Towards Inclusive Video Commenting: Introducing Signmaku for the Deaf and Hard-of-Hearing

Si Chen
sic3@illinois.edu
School of Information Sciences,
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Haocong Cheng
haocong2@illinois.edu
School of Information Sciences,
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Jason Situ
junsitu2@illinois.edu
Computer Science, University of
Illinois Urbana-Champaign
Urbana, Illinois, USA

Desirée Kirst
deskirst@gmail.com
Gallaudet University
Washington, District of Columbia
USA

Suzy Su
xiaoyus4@illinois.edu
School of Information Sciences,
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Saumya Malhotra
saumyam2@illinois.edu
College of Liberal Arts & Sciences,
University of Illinois
Urbana-Champaign
Urbana, Illinois, USA

Lawrence Angrave
angrave@illinois.edu
Computer Science, University of
Illinois Urbana-Champaign
Urbana, Illinois, USA

Qi Wang
qi.wang@gallaudet.edu
Gallaudet University
Washington, District of Columbia
USA

Yun Huang
yunhuang@illinois.edu
School of Information Sciences,
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

ABSTRACT

Previous research underscored the potential of danmaku–a text-based commenting feature on videos–in engaging hearing audiences. Yet, for many Deaf and hard-of-hearing (DHH) individuals, American Sign Language (ASL) takes precedence over English. To improve inclusivity, we introduce "Signmaku," a new commenting mechanism that uses ASL, serving as a sign language counterpart to danmaku. Through a need-finding study (N=12) and a within-subject experiment (N=20), we evaluated three design styles: real human faces, cartoon-like figures, and robotic representations. The results showed that cartoon-like signmaku not only entertained but also encouraged participants to create and share ASL comments, with fewer privacy concerns compared to the other designs. Conversely, the robotic representations faced challenges in accurately depicting hand movements and facial expressions, resulting in higher cognitive demands on users. Signmaku featuring real human faces elicited the lowest cognitive load and was the most comprehensible among all three types. Our findings offered novel design implications for leveraging generative AI to create signmaku comments, enriching co-learning experiences for DHH individuals.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in accessibility**; **Web-based interaction**; • **Social and professional topics** → **People with disabilities**; • **Applied computing** → **Interactive learning environments**.

KEYWORDS

DHH, Social Interactions, Danmaku, Signmaku

1 INTRODUCTION

"Danmaku" is a video-commenting feature where users overlay their *text* comments directly onto videos. Its effect on learning lies in fostering a more interactive and engaging learning environment by allowing learners to view comments from their peers, encouraging active participation and social connection during the video-based learning process [73]. However, text is less accessible for Deaf and hard-of-hearing (DHH) users, as many of them prefer signed language over text, in our context, American Sign Language (ASL) over English [45]. More than 500,000 people in the United States use American Sign Language (ASL) as their primary mode of communication [49].

Existing research focuses on improving the accessibility of video-based learning for DHH populations via generating closed captions
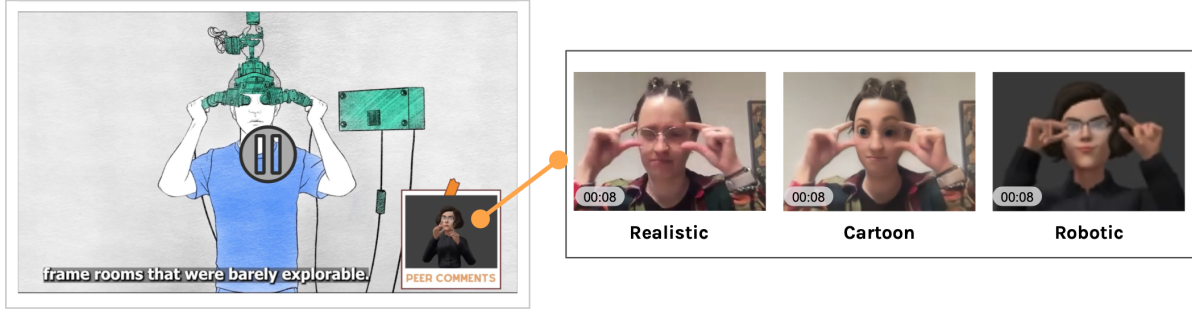
Figure 1: Three Styles of signmaku designs explored in our study. *realistic* **(unfiltered),** *cartoon* **(filtered face with real torso and hands), and** *robotic* **(filtered face, torso, and hands) leverages AI technology to filter signed video clips, considering different styles of privacy preservation.** *Cartoon* **filter's background was selected to contrast an individual's skin color for more visible hand movements using VToonify [79].** *Robotic* **filter's appearance was customized by the DHH individual in DeepMotion.**

and text versions of the audio content. Despite accounting for reading proficiency, the differences in background knowledge and information processing strategies make text-based information less accessible for DHH than hearing [46]. On the other hand, sign language is a form of visual communication that employs gestures, facial expressions, and body language. It is a first-class language with the same level of complexity as spoken English, but it uses a different sentence structure [57]. Little is known about how DHH users perceive and share danmaku comments on videos in ASL rather than text.

Thus, in this paper, we introduce a new ASL-based commenting feature called **Signmaku**, which is like a sign language version of danmaku. Unlike text-based comments, signmaku offers ASL-based comments that contain visual information with facial expressions and hand movements, as shown in Fig. 1. The unique challenge of sign language, compared to text, is its heightened privacy concerns of revealing individuals' faces. Therefore, we designed three visual styles by applying different generative AI techniques: 1) *Realistic* (unfiltered or original ASL recording with users' real faces); 2) *Cartoon* (filtered faces with real torso and hands using VToonify [79]); and 3) *Robotic* (filtered faces, torsos, and hands using DeepMotion[1]) that address different levels of privacy concerns. We aim to explore how it can improve the inclusivity of video-based learning for DHH. While our current signmaku design could not fully replicate the experience with text-based danmaku, we expect signmaku to enrich the learning experience of DHH students by allowing them to make and view comments for and from their peers in sign language, fostering a pseudo-synchronous co-learning environment.

In order to design and explore the effect of signmaku on DHH's video-based learning, we completed a two-phase study. In Phase I, we conducted a need-finding study with 12 DHH participants, which allowed us to identify several key requirements of the signmaku feature and collect high-quality signmaku comments that were used in the next phase. In Phase II, we generated differently styled signmakus and designed a within-subject experiment with 20 DHH participants, during which they viewed content with the

collected signmakus while watching a video. Participants were also asked to create and share their own signmaku comments and choose their preferred style for sharing. The design aim for the two filtered styles is to provide options for learners to remain anonymous if they so choose, and thereby eliciting their willingness to make and share their own ASL comments. The experimental study addresses:

- **RQ1**: *How does viewing three styles of signmaku impact DHH learners' video-based learning?*
- **RQ2**: *How do DHH learners perceive the creation and sharing of the three styled signmaku?*

Our research makes significant contributions: 1) we propose a novel interaction design, Signmaku, which not only improves the inclusivity of video-based learning for DHH but also enhances their sense of social connectedness; 2) we provide empirical evidence of the effectiveness of cartoon-styled signmaku on learners' improved engagement and their reduced efforts of sharing comments with their peers, with fewer privacy concerns than the other two signmaku styles; and 3) we reinforce the importance of reproducing movements of certain facial areas for delivering meaning and emotions of DHH users with sign language, which is often overlooked by generative AI technologies. Our study primarily focuses on interaction design and understanding user behavior. We do not claim contributions to technical advancements.

2 RELATED WORK

2.1 Challenges Faced by DHH Learners in Video-based Learning

Video-based Learning is a remote learning approach that relies on live or prerecorded video to teach new skills and knowledge. Many video-based learning platforms have attracted hundreds of thousands of students, such as MOOC, and YouTube. They allow people from around the globe to participate in courses via videos. For DHH learners, video-based learning is not yet fully accessible. The major reason is because online videos in general lack good-quality captions, if any. Some research shows DHH people used filtering tools to find videos with captions for them to watch [40].

[1]https://www.deepmotion.com

Even when captions exist, many are erroneous as they are auto generated from automatic speech recognition algorithms [13]. It is not unusual for these algorithms to miss some words due to noise and to recognize the incorrect word when generating captions [12, 35]. Human efforts are needed to correct those errors; however, video content creators were frequently unwilling or unable to do so [40, 47]. Beyond the accuracy of captions, the appearance of caption is also important to learners' understanding [11]. Low-quality captions for online learning videos created great challenges for DHH learners to understand the learning content when studying remotely from school through video lectures during the COVID-19 period, which aggravated mental health issues [5]. Furthermore, the United Nations acknowledged that COVID-19 exacerbates inequalities between people with and without disabilities, including access to education and inclusion in the community [51].

Social interaction encouraged within a video-based learning platform can help students stay engaged and enhance their own learning [19, 60, 78]. Social engagement with peers is paramount for all learners, but it holds even greater significance for DHH learners. This importance is magnified by the fact that a considerable number of DHH individuals have experienced or perceived exclusion within mainstream education (Mainstream is the practice of placing students with special education needs in a general education classroom.) Research indicates that deaf children can have social difficulties compared with their hearing peers [23, 66]. The video-based learning experience during the COVID-19 pandemic has worsened social exclusion among deaf students, especially with the disruption of daily interactions with other people and inadequate sign language interpreters [5, 69]. Research suggested parents of DHH children look for suitable online educational programs, and opportunities for their children to socialize with their DHH peers [36]. Given current technological advancements, there is a great potential to conduct technology research, targeting inclusion and proper education for DHH students by enabling positive social interactions and collaboration with their peers [9].

