interpreters were recruited from the same partner organization for some sessions, our observations and findings could have been influenced by the comfort of participants already knowing each other. Each session also had the same ratio of DHH to the interpreter to hearing participants, which may or may not reflect a real-world group conversation. Finally, though our study design was motivated by real-world situations, the limited time people spent on *Jod* was insufficient to recreate diverse group contexts that could have led to communication challenges. For example, though the DHH participants favored *Jod*'s *accessibility indicators* feature, given they might have known each other would have made the feature less valuable during the study. Our work can inform future research on conducting large-scale longitudinal studies and exploring different group compositions across session activities.

7 CONCLUSION

We designed and built a videoconferencing platform for mixed hearing ability conversations between DHH signers, interpreters, and hearing users. We revealed how *Jod*'s features interact with each other by simulating real-world conversations in user studies. Our study participants desired a balance between flexibility and system-provided automated defaults, and raised a need for acknowledgments and prioritization of received messages based on the group communication context. Based on our findings, we identified design guidelines for future videoconferencing platforms that could enhance virtual communication in mixed hearing groups.

ACKNOWLEDGMENTS

We thank Shivang Chopra, Aashaka Desai, Jon E. Froehlich, Shreeshail Hingane, Nidhi Kulkarni, Emma J. McDonnell, Saumay Pushp, and Harsh Vijay for their contributions and feedback in this research. We would also like to thank our partner organizations–AI4Bharat at IIT Madras, the National Institute of Speech and Hearing (NISH), and the WinVinaya Foundation–for their support and feedback. This work was supported in part by Microsoft Research India, the University of Washington Center for Research and Education on Accessible Technology and Experiences (CREATE), and the United States National Science Foundation through award IIS-1702751.

REFERENCES

[1] Rahaf Alharbi, John Tang, and Karl Henderson. 2023. Accessibility Barriers, Conflicts, and Repairs: Understanding the Experience of Professionals with Disabilities in Hybrid Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. https://doi.org/10.1145/3544548.3581541

[2] Sara Atske. 2021. The Internet and the Pandemic. https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/

[3] Steve Bennett. 2023. Video Conferencing Statistics 2022 - Everything You Need to Know. https://webinarcare.com/best-video-conferencing-software/video-conferencing-statistics/

[4] Jamie Berke. 2023. Name Signs in the Deaf Community. https://www.verywellhealth.com/using-name-signs-for-personal-names-1048725 Section: Verywell.

[5] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 16–31. https://doi.org/10.1145/3308561.3353774

[6] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Place: United Kingdom Publisher: Hodder Arnold.

[7] Indian Sign Language Research and Training Center. 2023. History. https://islrtc.nic.in/history-0

[8] Stephanie Chan. 2021. Usage of Mobile Video Conferencing Apps Including Zoom Grew 150% in the First Half of 2021. https://sensortower.com/blog/video-conferencing-apps-mau-growth

[9] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J.D. Bolter, and M. Gandy. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (Oct. 2005), 18–26. https://doi.org/10.1109/MPRV.2005.93 Conference Name: IEEE Pervasive Computing.

[10] Dhruv Jain, Audrey Desjardins, Leah Findlater, and Jon E. Froehlich. 2019. Autoethnography of a Hard of Hearing Traveler. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Pittsburgh PA USA, 236–248. https://doi.org/10.1145/3308561.3353800

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

[12] Yeon Soo Kim, Hyeonjeong Im, Sunok Lee, Haena Cho, and Sangsu Lee. 2023. "We Speak Visually": User-Generated Icons for Better Video-Mediated Mixed-Group Communications Between Deaf and Hearing Participants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3581151

[13] Yeon Soo Kim, Sunok Lee, and Sangsu Lee. 2022. A Participatory Design Approach to Explore Design Directions for Enhancing Videoconferencing Experience for Non-signing Deaf and Hard of Hearing Users. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS*