Social media features have proven to facilitate interaction not just among hearing learners [19, 62], but for those who are DHH. A survey on predominant social media usage patterns of DHH teenagers in Saudi Arabia revealed that a majority engaged with social media daily [50]. Additionally, another survey highlighted how social media empowers DHH students to foster connections and collaborations on information discovery, content creation, and sharing [4]. The positive impact of social media on academic performance was further emphasized by students, who reported heightened interaction, support, and feedback [70]. Notably, researchers found that individuals within the DHH community predominantly engage in sharing content through written English, despite a preference for sign language. This inclination can be attributed to challenges such as difficulties in captioning signed videos and obstacles related to recording, such as device limitations, lighting, storage space, etc. [17, 45]. In summary, these social media features are beneficial for social interaction but have limitations due to the lack of support for the native language of the DHH community - sign language. In our research, we provide social interactions for DHH learners by supporting the use of sign language.

2.2 Danmaku for Social Interaction and Knowledge Sharing in Video-based Learning

Danmaku is a commentary design for online videos that originated in Japan and is now popular on Asian video websites, e.g., Bilibili.com [73, 74, 80]. In contrast to traditional video forums, where comments are asynchronously displayed below the video, danmaku comments overlay the video screen. These comments align with specific scenes chosen by online users and scroll from right to left across the video screen. Users can **anonymously send and view** danmaku comments in a synchronous way while watching videos. Wu et al. compared danmaku comments with traditional video forums and found the danmaku design significantly promoted user participation and social interactions [73]. The work also found danmaku and forum complement each other in the realm of knowledge sharing. Danmaku comments primarily facilitate explicit (know-what) knowledge sharing, while forum comments tend to showcase more tacit (know-how) knowledge sharing. He et al. found that viewers were more emotional and expressive in danmaku comments [33].

Research on online learning using danmaku, found that watching videos with danmaku and providing danmaku is a way for many users to entertain themselves and increase their sense of belonging among other audiences [21, 43, 44, 77]. Danmaku allows students to collaborate with their tutors and peers to create learning content by adding comments that help explain learning materials [77]. For teachers, danmaku can bridge the gap between teachers and students, and reduce learners' feelings of loneliness [41]. Conversely, some research pointed out that danmaku is distracting due to its rich abundance of information content and uneven content quality [21]. Furthermore, using danmaku as learner-sourced material raises the possibility that the content created will be ineffective, inappropriate, or incorrect [1].

Meanwhile, research investigating the use of danmaku in video learning with DHH students remains sparse, particularly due to the fact that the current danmaku is text-based which is not fully accessible to DHH people as their native languages are sign and gesture-based languages. Despite accounting for reading proficiency, the differences in background knowledge and information processing strategies make text-based information less accessible for DHH people than hearing people [46]. Our study addresses this gap by accommodating sign language preferences in redesigning text-based danmaku into signmaku (sign language danmaku). Though signmaku is a new concept, a few known or generally accepted facts support the possibilities for DHH students to benefit from synchronous ASL videos along with the original video content. Firstly, many TV shows have a sign language interpreter next to the video content and some video services have a closed ASL interpreting [34, 37]. Secondly, instructors believe DHH students have relatively better visual skills than hearing [58] and DHH students frequently switch their visual attention between the board, slides, instructors, and interpreters [58].

2.3 Anonymization of Sign Language Videos for Privacy Concerns

The ability to communicate anonymously using one's preferred language holds vital importance across diverse social, professional,

and societal settings. As outlined earlier, anonymous comment sharing is a crucial aspect of danmaku interaction and we, too, would need to provide anonymization features in signmaku in our study. While achieving anonymous written communication online is relatively uncomplicated for users of written languages, such options have not been readily available to users of sign languages. For video-based communication, individuals who use spoken languages can conceal their faces on online video-sharing platforms; however, sign-language users cannot do so due to the crucial linguistic information conveyed through facial expressions [8, 15, 25, 39]. Respondents expressed willingness to view videos created by individuals with their faces disguised as long as there was an ethical purpose [39].

Sign language is a form of visual communication that employs gestures, facial expressions, and body language. It is a complete first-class language with the same level of complexity as spoken English, but it does not use the same sentence structure [57]. Anonymizing sign language videos poses challenges as it is not feasible to directly cover their face. Recent efforts have emerged to assist sign language users in automatically concealing their faces in videos while retaining vital facial expressions; however, this task remains challenging and none of the current approaches offer a sufficiently effective solution. Lee et al. presented several prototypes to DHH ASL signers and revealed users' perception of key trade-offs among three dimensions–understandability, naturalness of appearance, and degree of anonymity protection [39]. Clearly representing facial expressions including eyebrow and mouth movements plays a crucial role in conveying content accurately in sign language. A limitation of their study is that they did not study sharers' perspective. Additionally, emerging research focused on sign language recognition [55, 57], animation-based generation [48, 61], and machine learning-driven translation systems [24, 67], demonstrating a promising trajectory for the future. To advance the development of sign language processing, the active participation of the Deaf community is imperative across all stages. This involvement is crucial for designing systems that align with user requirements, ensuring usability, and promoting technology adoption [14, 54, 71].

3 METHOD

Inspired by the idea of text-based danmaku [74] and its effect of creating a pseudo-synchronous learning experience to engage users [73], we introduce **Signmaku**, which displays peer learners' comments in sign language alongside the video, fostering a pseudo-synchronous co-learning experience. Note that in this study, we used American Sign Language (ASL) as the sign language, and all signmakus were created in ASL.

We conducted a two-phased study to design and evaluate the online learning experience with signmaku. Phase I, described in Section 3.1, is a two-round formative design study (N=12) to explore the design parameters of signmaku. Round 1 focused on (1) understanding willingness to view (2) filtering styles to maintain anonymity if needed (3) collect ASL data signmakus to be used in round 2. Round 2 focused on fine-tuning the signmaku design. Phase II, detailed in Section **??**, is an evaluation study (N=20) to evaluate how DHH students view and create signmaku in online learning, with a focus on three styles.

Positionality of Research Team The research team consisted of members with diverse backgrounds, including hearing and Deaf individuals: one Deaf research assistant, one hearing research assistant who is a Ph.D. student in linguistics and also an ASL interpreter, one hearing research assistant that is a beginner in ASL. All study sessions were supervised by a hearing professor with 30 years of experience in teaching DHH students at the collegiate level. The data analysis primarily involved four hearing coauthors who had prior experience in both quantitative and qualitative analysis. They regularly held discussions with Deaf research assistants and the hearing professor to refine the codebook and ensure comprehensive analysis.

3.1 Phase I: Designing Signmaku–Video Comments in Sign Language

To delve into the design of video-based learning experience with signmaku, we conducted a university Institutional Review Board (IRB)-approved pilot study with 12 DHH students (8 males, 4 females), including 8 d/Deaf students and 4 hard-of-hearing students, all from a US higher education school for DHH students. Throughout the study, we engaged in interviews and follow-up email conversations with these students to identify the essential design considerations necessary for the concepts to be effective.

The formative study consisted of two rounds. Round 1 was used to understand the viewing and sharing needs of signmaku and collect data to be presented in round 2. Round 2 was used to iteratively inform the actual design of signamku (Figure 1 shows the final design) via interviews to reveal key design parameters. All 12 participants were recruited in round 1, and several weeks later, round 2 included six of these participants. In Round 1, each participant was compensated $40 for their involvement in the 2-hour session. No compensation was offered for follow-up participation in Round 2. We analyzed the interview transcript and discussed the findings as well as improvements to the signmaku design at the end of each round.

3.1.1 Round 1: Needfinding of Signmaku for Social Connectedness. In the initial session, the participants were presented with a 15-minute video about Augmented Reality (AR), accompanied by error-free open captions. Following the video viewing, the participants were requested to provide their comments regarding the video content, using either ASL (as signmaku), English, or a combination of both. Additionally, they were asked if they were willing to share their comments for future learners, which would be utilized as supplementary learning material. We inquired about the factors that influenced their willingness to share. Participants were given the opportunity to express their preferences regarding viewing others' comments in either ASL or English as well as the manner in which they would like to access such comments during video learning. On average, each participant made 9 comments, with two-thirds in ASL and one-third in English. ASL-based video comments lasted between 5 and 45 seconds, with longer ASL-based comments including multiple sentences.

Our round 1 findings indicated that incorporating comments, both as viewers and as providers, has the potential to foster increased social connectedness among students during online sessions by addressing feelings of loneliness that may arise. Notably, it

was observed that ASL-based comments were generally perceived to be more comprehensible and engendered a stronger sense of connection among DHH participants. Especially as a minority group, DHH community members used their own language to fully express themselves and feel included in the learning process. However, some participants expressed concerns about privacy, because providing ASL-based comments entailed revealing their faces, which can hold grammatical meanings in ASL and cannot be directly removed. Consequently, some participants suggested implementing filters to mask/swap their faces, while others exhibited a preference for more comprehensive masking, such as swapping their entire torso and face. In summary, participants showed a positive inclination towards participating in such commenting activities, and anonymity was essential for some participants.

3.1.2 Round 2: Specifying Design Features of Signmaku. Based on the literature review and the results from the first round, we synthesized and learned that signmaku could serve as a valuable paradigm for promoting social connectedness and knowledge sharing. Subsequently, a second round was conducted 1-2 months after the first round to identify key design parameters of signmaku, particularly its display during video consumption. For this round, all 12 participants from the previous interview were re-invited and 6 chose to participate. We selected 15 signmakus from Round 1 and embedded them into the lower right corner of the same 15-minute video about AR. This gave participants on average one signmaku per minute. We then asked our participants to watch this video with signmakus.

The visual layout design underwent several rounds of improvement process based on feedback from each of the six participants. Once participants viewed the video with signmaku, we asked for suggestions to enhance its effectiveness for better comprehension and learning. The draft version is similar to prior literature [20]. In that work, Chen et al. found that with the *embedded* layout, learners altered their attention more frequently between two comments and remembered more surface-level comments. Therefore, we selected the *embedded* layout for signmaku.

The participants' feedback was carefully incorporated into the design, and subsequent emails were sent to them to confirm their willingness to provide additional inputs. The original layout for video-watching underwent multiple refinements to arrive at the final layout used in the subsequent formal study. The final design of signmaku is shown in Figure 1. Below, we provide an outline of the key parameters that were manually adjusted during the improvement process, along with the reasons behind those adjustments. Round 2 was dedicated to adjusting the parameters for our second study; however, these parameters might not be universally applicable to signmaku in different video contexts. Future research should explore methods for automatically completing the adjustment of design parameters.

- **Duration**: Participants were required to alternate between reading signmaku and captions while also attending to the video's visual content. Consequently, it was crucial for the signmaku duration to strike a balance. If it was too lengthy, it would interrupt the participants' **comprehension flow and attention span**, but if it was too short and limited to a single sign, it might ineffectively convey the intended message. Thus, we reduced the recommended duration from 30 to 10 seconds. This allowed DHH individuals to complete signing a full sentence in ASL. It is important to note that this parameter was tailored and refined based on the pace of the selected video content. It may require further exploration and adaptation when applied to different video contexts with varying tempos.

- **Dimension**: Participants found it crucial for the signmaku to be sufficiently large, ensuring clear visibility of the signer's **facial expressions, fingers, and elbows** for the participants to understand the language. Simultaneously, it is essential to avoid overlap with captions and video content. To address this, we positioned the signmaku in the lower right corner, aligning it with the height of the text captions. This placement allowed learners to seamlessly **switch their visual focus** between the two sources with ease. In the initial setup, an additional panel for text-based comments was incorporated alongside the video panel. This panel automatically scrolled during video playback while showing text-based peer comments. However, we later removed this panel from the formal study to provide a larger video player that could ensure better visibility of signmaku and captions.

- **Placement in relation to the video content**: Participants expressed a desire for the signmakus to accurately correspond to the video's content, particularly when referencing specific elements like equations. However, they also suggested that the pop-up does not necessarily need to appear directly next to the referenced content. They felt that the unpredictability of where the signmaku might appear could add to the confusion and pressure, leading them to prefer a **consistent presentation**, allowing them to know clearly where to expect the signmaku.

- **Categories**: We found that some signmakus from round 1 focused more towards complaints of video content and less towards the knowledge of the video (e.g., "When can this video end, I am tired."). For knowledge-sharing purposes, we manually selected signmakus that commented on the actual video content instead of emotion-only content, in a similar approach to previous danmaku research [74] (e.g., "The prototype would not fly nowadays. But I will admit it does look pretty cool.")

- **Two Filters vs Unfiltered**: Following participants' feedback, we applied filters on signmakus with advanced face-swap technology, similar to a prior study [39] that used several prototypes to automatically disguise the face in ASL videos while preserving essential facial expressions and natural human appearance. We explored multiple technologies of face filtering and eventually decided to examine in-depth the two most promising filters. The first filter was the *cartoon* filter using VToonify [76], a portrait style transfer tool, which retains facial movements while giving the face a cartoon-like appearance and providing a basic level of privacy. The original human hands are preserved in this filter. The second filter was the *robotic* filter using DeepMotion, an AI-based motion capture and body tracking tool, which masks the entire body and utilizes artificial hands and faces to create a robotic appearance. This filter provides an increased level of privacy than the *cartoon* filter.

- **Quantity**: Upon further discussion, we retained 15 sign-makus with the average of one signmaku per minute. With 2 filter options and unfiltered signmakus, this allowed us to split the signmakus into 3 groups of 5 signmakus and distribute them evenly across three 5-minute segments of the video, i.e., the beginning, the middle segment, and the final minutes of the video. For each group of signmakus, we included diverse emotions (two negatives, two positives, one neural). Notably, participants mentioned how having all 3 styles of signmakus randomly popping up could be distracting as the viewer would never know which style would appear next. Therefore, each group had the same style before switching to the next style of signmaku, so that participants would only encounter two changes in styles of signmaku throughout the video.
- **Quality and Clarity**: The signing quality and clarity were manually assessed by four researchers on the team. When creating filtered signmakus from round 1, we noticed variations in video signing quality and clarity among participants. Some participants' filtered signmakus were difficult to understand. Upon further discussion, we retained signmakus from only one of the participants that had best quality and clarity in both unfiltered and filtered signmakus. This participant had made 11 signmakus that spread across the 15-minute video timeline, and we cut their longer ASL recordings into several signmakus and remapped smaller signmakus back to the video timestamp based on the contents in her signmaku. Other factors related to signing quality and clarity were left for future exploration.

3.1.3 Participants Information. A total of 20 DHH participants were recruited. All participants in our study were students currently enrolled in college, ranging in age from 18 to 43. All participants were from Gallaudet University, a private university in the US focusing on education for DHH students. None of the participants were involved in the formative study described in Section 3.1. Out of the 20 participants, seven were identified as female, one as non-binary, and the remaining 12 as male. The majority, 17, identified their hearing status as d/Deaf within the DHH community, while the other three participants identified as hard of hearing. Seven participants were from a family where no immediate family member were DHH, whereas the other 13 participants had one or more DHH family members. All participants had learned ASL for at least one year, including 11 participants who started learning ASL from birth. Regarding their classroom language preferences, 11 participants felt comfortable with using ASL only, one participant felt comfortable with English only, and the other 8 participants were comfortable with both ASL and English in the classroom. As for ethnicity, 2 participants identified as Black, 1 as Asian, 11 as White, 5 as Multiracial, and 1 as Portuguese. Most were non-STEM majors e.g., business. See Table 1 for the demographics of the participants. Each participant was compensated at 20 USD per hour. The duration of each study ranged from one hour to two hours. The study plan was approved by the university IRB.

3.1.4 Experimental Design. We conducted a two-step user study, as shown in Fig. 2. The first step addressed RQ1, while the second step was to address RQ2. The study was followed by a survey and concluded with an interview. The prototype used for this study was developed based on the proposed design from the formative study, as shown in Fig. 2 (outcome of round 2). To help participants navigate through the prototype and study procedure, we prepared instructions in a pre-recorded ASL video, spoken and written English to the participants throughout the study. Participants could also ask questions in ASL or by typing in the chat box in Zoom. The researchers turned off their cameras while participants completed each step to avoid distraction and free up visual space, but cameras were turned on if the participant requested help or strayed from the procedure. All but one participant preferred to use ASL, while the remaining participants preferred to use spoken English. During the onboarding process, participants were introduced to the overall goals, procedural steps, and visual format, (e.g., where and how sign-maku would pop up). This preview gave participants a clear idea of where to look and what to expect. Each step of the procedure was reintroduced before each section, granting the participants additional opportunities to review the process and ask clarifying questions. The Deaf individual and hearing research assistants with interpreting experience took turns leading 19 out of 20 user study sessions (T1 - T19), and all recordings were transcribed by these same assistants. The remaining session was led by the first author with an interpreter.

Step 1 - Video Watching with Signmaku. First, participants watched a 15-minute error-free open-captioned video about AR, the same video used during the formative study described in Section 3.1. This step allowed us to observe how participants viewed the signmakus. Participants kept their webcams turned on and consented for AI to recognize their facial expressions while watching the video. However, participants could opt to hide their webcam viewer on their interface if they did not wish to see themselves while watching the video. During the video, 15 peers' signmakus, as collected from the formative study, appeared on the bottom right of the video - on average one signmaku per minute similar to the formative study. Each signmaku popped up in real-time and related to the information in the video at that moment.

During the formative study, participants suggested using filters to preserve privacy while keeping the content in a signmaku. To explore how the styles of signmakus may impact how participants perceive these signmakus, we used the *cartoon* and *robotic* filters, in addition to the *realistic* unfiltered signmakus, as shown in Fig. 2. There were 15 total signmakus, grouped into three sets of five signmakus, that each participant viewed at the same timestamps in the video. We detailed the process of choosing and refining signmakus used in this phase of the study in section 3.1. The order of appearance of the styles in the three sets of signmakus was randomized. Table 1 includes the order of styles received by each participant.

Step 2 - Creating Signmaku during Video Reviewing. After finishing the video for the first time, participants could review the video in part or in whole in order to add their own signmaku or text comments. This step provided insights into what options participants might consider when sharing signmaku with other learners. Participants saw all three signmaku styles before they decided which style they wanted to share. This provided an informed

Table 1: Demographics of the participants. All participants were recruited from a university that focuses on education for DHH. Each participant experienced the three conditions, featuring stylistically relevant styled signmaku comments, in a randomly assigned order while watching the video.

| PID	Gender	Hearing Status	Age	Ethnicity	Randomly Assigned Conditions
T1	Male	Deaf	32	Black	Cartoon - Robot - Raw
T2	Male	Hard of Hearing	19	Black	Raw - Robot - Cartoon
T3	Female	Deaf	19	Asian	Cartoon - Robot - Raw
T4	Female	Deaf	22	White	Raw - Robot - Cartoon
T5	Female	Deaf	43	White	Raw - Robot - Cartoon
T6	Male	Deaf	22	Portugal	Robot - Raw - Cartoon
T7	Male	Deaf	19	White	Robot - Raw - Cartoon
T8	Male	Deaf	22	Multiracial	Robot - Raw - Cartoon
T9	Male	Deaf	22	White	Robot - Cartoon - Raw
T10	Female	Deaf	21	Multiracial	Cartoon - Raw - Robot
T11	Female	Deaf	31	White	Cartoon - Raw - Robot
T12	Male	Deaf	18	White	Robot - Cartoon - Raw
T13	Male	Deaf	28	White	Cartoon - Raw - Robot
T14	Female	Deaf	20	Multiracial	Raw - Cartoon - Robot
T15	Male	Deaf	19	White	Robot - Cartoon - Raw
T16	Female	Deaf	43	White	Raw - Cartoon - Robot
T17	Non-binary	Deaf	28	Multiracial	Raw - Cartoon - Robot
T18	Male	Deaf	22	Multiracial	Raw - Robot - Cartoon
T19	Male	Hard of Hearing	26	White	Robot - Raw - Cartoon
T20	Male	Hard of Hearing	31	White	Cartoon - Robot - Raw
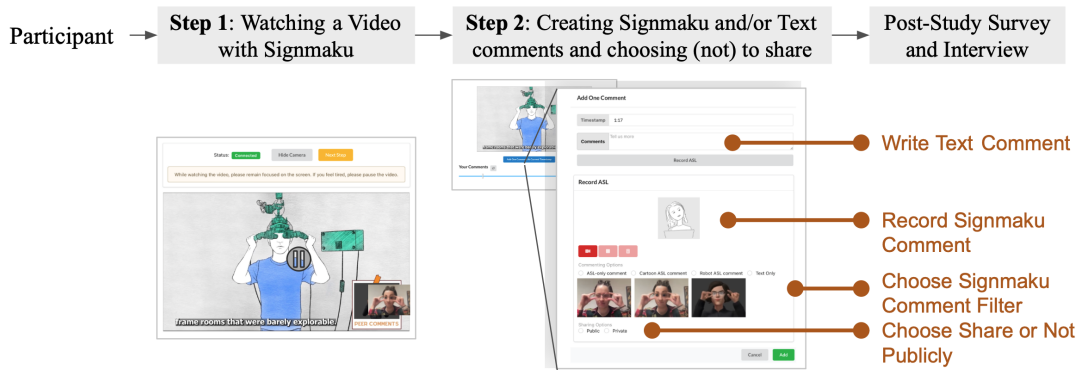
Figure 2: Experiment Design. Each participant completed two activities during the study. First, they watched a video about augmented reality (AR) with error-free open captions and signmakus (RQ1). Second, they provided their comments in signmaku (ASL comment) or text comments (RQ2). A comment may be consisted entirely of text, solely using ASL, or, on very rare occasions, include a mix of both. They then completed post-study survey and interviews. Participants were ask to provide signmaku only after they finished watching the video for the first time.