'22). Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3517428.3550375

[14] Raja S. Kushalnagar and Christian Vogler. 2020. Teleconference Accessibility and Guidelines for Deaf and Hard of Hearing Users. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3373625.3417299

[15] Harlan Lane. 2005. Ethnicity, Ethics, and the Deaf-World. *The Journal of Deaf Studies and Deaf Education* 10, 3 (July 2005), 291–310. https://doi.org/10.1093/deafed/eni030

[16] Kelly Mack, Danielle Bragg, Meredith Ringel Morris, Maarten W. Bos, Isabelle Albi, and Andrés Monroy-Hernández. 2020. Social App Accessibility for Deaf Signers. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–31. https://doi.org/10.1145/3415196

[17] 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 (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3441852.3471199

[18] Kelly Mack, Emma McDonnell, Venkatesh Potluri, Maggie Xu, Jailyn Zabala, Jeffrey Bigham, Jennifer Mankoff, and Cynthia Bennett. 2022. Anticipate and Adjust: Cultivating Access in Human-Centered Methods. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–18. https://doi.org/10.1145/3491102.3501882

[19] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 434:1–434:25. https://doi.org/10.1145/3479578

[20] Emma J McDonnell, Soo Hyun Moon, Lucy Jiang, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2023. "Easier or Harder, Depending on Who the Hearing Person Is": Codesigning Videoconferencing Tools for Small Groups with Mixed Hearing Status. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. https://doi.org/10.1145/3544548.3580809

[21] Kathryn P. Meadow. 1977. Name Signs as Identity Symbols in the Deaf Community. *Sign Language Studies* 16 (1977), 237–246. https://www.jstor.org/stable/26203239 Publisher: Gallaudet University Press.

[22] 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 (ACM-SE 45)*. Association for Computing Machinery, New York, NY, USA, 501–506. https://doi.org/10.1145/1233341.1233431

[23] Anna Mindess. 1990. What Name Signs Can Tell Us About Deaf Culture. *Sign Language Studies* 66 (1990), 1–23. https://www.jstor.org/stable/26204041 Publisher: Gallaudet University Press.

[24] World Federation of the Deaf. 2023. Our Work. https://wfdeaf.org/our-work/

[25] World Health Organization. 2023. Deafness and hearing loss. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[26] Jazz Rui Xia Ang, Ping Liu, Emma McDonnell, and Sarah Coppola. 2022. "In this online environment, we're limited": Exploring Inclusive Video Conferencing Design for Signers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3491102.3517488

[27] Chris Sano. 2022. Introducing Sign Language View for Teams Meetings. https://techcommunity.microsoft.com/t5/microsoft-teams-blog/introducing-sign-language-view-for-teams-meetings/ba-p/3671257 Section: Microsoft Teams Blog.

[28] Matthew Seita, Khaled Albusays, Sushant Kafle, Michael Stinson, and Matt Huenerfauth. 2018. Behavioral Changes in Speakers who are Automatically Captioned in Meetings with Deaf or Hard-of-Hearing Peers. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 68–80. https://doi.org/10.1145/3234695.3236355

[29] 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 (W4A '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3430263.3452430

[30] Matthew Seita, Sooyeon Lee, Sarah Andrew, Kristen Shinohara, and Matt Huenerfauth. 2022. Remotely Co-Designing Features for Communication Applications using Automatic Captioning with Deaf and Hearing Pairs. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3491102.3501843

[31] Advaith Sridhar, Roshni Poddar, Mohit Jain, and Pratyush Kumar. 2023. Challenges Faced by the Employed Indian DHH Community. In *Proceedings of the 19th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT)*. Springer.