selection and separated the viewing (RQ1) and sharing (RQ2) effects of signmaku individually. Participants were allowed to add text comments, which emphasized the differences between recording an ASL comment and typing an English comment.

The video with embedded signmakus was identical in both step 1 and 2. When adding a comment, participants could jump directly to the desired timestamp and select a language preference: type a written English comment or record an ASL signmaku comment. We encouraged participants to add at least one signmaku and one

text comment, but they were not required to add both. Signmaku selections had two additional options: filters and sharing. Participants could apply one of three filters: *robotic*, *cartoon*, and *realistic*. Signmakus could be made public or kept private. Such customization enhanced participants' privacy control over each comment, which was a concern raised in the formative study.

Note that the focus of our study is on understanding user behavior and does not aim to make any technical contribution. Our prototype did not implement real-time filtering due to the lengthy

processing time. If time permitted during the interview, some participants' ASL comments were processed outside our current prototype, allowing them to provide additional feedback on the filters.

3.1.5 Measurements and Data Analysis. For each user study conducted to gather and assess feedback on the prototype, we collected the following data: responses from a post-study survey, emotions detected via webcam during video viewing, action logs, and interview responses.

Post-Study Survey: The survey consisted of two major parts: Part 1 focused on the viewing experience (RQ1), and Part 2 concentrated on the creating experience (RQ2). Additionally, we collected participants' demographic information as well as their preferred language in learning. Please refer to the appendix (Appendix A.1) for full survey questions.

Part 1- viewing experience: Participants were asked to rate their subjective task load *while* viewing the three signmaku styles. We focused on participants' opinions toward the task load (mental demand, physical demand, temporal pressure) *while* watching videos with different styles of signmaku. We adapted NASA Task Load Index [31] (Likert scale 1-10), as it has been used to study DHH interactions with technology, e.g., [29, 42]. Each participant responded to nine questions, divided equally among three styles (three questions per style). For example, a sample question was "How mentally demanding was viewing the pop-up ASL-only comments?"

After collecting the survey scores, we performed three repeated measures ANOVA tests with a randomized effect to compare the effect of signmaku style on (1) mental demand, (2) physical demand, and (3) time pressure. A random effect, "1/PID," was included to account for individual differences among participants that could not be explained by the fixed effects in the model [16], where "PID" was Participant ID as shown in Table 1 column 1. The power analysis was conducted in R using the WebPower package. Note that all statistically significant tests with this part of survey data remained significant ($p\_adj < 0.05$) after a Bonferroni correction; we still used "p" instead of "p_adj" when reporting data analysis in the findings.

Part 2 - creating experience: Participants were asked to answer 20 questions about their perceived sense of community when using our prototype design to view and share to answer RQ2. These questions were based on the Classroom Community Scale [26, 59]. After collecting participants' ratings, we first added up the survey responses to two composite scores of sense of community in learning - "sense of connectedness" and "sense of learning". Then, we ran correlations between survey responses on task load and survey responses on sense of community because we were curious if the increased load of signmaku, especially *robotic* found in RQ1 (presented in section 4.1.1), had any positive or negative relation with the overall experience. We first ran the correlation between two composite scores and each style's three questions for task load. The results were overall similar across the three styles, so we further conducted correlation analyses on each of the 20 questions. We compared each of the 20 questions with the three task load questions specific to each style.

Learners' Emotions Automatically Detected while Viewing: Prior works showed that learners' emotions could be estimated using facial recognition technologies continuously [38, 52]. In our study, we not only gathered data on participants' subjective perceptions but also utilized facial recognition technology, as described by [27], to identify their emotional responses during the learning process. To address RQ1, we conducted a statistical analysis to examine changes in participants' emotions before and after they were exposed to various styles of signmaku in Phase 1. Specifically, we looked at the "valence changes" in participants' emotional valence triggered by the appearance of a signmaku on the video display. We calculated the valence change by taking the value from the first two seconds of the signmaku popup and subtracting the average valence value over the entire 15 minutes of video viewing.

Each participant contributed a total of 15 emotion expression changes, comprising five times for each of the three styles. We utilized repeated measures ANOVAs to assess the impact of signmaku style on valence changes, which ranged from -1 to 1 (representing a spectrum from negative to positive, with 0 indicating a neutral state). To account for individual differences among participants, "1/PID" was also included as a random effect in the model, similar to the survey analysis. Since these technologies are not always 100% accurate in real-life applications, we provided participants with visualizations of their emotions over the video timeline, derived from facial data (as in [17]). At the beginning of the interview, we explained the data usage, allowing participants to review it and opt-out if desired. Participants were generally comfortable with our approach, suggesting it was adequate to conduct within-subject comparison. Note that all statistically significant tests with facial data remained significant ($p\_adj < 0.05$) after a Bonferroni correction; we still used "p" instead of "p_adj" when reporting data analysis in the findings.

Action Log: We used participants' action logs to understand the learning behaviors when providing signmaku. More specifically, we collected the following logs: the written text in English, the duration of typing text, the duration of crafting each signmaku, the number of re-recorded signmaku, the option of choosing the three styles of signmaku filter, and the choice of sharing or not. We compared counts of comments between the three types using the Wilcoxon Signed-Rank Test, with the results presented in section 4.2.2. We also compared the time spent per person and per clip on private/shared options using the same text, with the results presented in section 4.2.3. Due to our limited sample size, we did not conduct a further statistical comparison of effort for each type in the context of private vs public options. Additionally, we compared time spent per person and clip on ASL/typed English options using the same text. The details of this comparison are presented in section 4.2.4.

Post Study Interview: During the semi-structured interviews, participants were asked about their thoughts when using the prototype, their reasoning behind certain actions during previous steps, and their suggestions toward prototype improvement. Participants either opted to use ASL or spoken English to complete the interview. Sample interview questions included, "*which filter do you prefer or not prefer when viewing others' signmakus and why;*" "*when sharing your signmakus to others, do you have any concerns about two filtered options.*" A thematic analysis of the transcriptions extracted the main themes [68]. Two coauthors independently coded three of the transcripts and discussed together to create an initial version of the codebook. One coauthor further completed the remaining

transcripts. The two coauthors discussed frequently to revise the codebook and group codes into themes. Some sample codes include positives and difficulties of viewing or sharing signmakus, strategies adapted when using the interface, and accessibility or usability suggestions.

4 FINDINGS

4.1 Engaging Learners through Viewing Three Styles of Signmaku (RQ1)

Overall, the majority of the interviewed participants reported that signmakus were able to reengage and amuse them after being tired, because reading text-only captions could be *"tedious"*, *"boring"* and *"tiring"*. And the signmakus made participants feel more connected, e.g., they wondered *"what's the next comment would say"* and felt they *"we were watching the video together at the same time...not just passively watching"* (T1). The signmaku provided participants more visual information via ASL when feeling *''struggled...like science, with big words'',* which in turn increased their motivation: *"the peer comment actually helped with "visualizing"... ASL helps me learn more!"* (T15). T11 hoped to have signmaku for long videos to make the learning process more engaging and motivating.

When asked if they found competition between different information sources, some disagreed and further said they used different data sources to double confirm their understanding and to "fill in the blank." T11 specifically mentioned that, as DHH, they were used to and were good at using different information sources to guess and disambiguate. The signmaku was an additional information source that provided cues to support the process. Notably, some DHH learners shared that they developed strategies to pay attention to multiple visual elements in videos, allowing them to compensate for the lack of auditory information.

However, challenges remained when watching videos with signmaku. Interview data revealed that the main challenges agreed upon among participants were to quickly form their own new visual attention management strategy to fully comprehend the video content together with captions and signmaku. They would constantly switch visual attention to viewing video content and caption and signmaku. They considered the ability to understand both at the same time as a skill that needed practicing: *"I can see the video and peers' comments simultaneously, but I am not that skilled. It is challenging and distracting... Need practicing"* (T16). T4 also described a conflict in attention between caption and signmaku, *"It was interesting when I saw the ASL comments. I was excited to see and watch those and I just stopped looking at the captioning. Yeah... When I was reading the comments, I found myself agreeing..."*

We next present a detailed comparison between the three styles of signmaku.

4.1.1 Significant differences in perceptions of viewing the three signmaku styles. Participants were asked to rank the three styles of signmaku for 'enjoying the signmaku with the video.' The majority (66%) selected *realistic* as their first choice, 33% chose *cartoon* as their first choice, and 90% considered the *robotic* version as the least enjoyable. To investigate further, we conducted a statistical test to determine if the different styles of signmaku had different task demands when viewing. Three repeated measures ANOVA tests
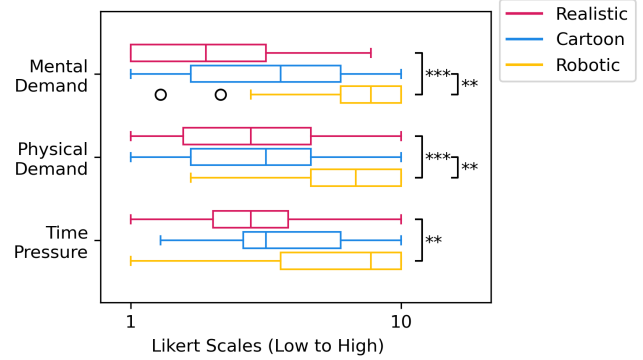

Figure 3: Boxplots of participants' reported mental demand, physical demand, and time pressure of viewing three styles of signmakus (N=20) on 10-point Likert scales (Low to High). The mental demand, physical demand, and time pressure were all significantly higher when viewing *robotic* ASL signmakus compared to the realistic styles. Note: ** signifies p< .01, and * signifies p<.001.**

with a randomized effect were performed to compare the effect of signmaku style on (1) mental demand, (2) physical demand, and (3) time pressure. A power analysis found sufficient statistical power for the study (0.81). Results are shown in Figure 3.

Mental Demand: There was a statistically significant difference in mental demand (F(2,57)= 17.25, p< .001) between at least two groups. On average, participants found the mental demand of viewing *realistic* to be very low on 10-point Likert scales (M=3.5, SD=2.7). The mental demand for the *cartoon* was slightly higher (M=5.7, SD=2.9) than *cartoon*. The *robotic* had very high mental demand (M=8.4, SD=2.3). A post hoc analysis found that the *robotic* signmakus had significantly higher mental demand than the *cartoon* and *realistic* signmakus (p< .01 and p< .001 respectively). The *cartoon* had similar mental demand as *realistic*.

Physical Demand: There was a statistically significant difference in physical demand (F(2,57)= 10.32, p< .001) between at least two groups. For the *cartoon*, the physical demands (M=5.3, SD=2.6) were medium and similar to *realistic* (M=4.7, SD=2.7). The *robotic* had very high physical demand (M=8, SD=2.1). A post hoc analysis found that the *robotic* signmakus had significantly higher physical demand compared to *cartoon* and *realistic* signmakus (p< .01 and p< .001 respectively) .

Temporal Pressure: There was a statistically significant difference in temporal demand (F(2,57)= 5.983, p< .01) between at least two groups. For the *cartoon*, the time pressure (M=6.0, SD=2.6) of viewing was similar to *realistic* (M=4.9, SD=2.6), and participants felt neither rushed nor relaxed. The *robotic* made participants feel very rushed (M=7.8, SD=2.8) had higher temporal pressure than *realistic* (p< .01) (post hoc analysis).

4.1.2 Realistic Style ASL yielded the least cognitive Load, followed by Cartoon ASL. Interview findings explain that real hands in both *cartoon* and *realistic* signmaku greatly contributed to the understandability of the content, which was absent in *robotic*. T1 explained that hands and lips heavily need improvement for *robotic*.

T4 found *cartoon* and *realistic* both looked like real human signing, but they "*didn't notice that it was at all humans*" and thought "*the robot is very hard to understand, hands are completely off...*"

Additionally, the use of facial expressions was key to helping one's understanding of the signed content including emotions, which was suppressed in *cartoon* compared to *realistic*. While the limitations of facial expressions of *cartoon* were noted by some of the participants. For example, T10 found *cartoon* understandable and commented on some facial movements:

> "*I liked the cartoon. The (cartoon's) signing was clear but it really wasn't the same as the real ASL signing. It's harder because the clarity of the cartoon's facial expressions was harder to understand. It (cartoon) wasn't as clear. But the (realistic) ASL comments were clear, and those facial expressions were clear.*""

When experiencing with *robotic* signmaku, participants created the metric, "smoothness," for measuring how well signing avatars presented different modalities together. T15 commented: "*The robot was not natural. It needs more flow. Not choppy. I wasn't sure what it said.*" T3 also complained: "*The robot is stiff, not flexible like a real person.*" T16 provided a more detailed explanation:

> "*I don't like a robot avatar at all. It is useless because it is chunky, and it is hard to understand its signing. I could not recommend using it at all for ASL. I learned an interesting tidbit from Dr. Annelies Kusters, who mentioned how you perceive sign language in texture. That got me thinking about ASL's texture, and it got me to understand that ASL is very smooth, flexible like water. So, the avatar for ASL should be curvy and smooth, whereas the robot is the opposite.*"

4.1.3 Inaccurate robotic ASL triggered the highest cognitive load, compared to cartoon and realistic being similar. Overall, *robotic* was hard to understand with fake and inaccurate hands and faces, which inclined some learners to zoom in and enlarge the signmaku–trying to see the signs more clearly. Also, the hands were using time and space awkwardly, making it hard to interpret semantic meaning. This was also associated with missing the chance to look at the original video content and the sense of pressure associated with feeling left out. Some eventually decide to ignore the *robotic*, as T5 explained:

> "*with the robot, I really had to concentrate. I didn't understand it at all. It wasn't worth the work. I felt like it ate up all of my energy that should have been on the content video. So, I started completely ignoring the robot. I continued to monitor the comments, but if it was the robot, I wouldn't look. I'd continue to look at the video.*"

Some agreed viewing *robot* required too much cognitive load that they intentionally decided to ignore it and to prioritize focus on the captions, which were more commonly mentioned for *robotic* than the two other styles. Some further suggested giving DHH learners more control so they can better know what to expect and where to look as they said without control "*I will be lost, and I don't know what to look at*" (T16). Some participants were attracted by the signmaku and forgot to look back at the captions. T4 found

robotic signmaku to be challenging to understand and eventually ignored it:

4.1.4 Cartoon signmaku brought a lot of joy. While participants watched the video, their facial movements were collected to estimate their emotional responses [27]. There was a statistically significant difference in valence (negative to positive) changes between at least two signmaku styles ($F_{(2,247)}$= 4.078, p< .05), indicating that the emotional changes from negative to positive were influenced by the style of signmaku being watched. No significant difference was found for arousal changes between signmaku styles. The power analysis indicates a good statistical power for the study (0.84).

Post hoc analyses revealed that *cartoon* increased positive emotions (M= 0.16, SD= 0.08) more than *robotic* (p< .05) (M= 0.008, SD= 0.05) and *realistic* (p< .05) (M= 0.007, SD= 0.05); and no significant differences were found between *robotic* and *realistic*. In summary, these results suggested that watching *cartoon* signmakus brought a positive emotional effect compared to watching *robotic* or *realistic* signmakus.

This result was aligned with participants' interview feedback, that the most common reason for *cartoon* to be preferred was that it felt fun to watch, as it "*reminds me of Disney due to big eyes.*" (T11). *Cartoon* was also regarded with high potential for customization and a wide range of choices it could offer, such as creating funny or role-playing content in varied learning and social contexts. For instance, T4 commented:

> "*...sometimes you can become an animal. Or your head can become a character? Or it'll make your face and body a different color...funny.*"

T16 also added that cartoon signmakus might not be preferred if the video was for a serious context, e.g., a meeting or business:

> "*The cartoon-style comments are fun to use. It depends on the seriousness of the context. Keeping a cartoon avatar would be a good idea to make the learning experience fun learning. For children, playing is learning, and it applies to any age. If you enjoy and you will enjoy learning. Using cartoon avatar depend on the context. In an informal environment, it could be fun to use.*"

Summary-RQ1 While signmaku, video, and captions might compete for DHH's visual attention, they provide complementary and visual information for disambiguating captions that are boring and tedious to follow. Participants' survey results showed that viewing *robotic* put the highest task load (mental demand, physical demand, time pressure) on them. According to their interview feedback, the *robotic* signmakus were the least accurate in delivering hand movements and facial expressions. On the contrary, *cartoon* signmakus did not pose more physical demand and time pressure than the *realistic* ones. Moreover, *cartoon* signmakus were favored by participants for their entertaining effect, which aligned with participants' facial expression changes captured by the automatic emotion recognition tool.

4.2 Engaging Learners via Creating and Sharing Signmaku (RQ2)

RQ2 reveals users' perceptions of creating and sharing signmaku *after* viewing signmaku in RQ1, using survey, system log analysis,

and interview. We compared the user experience of creating and sharing ASL comments to text comments, uncovering design opportunities. In total, our participants created 68 signmakus with an average of 3.4 signmakus per participant (SD= 3.8).

4.2.1 Creating and sharing cartoon signmaku promoted a sense of learning community. During the post-study survey, participants reported an increase in "sense of connectedness" (average score= 33.8/50, SD= 3.9) and "sense of learning" (average score= 28.8/50, SD= 5.9), when they were asked to compare the proposed design against video-based learning without viewing or creating signmaku themselves. We observed that the two composite scores, "sense of connectedness" and "sense of learning," are significantly positively correlated with the mental, physical, and temporal demands of all three types in an overall similar manner, especially for *cartoon* and *realistic* signmakus. Among all three styles, increased mental and physical demand positively correlates with a "sense of connectedness" but not with temporal pressure. Increased physical demand positively correlates with "sense of connectedness" among *realistic* (ρ= 0.61, p< .01) and *cartoon* (ρ= 0.50, p< .05), but not for *robotic* (ρ= 0.25, p> .05). Increased mental demand positively correlates with "learning opportunities" among *realistic* (ρ= 0.50, p< .05) and *cartoon* (ρ= 0.53, p< .05), but not for *robotic* (ρ= 0.33, p> .05) .