[32] Microsoft Support. 2023. Accessibility tools for Microsoft Teams. https://support.microsoft.com/en-us/office/accessibility-tools-for-microsoft-teams-2d4009e7-1300-4766-87e8-7a217496c3d5

[33] Zoom Support. 2023. Adding and Sharing your Pronouns. https://support.zoom.us/hc/en-us/articles/4402698027533-Adding-and-sharing-your-pronouns

[34] Zoom Support. 2023. Adjusting Your Video Layout During a Virtual Meeting. https://support.zoom.us/hc/en-us/articles/201362323-Adjusting-your-video-layout-during-a-virtual-meeting

[35] Zoom Support. 2023. Enabling Sign Language Interpretation View. https://support.zoom.us/hc/en-us/articles/9513103461005-Enabling-Sign-Language-interpretation-view

[36] John Tang. 2021. Understanding the Telework Experience of People with Disabilities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 30:1–30:27. https://doi.org/10.1145/3449104

[37] Christian Vogler, Paula Tucker, and Norman Williams. 2013. Mixed Local and Remote Participation in Teleconferences from a Deaf and Hard of Hearing Perspective. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2513383.2517035

[38] Zoom. 2023. Zoom is for Everyone. https://explore.zoom.us/en/accessibility/

A APPENDIX

A.1 Task-based Exploration

We asked participants to explore system features through the following task prompts:

(1) Make a participant's video tile larger or smaller.
(2) Move another participant's video tile, anywhere on the screen.
(3) Remove a participant's video. Add it back by clicking on the Add button in the participants' list, on the right side.
(4) You can also lock a participant's video. If you lock someone's tile you will not be able to move them around or resize it (Optional).
(5) Can you ask me (the researcher) to turn ON my lights?
(6) Can you try requesting Participant X to speak slowly?
(7) Try clicking on notifications and see what happens.
(8) Try gestures like raising your hand, clapping, ok, and thumbs-up. You can also click on the icons on the top right to communicate these reactions.
(9) Can you try sending a message on the chat?
(10) Can you try resizing and moving the closed captions box?

A.2 Jod's Features and Addressed Accessibility Barriers

***Jod*'s Features**	**Addressed Accessibility Barriers**
Customizable Visual Layout. Allows users to completely customize their visual layout by resizing, rearranging, and removing video tiles. They can reposition and resize the captions box too.	Speechreading is challenging due to lack of eye contact and because speaker's gestures and facial expressions can get inaccessible [13]. Suggested Design Direction: Ability to zoom in on the speaker and remove passive participants [13]. DHH individuals need to rely on captions when speechreading becomes difficult [13]. Suggested Design Direction: Keep captions near the speaker [13]. Videoconferencing platforms offer limited support to customize visual elements but DHH users' needs to rearrange and resize the elements on their screen are unique [14]. Maneuvering multiple sources of information during video conferencing e.g. slides or screen share, signing interpreter, speaker video [22]. Suggested Design Direction: Semi-transparent video which can be overlaid over a shared screen [22].
Preset Feedback Messages. Participants can request others to look at them, keep their upper body visible, sit in well-lit areas, speak slower, use easier language, and repeat themselves. **Accessibility Indicators.** Help guage accommodations and preferences in mixed hearing settings.	Poor lighting and busy visual backgrounds can make it hard for DHH users to speechread or follow signing [14, 36]. Bad camera adjustments may lead to less eye contact which can be perceived as a lack of engagement [14] Hearing users' behaviors may negatively affect DHH users' conversation experience (e.g. speaking at a low volume or speaking too fast) [11, 28, 29]. Suggested Design Direction: Notification systems to influence hearing users' behavior [30].
Active Signer Identification. Focus on DHH individuals who are signing instead of interpreters who are voicing for them.	Difficulty in speaker identification [13, 14, 26, 37] and DHH signer indentification through the voice of the interpreter [37]. Suggested Design Direction: Dedicated location for essential elements such as speaker and captions [13].
Enhanced Transcriptions. Ensure all users have a shared conversational context through ASR outputs of past conversations, emojis, and start-stopped signing tags.	If a speaker speaks too fast, captions may disappear faster than someone's reading speed [14]. DHH users may miss content and lose conversation context if they look away from their screen and miss reading captions [14].

Table 2: Summary of Related Work: *Jod*'s Features and Addressed Accessibility Barriers