The interview findings suggested that being able to create and share comments also fostered a feeling of connectedness. For example, T1 was one of the many participants who re-watched all signmaku once again when providing their own comments and tried to reply to others. When asked why, T1 said,

> *"...I felt connected. It was like, 'I see you!' and I wanted to add to that in my comment. I felt like I could relate. 'I feel you...' Connected to that comment... Enough that you felt motivated to provide your own comment..."*

4.2.2 Cartoon offered a balance between anonymity and understandability. When creating signmaku, the options for *realistic* and *cartoon* were preferred over *robotic*. User interaction action logs recorded that our participants (N=20) created 33 *realistic* signmakus, 25 *cartoon* signmakus, and 10 *robotic* signmakus. There were significantly fewer *robotic* comments than the two other styles (p< .05). Among the three styles, 25 uploaded *cartoon* comments of which 11 were private and 14 public, while for *realistic* 21 were private and 12 public, and for *robotic* 3 were private and 7 public, respectively. In each style, approximately half of the comments were selected for sharing, and half were kept private. Regarding the sharing rate of various signmaku styles, there were no significant differences among the three styles when using a chi-square test. The limited sample size did not support further comparative analysis of each style.

The interview results found that the *robotic* style prioritized privacy at the expense of understandability in sharing, while the *cartoon* style was an appropriate compromise between both requirements. In contrast, the *realistic* style failed to maintain any level of privacy. Better ASL understandability is necessary for self (used as person note-taking for further reviews) and others (sharing publicly). For example, T15, who found *robotic* signmakus more

mentally challenging and harder to interpret than *cartoon* and *realistic*, switched between the two styles but eventually decided to use *robotic* for anonymity when providing signmakus:

> *"... I wasn't sure what it said, it's a lot of work...(for sharing) I prefer that one [realistic] because it feels more natural, like peer-to-peer...However, for those people who don't feel comfortable revealing themselves, I would recommend the robot... I think the robot is the best (for staying anonymous in sharing)."*

While having good understandability, *cartoon* also preserved an acceptable anonymity ability for some participants. T3, T5, T7, and T11 only made *cartoon* with 89% of these signmakus were made public. T11 said:

> *"In a classroom setting, it would be fine not to use the privacy options because I know the people. But if it's on a platform such as YouTube, I would use the cartoon. ... I prefer the avatar option because I'm shy.".*

Though our participants were instructed to generate and share ASL comments, their primary focus remained on comprehending ASL content, with ASL anonymization considered important but less so in comparison to understanding ASL content. In these cases, *cartoon* met their needs as a balance of both needs. From the interview comments, the rationale behind not sharing, instead of privacy and anonymity concerns, was that some participants believed that certain comments content were less beneficial for peer learners and were more closely tied to their individual learning experiences, serving as personal notes.

4.2.3 Sharing realistic signmaku increase effort in self-representation. The time spent on shared ASL clips (M= 108.8 seconds per person, SD= 91.3) was longer than those set as private (M= 97.7 seconds per person, SD= 200.8) (p= .052), indicating signing shared comments increased users' effort.

Sharing signed comments was also accompanied with the needs in self-representation. Five participants re-recorded their ASL videos, after deciding they would *share* the signmaku with others. After including the ASL re-recording, the total time for crafting ASL increased. The total time for crafting ASL comments averaged 168.7 seconds (SD=108.0) per participant, similar to the time for typing text. On average, participants had 0.8 re-recorded attempts per participant. The interview result explains they re-recorded the ASL clips for better *"precision, clarity and to be more related to the video context"* and *"better lighting and placing in the camera."* The two participants who favored providing text-only comments (all text comments were shared). They mentioned that text commenting was for the purpose of sharing; typing text allowed them to avoid the stress associated with impression management, expedited the study process, and enabled more precise expression through text editing. And the process of "editing" ASL comments may necessitate re-recording attempts, as elaborated in the quotes below:

> *"I want to sign of course, but sometimes I feel like I have to set everything up, look nice with the right clothes, have my makeup and hair styled, and all that. Ugh! It seems to take so much time when I could just type it all up and edit it easier and faster...With ASL, I have to*

think it through to make sure I use the right sign and the correct facial expression." – T4

Participants found that using two proposed filters for sharing required less self-representation effort compared to the *realistic* style. After including re-recording, time, the time spent on *realistic* (M= 45.8 seconds per comment, SD= 30.8) was longer than time spent on *robotic* (M=20.5 seconds per comment, SD= 17.1) (p< .05). Some were interested in using two proposed filters for self-presentation, however, the current prototype lacked real-time filtered views, leaving participants uncertain about how well the filters suited them. Participants suggested that having a preview of the filtered video would boost their willingness to use and share with two AI filters by allowing them to check for clear signing and facial movements. We did not conduct further comparison of effort for each type in the context of private vs public options due to our limited sample size.

4.2.4 Signmaku reduced the cost of creating compared to text-based commenting. Each participant was encouraged to create at least one text comment and one ASL comment. They had the choice to create comments solely in text, exclusively in ASL, or, on rare occasions, to opt for a combination of both. The prototype tool showed text-only comments anonymously by default, and participants were informed of the options during on-boarding.

Among all the participants, three preferred ASL comments only, and two preferred to create text comments only. We then compared the effort of participants' signmaku and text comments. After dropping five participants who preferred only one commenting method only, the remaining participants (N=15) uploaded 2.6 *realistic* ASL comments and 2.5 text comments on average. Given the time cost of converting the *realistic* ones to the other two styles and the constraint of each study, we did not have the participants convert their realistic comments to other styles. Notably, a paired-samples Wilcoxon signed-rank test indicated that the total time length of ASL comments (M= 74.0 seconds, SD= 66.7) was significantly lower than that for typing text comments (M= 190.6 seconds, SD= 153.6) (p< .05). Similarly, the average duration for each ASL comment (M= 26.9 seconds, SD= 22.8) was considerably shorter than the typing time for text comments (M= 94.4 seconds, SD= 80.6) (p< .001). After including ASL redo, time spent per clip on ASL averaged 61.0 seconds (SD= 19.8), and time spent per person on ASL averaged 168.7 seconds (SD= 108.0). Both were comparable to typing text per clip and per person, respectively (paired-samples Wilcoxon signed-rank test, p>.05).

In other words, time spent on ASL, including redo, was comparable to the time taken for text comments, both longer than ASL without redo. However, despite the similarity in time between ASL with redo and text comments, participant interviews shed light on the distinct differences: the majority of participants displayed more fluency and expressiveness in ASL than in typing text comments. Furthermore, their preference for ASL over text for language input was evident from their background surveys. A notable statement from T15 encapsulates this sentiment: *"As a Deaf person, fully expressing myself in English is challenging. In ASL, I can do so fluently and effectively"*

The content between text and ASL comments was similar. Both commenting mechanisms were used to share emotions and to relate

personal experiences towards the video content as well as peers' signmakus. For example, T8 commented on the video content using text comment: *The use of augmented reality in retail and shopping is intriguing.* They also created a signmaku to reply to peer's signmaku in the same context: *At 4:20 in the video, the student made a comment that I really agree with. I like augmented reality shopping. It's cool that you can pan you phone around the room to see if a couch can be moved around in a room. Once you know it fits, then you can buy it in person* (ASL clip transcribed to English).

Summary-RQ2 After participants created their own signmaku comments, we found a positive correlation between participants' perceived physical load when viewing the cartoon signmaku and some questions about perceived social connectedness and enthusiasm for learning. Such a relationship was not found when participants viewed the other two signmaku styles. Additionally, compared to traditional text-based commenting, commenting using ASL costs less time and is more expressive for DHH to create, though they would spare more time on the setup for self-representation, e.g., right clothing, if to share realistic signmaku.

5 DISCUSSION

5.1 Signmaku as an Edu-tainment Feature Promoting an Inclusive Video-based Learning Community

In our within-subject study exploring the impact of three styles of signmaku (*realistic*, *cartoon*, and *robotic*) on learning, the findings from RQ1 and RQ2 revealed that *cartoon* were preferred over *robotic* and *realistic*. *Cartoon* provided good understandability (RQ1 & RQ2) while preserving some anonymity (RQ2) which meets their need to share own comments, increases sense of learning community (e.g., inter-dependency) and further elicits the most positive emotion change (RQ2).

In RQ1, additional remarks highlighted that signmaku, regardless of style, enhances the participation and re-engagement of DHH during the learning process. It also imparts a greater wealth of visual information through sign language, a dimension that is absent in written English. Additionally, it assists in alleviating boredom often associated with reading text-only captions. Our discoveries provide a valuable expansion to the current video-based learning experience for DHH, which predominantly concentrates on improving video caption accuracy and designs (as seen in [13]).

RQ1 and RQ2 results suggest that signmaku can enhance inclusivity in video-based learning communities through entertainment and emotional engagement. These findings empirically reinforce the value of signmaku as an "edu-tainment" feature, aligning with Buckingham and Scanlon's concept of captivating learners' attention and evoking emotions to promote inclusive video learning [2]. We contribute signmaku as a new "edu-tainment" and learner-sourcing feature and reveal its applicability in a minority online learner population- DHH. In this case, these signmaku comments could be viewed as a co-created and sustained outcome stemming from the relationship between DHH and the video. Existing "edu-tainment" features in video-based learning include vibrant animations [2], creators offering authentic, casual, and personalized content [75], and teachers incorporating entertainment elements on live-streaming platforms [19]. Earlier studies have explored the use

of "edu-tainment" features to make sign language learning material more enjoyable for DHH children [56].

Our findings show the promise of signmaku in enhancing inclusive video-based learning through the integration of social interaction via a more accessible commenting mechanism and an improved sense of social connectedness. Specifically, our research builds on previous research by introducing DHH learners to a pseudo-synchronous learning experience and allowing them to use their preferred language (sign language) to interact with online learners they do not know, if needed, anonymously. As highlighted in previous research, social interactions have always held importance but have been lacking for DHH individuals during their formative years [23, 63, 66]. How will signmaku be adopted in the wild? Similar challenges have been identified among other minority learners as well (e.g., [64]). How may this new commenting feature be adopted by a broader audience? How can it be utilized to promote interaction between learners who have varied accessibility challenges? Our work has paved the way for greater inclusivity in online learning, marking a vast and promising avenue for further exploration and research.

5.2 Design Implications for Inclusive, Visual-Centric Video-based Learning

Our findings demonstrate that ASL offers a more accessible means for individuals who are DHH to both view their peers' comments and make their own comments. Additionally, ASL provides a visually richer source of information compared to text-based comments. Culturally Deaf individuals primarily rely on visual communication channels rather than auditory ones, as supported by previous research [10, 30, 53]. This underscores the significance of accommodating DHH users with ASL and visual learning in video-based learning, paralleling the importance of audio design for hearing users in auditory-focused environments, as demonstrated by previous research. For example, Arakawa et al. developed a video-based learning system that perturbs the voice in the video to help mind-wandering hearing users refocus their attention without consuming their conscious awareness [7]. Chen et al. studied audio note-taking in video-based learning tends to be more expressive and lengthier in comparison to text-based note-taking [18]. Our findings provide new design implications centering on how generative AI could be applied for inclusive video-based learning.

5.2.1 The Impact of AI-Generated Facial Expressions on Sign Language Comprehension. The comment made in English could be translated to ASL for DHH learners. Vice versa, ASL comments should be translated into English for general hearing users. With recent advancements in AI, the real-time generation for ASL is possible, e.g., [28]. It also allows DHH learners to break the circle and communicate with users in another language in their own preferred modality. ASL generated from English is also needed from those DHH who prefer to type or sign only, due to strong English & ASL skills and ease in editing typed English compared to editing or re-recording signed comments (RQ2).

Language comprehension is fundamental, and participants further contributed the metric "smoothness" for signing avatars, signifying the AI-generated signing avatar's capacity to seamlessly integrate various modalities. When formulating specific metrics

for the development of AI-generated ASL content, it is crucial to be included as evaluation criteria. It's also important to note that previous research has highlighted distinctions in temporal preferences when interacting with human signers versus AI-generated signing avatars. For example, DHH users may require faster signs and slower transitions than those typically employed by human signers [3]. Therefore, while the metrics for human signers can serve as a foundational reference, they may not directly apply to AI-generated signing avatars, necessitating tailored adjustments and refinements.

5.2.2 Expressiveness of Facial Movements Generated by AI for Sign Language Understandability. Our findings demonstrate a preference for *realistic* and *cartoon* signmaku over *robotic* signmaku in both viewing, creating, and sharing due to natural hand and body movements and at least some facial expressions making the language possible to understand. The *cartoon* signmaku has some facial movement with unfiltered hands, and the task demand was similar to *realistic* in all three dimensions, indicating that when facial cues are lessened, the communication is still understandable. And when hand, body, and facial movements are all lessened, the message becomes hard to understand. The above implies the fundamental importance of hand and body movements.

Meanwhile, as shown in RQ2, *realistic* is still preferred over *cartoon* for its understandability. It reveals that facial movement makes it not only understandable but truly expressive. For example, facial expression and body posture are often used to convey emotional magnitude (surprised vs. shocked. vs. aghast) and eyebrows may be raised or lowered depending on the form of question being asked [72]. The importance of the expressive body has been brought up by researchers in the accessible computing field in general [65]. Expanding on previous research that focused on how audiences perceive various ASL filtering styles [39], our study also considered the perspective of those sharing content. We discovered that users are less inclined to create and share signmakus with filtered styles that lack facial movements.

5.2.3 The Role of AI-Generated "Make-Up" Filters in Enhancing Sharing. RQ2 results suggested that the quality of ASL comments influences participants' williness to share. Some opt for re-recording to enhance quality, and T4 emphasizes the importance of reviewing and improving, particularly in facial movements. Therefore, 'Make-up' filters that automatically improve recording quality could improve willingness to share by reducing creation efforts. For example, filters can improve signed comments' expressiveness by adding vivid facial and eyebrow movements. Other non-linguistic features could be useful as well, such as lighting improvements. These filters underscore the potential of generative AI technology in providing understandable and expressive ASL content.

In RQ1, participants emphasized the desire for self-presentation options, including fun filters, alongside the anonymity feature, consistent with findings in our formative study, RQ2, and prior research [39]. Future research should explore diverse filter options for facial expressions, body language, and signing styles. As RQ2 suggests, sign language is rich in cultural and linguistic diversity, with variations in regional dialects, signing styles, and personalized expressions. Further generative AI technology should allow identity

expression within sign language avatars enabling learners to show-case their unique cultural and linguistic backgrounds, preserving the diversity of sign language communities and authenticity of the individuals. This dataset might also be interesting to researchers who want to contrast fluent and non-fluent signing – [32]. Similar to a study by Arakawa et al. [6], which discovered that voice conversion impacted speech ownership and implicit bias in self-presentation, it is essential to consider potential biases introduced by sign language filters.

5.2.4 Scalability of Signmaku Design Parameters. A significant area where AI can make valuable contributions is in automating the fine-tuning of design parameters during round 2 of formative studies (Phase I). Specifically, design aspects like "duration," "dimension," and "placement in relation to the video content" can be optimized by harnessing large-scale data and adapting to various user interface layouts. For example, among these parameters, managing the duration can be particularly challenging. We observed that participants comfortably diverted their attention from the learning content for up to 10 seconds and still easily re-assumed the content without feeling stressed. Typically, for educational ASL sentences, one full sentence, as determined in our initial formative study, takes approximately 10 seconds. However, it's worth noting that learners might prefer signing for a longer duration, and in our evaluation study (Phase II), our participants' signed comments lasted about 30 seconds. In our current approach, we manually edit the videos, but there is potential for designing a tool that offers auto-editing and streamlines this process. Also, our current design only presents one signmaku at one particular time of the video, future research might explore the possibility of displaying multiple signmaku simultaneously.

We want to emphasize that the parameters we selected for each design aspect are solely for our evaluation study (Phase II), which is based on our selected video on AR. We do not claim these parameters are generalized to broader contexts. Importantly, future research should investigate how participants' experiences are impacted when multiple users' comments are presented, as danmaku in nature allows multiple users commenting on the same video scene [22, 74]. Additionally, future research would need to evaluate how filters work for some signers but not for others, and how the filters can be further personalized based on how individuals express differently. This is particularly relevant in light of our needfinding, which revealed significant variations in individuals' signing quality and clarity when using filtered signmaku.

5.2.5 Promoting Deaf Cultural Identity in Online Social Interactions. It's crucial to recognize Deaf as a cultural identity, not as a disability, and promote mutual inclusivity through intercultural dialogue. To foster a sense of belonging within the Deaf culture, future video systems can consider encouraging DHH individuals to teach ASL during interactions with hearing people. Mack et al. found that Deaf signers want to share their language and culture with both hearing family members and friends [45], highlighting the need for online learning technology to facilitate intercultural engagement and language learning between Deaf and hearing individuals. Future research should also investigate hearing individuals' perception of signmaku, exploring design features like signmaku frequency control and AI-generated text captions using recent AI

technology [15]. Additionally, further research is required to understand meaningful two-way translation using AI, moving beyond word-for-word translations.

5.3 Limitations

We acknowledge a few limitations in our study. First, our participants are from a higher education school specifically for DHH students. DHH students from mainstream higher education schools might have different perspectives toward their interactions with other students and technology. Future studies should conduct studies with more DHH students with a more diverse background. Second, our study is conducted using American Sign Language (ASL), and the results may be different using other sign languages. Future studies may explore how different sign languages may impact the implementation of signmaku for online learning. Third, our current signmaku design did not allow participants to customize the parameters of signmaku. During the interview, participants suggested more ways to customize the signmaku experience, such as changing size and location of the signmakus or turn on or off signmakus. Future studies may explore how these interactions may impact the experience and effects of using signmakus in video-based learning. Future research could build beyond our work on understanding the sharing effect *after* video watching to *during* video watching. Fourth, we used facial expression recognition technology to automatically detect participants' emotion while watching the video. We did not use it to examine the detected emotions, rather we only compared the emotion changes. Future studies could apply more advanced tools to capture learners' emotions more accurately and re-evaluate our findings.

6 CONCLUSION

We introduced a new video commenting feature, Signmaku, which demonstrated promising results in enhancing video-based learning for DHH students. The *cartoon* signmaku created a more engaging and interactive learning environment, positively impacting users' perceived learning engagement and social interaction. Our novel interaction design has the potential to encourage learners to actively participate in the learning process and empower them to express their opinions openly. Our research expands the design space for online learning, thus addressing the specific needs of DHH individuals and providing access to a wide range of educational resources. We emphasize the need to address the limitations of generative AI techniques in creating sign language to ensure accurate and comprehensive content. Adopting signmaku and prioritizing inclusive online learning can unlock significant learning opportunities for DHH students, thereby promoting equity in education and enabling DHH students to thrive alongside their hearing peers.

research assistant, Samantha C. Moore, as well as all the participants for their thoughtful and reflective feedback.

REFERENCES

[1] Solmaz Abdi, Hassan Khosravi, Shazia Sadiq, and Gianluca Demartini. 2021. Evaluating the Quality of Learning Resources: A Learnersourcing Approach. *IEEE Transactions on Learning Technologies* 14, 1 (2021), 81–92. https://doi.org/10.1109/TLT.2021.3058644

[2] Nalan Aksakal. 2015. Theoretical view to the approach of the edutainment. *Procedia-Social and Behavioral Sciences* 186 (2015), 1232–1239.

[3] Sedeeq Al-khazraji, Becca Dingman, Sooyeon Lee, and Matt Huenerfauth. 2021. At a Different Pace: Evaluating Whether Users Prefer Timing Parameters in American Sign Language Animations to Differ from Human Signers' Timing. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[4] Muhammad Ali and Anza Nasir. 2023. Investigating the Social Media Practices among Aurally Challenged Students (ACS): The Challenges and Issues. *Human Nature Journal of Social Sciences* 4, 1 (2023), 518–532.

[5] Wajdi Aljedaani, Rrezarta Krasniqi, Sanaa Aljedaani, Mohamed Wiem Mkaouer, Stephanie Ludi, and Khaled Al-Raddah. 2022. If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: a literature review. *Universal access in the information society* (2022), 1–20.

[6] Riku Arakawa, Zendai Kashino, Shinnosuke Takamichi, Adrien Verhulst, and Masahiko Inami. 2021. Digital Speech Makeup: Voice Conversion Based Altered Auditory Feedback for Transforming Self-Representation. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 159–167.

[7] Riku Arakawa and Hiromu Yakura. 2021. Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[8] Charlotte Baker-Shenk. 1985. The facial behavior of deaf signers: Evidence of a complex language. *American Annals of the Deaf* 130, 4 (1985), 297–304.

[9] Georgina Batten, Peter M Oakes, and Tim Alexander. 2014. Factors associated with social interactions between deaf children and their hearing peers: A systematic literature review. *Journal of deaf studies and deaf education* 19, 3 (2014), 285–302.

[10] H Dirksen Bauman and Joseph Murray. 2009. Reframing: From hearing loss to deaf gain. *Deaf Studies Digital Journal* 1, 1 (2009), 1–10.

[11] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. 2019. Preferred appearance of captions generated by automatic speech recognition for deaf and hard-of-hearing viewers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[12] Larwan Berke, Matthew Seita, and Matt Huenerfauth. 2020. Deaf and hard-of-hearing users' prioritization of genres of online video content requiring accurate captions. In *Proceedings of the 17th International Web for All Conference*. 1–12.

[13] Bhavya Bhavya, Si Chen, Zhilin Zhang, Wenting Li, Chengxiang Zhai, Lawrence Angrave, and Yun Huang. 2022. Exploring collaborative caption editing to augment video-based learning. *Educational technology research and development* 70, 5 (2022), 1755–1779.

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

[15] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring collection of sign language datasets: Privacy, participation, and model performance. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14.

[16] Joseph Chee Chang, Amy X Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S Weld. 2023. CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[17] Si Chen, Desirée Kirst, Qi Wang, and Yun Huang. 2023. Exploring Think-Aloud Method with Deaf and Hard of Hearing College Students. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1757–1772. https://doi.org/10.1145/3563657.3595980

[18] Si Chen, Dennis Wang, and Yun Huang. 2021. Exploring the complementary features of audio and text notes for video-based learning in mobile settings. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[19] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. "I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.

[20] Yue Chen, Qin Gao, and Ge Gao. 2022. Timeline-anchored comments in video-based learning: The impact of visual layout and content depth. *International Journal of Human–Computer Interaction* 38, 9 (2022), 868–883.

[21] Yue Chen, Qin Gao, and Pei-Luen Patrick Rau. 2015. Understanding gratifications of watching danmaku videos–videos with overlaid comments. In *Cross-Cultural Design Methods, Practice and Impact: 7th International Conference, CCD 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I 7*. Springer, 153–163.

[22] Yue Chen, Qin Gao, and Pei-Luen Patrick Rau. 2017. Watching a movie alone yet together: understanding reasons for watching Danmaku videos. *International Journal of Human–Computer Interaction* 33, 9 (2017), 731–743.

[23] Vaso Constantinou, Andri Ioannou, Iosif Klironomos, Margherita Antona, and Constantine Stephanidis. 2020. Technology support for the inclusion of deaf students in mainstream schools: a summary of research from 2007 to 2017. *Universal Access in the Information Society* 19 (2020), 195–200.

[24] Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. 2023. Machine translation from signed to spoken languages: state of the art and challenges. *Universal Access in the Information Society* (apr 2023). https://doi.org/10.1007/s10209-023-00992-1

[25] Geoffrey Restall Coulter. 1979. *American sign language typology.* University of California, San Diego.

[26] Shane Dawson. 2006. A study of the relationship between student communication interaction and sense of community. *The Internet and Higher Education* 9, 3 (2006), 153–162.

[27] Didan Deng, Zhaokang Chen, and Bertram E Shi. 2020. Multitask emotion recognition with incomplete labels. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 592–599.

[28] Aashaka Desai, Lauren Berger, Fyodor O Minakov, Vanessa Milan, Chinmay Singh, Kriston Pumphrey, Richard E Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, et al. 2023. ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition. *arXiv preprint arXiv:2304.05934* (2023).

[29] A'di Dust, Carola Gonzalez-Lebron, Shannon Connell, Saurav Singh, Reynold Bailey, Cecilia Ovesdotter Alm, and Jamison Heard. 2023. Understanding Differences in Human-Robot Teaming Dynamics between Deaf/Hard of Hearing and Hearing Individuals. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 552–556.

[30] Michele Friedner and Annelies Kusters. 2015. *It's a small world: International deaf spaces and encounters.* Gallaudet University Press.

[31] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[32] Saad Hassan, Matthew Seita, Larwan Berke, Yingli Tian, Elaine Gale, Sooyeon Lee, and Matt Huenerfauth. 2022. ASL-Homework-RGBD Dataset: An annotated dataset of 45 fluent and non-fluent signers performing American Sign Language homeworks. *arXiv preprint arXiv:2207.04021* (2022).

[33] Changyang He, Lu He, Tun Lu, and Bo Li. 2021. Beyond Entertainment: Unpacking Danmaku and Comments' Role of Information Sharing and Sentiment Expression in Online Crisis Videos. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–27.

[34] Carl Jensema, Ralph McCann, and Scott Ramsey. 1996. Closed-captioned television presentation speed and vocabulary. *American Annals of the deaf* (1996), 284–292.

[35] Sushant Kafle and Matt Huenerfauth. 2016. Effect of speech recognition errors on text understandability for people who are deaf or hard of hearing. (2016).

[36] Karen L Kritzer and Chad E Smith. 2020. Educating deaf and hard-of-hearing students during COVID-19: What parents need to know. *The Hearing Journal* 73, 8 (2020), 32.

[37] Raja Kushalnagar, Matthew Seita, and Abraham Glasser. 2017. Closed ASL interpreting for online videos. In *Proceedings of the 14th International Web for All Conference*. 1–4.

[38] Imane Lasri, Anouar Riad Solh, and Mourad El Belkacemi. 2019. Facial emotion recognition of students using convolutional neural network. In *2019 third international conference on intelligent computing in data sciences (ICDS)*. IEEE, 1–6.

[39] Sooyeon Lee, Abraham Glasser, Becca Dingman, Zhaoyang Xia, Dimitris Metaxas, Carol Neidle, and Matt Huenerfauth. 2021. American sign language video anonymization to support online participation of deaf and hard of hearing users. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[40] Franklin Mingzhe Li, Cheng Lu, Zhicong Lu, Patrick Carrington, and Khai N Truong. 2022. An exploration of captioning practices and challenges of individual content creators on YouTube for people with hearing impairments. *arXiv preprint arXiv:2201.11226* (2022).

[41] Shugang Li, He Zhu, Ying Qian, Shiqi Ren, and Bing Fang. 2022. Classification and quantification of Danmaku Interactions in online video lectures: An exploratory study. https://www.hindawi.com/journals/wcmc/2022/5656669/

[42] Ziming Li, Shannon Connell, Wendy Dannels, and Roshan Peiris. 2022. Sound-VizVR: Sound Indicators for Accessible Sounds in Virtual Reality for Deaf or

Hard-of-Hearing Users. In *Conference on Computers and Accessibility (ASSETS'22)*.

[43] Junfeng Liao, Rundong Li, Huizhong Liao, and Junkai Huang. 2023. Research on the Influencing Factors of Online Video Danmaku Watching and Participation. In *2023 4th International Conference on Education, Knowledge and Information Management (ICEKIM 2023)*. Atlantis Press, 711–721.

[44] Xi Lin, Mingyu Huang, and Leslie Cordie. 2018. An exploratory study: using Danmaku in online video-based lectures. *Educational Media International* 55, 3 (2018), 273–286.

[45] 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 (2020), 1–31.

[46] Marc Marschark, Patricia Sapere, Carol Convertino, and Rosemarie Seewagen. 2005. Access to postsecondary education through sign language interpreting. *Journal of Deaf Studies and deaf education* 10, 1 (2005), 38–50.

[47] Emma McDonnell. 2022. Understanding Social and Environmental Factors to Enable Collective Access Approaches to the Design of Captioning Technology. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.

[48] Nayan Mehta, Suraj Pai, and Sanjay Singh. 2020. Automated 3D sign language caption generation for video. *Universal Access in the Information Society* 19 (2020), 725–738.

[49] Ross Mitchell, Travas Young, Bellamie Bachleda, and Michael Karchmer. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies* 6 (03 2006). https://doi.org/10.1353/sls.2006.0019

[50] Huda Shaaban Awed Mohammad Ahmed Hammad. 2023. Investigating Social Media Usage Among the Deaf and Hard of Hearing Students. *Journal of Southwest Jiaotong University* 58, 1 (2023).

[51] United Nations. 2022. https://www.un.org/en/coronavirus/disability-inclusion

[52] Omid Noroozi, Héctor J Pijeira-Díaz, Marta Sobocinski, Muhterem Dindar, Sanna Järvelä, and Paul A Kirschner. 2020. Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies* 25 (2020), 5499–5547.

[53] Carol Padden, Tom Humphries, and Carol Padden. 2009. *Inside deaf culture*. Harvard University Press.

[54] Soraia Prietch, J Alfredo Sánchez, and Josefina Guerrero. 2022. A Systematic Review of User Studies as a Basis for the Design of Systems for Automatic Sign Language Processing. *ACM Transactions on Accessible Computing* 15, 4 (2022), 1–33.

[55] Srijith Radhakrishnan, Nikhil C Mohan, Manisimha Varma, Jaithra Varma, and Smitha N Pai. 2022. Cross Transferring Activity Recognition to Word Level Sign Language Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2446–2453.

[56] Hidayah Binti Rahmalan, Zuraida Binti Abal Abas, Zaheera Binti Zainal Abidin, Abdul Samad Bin Shibghatullah, Ahmad Fadzli Nizam Bin Abdul Rahman, Chong Wei Shin, et al. 2018. Edutainment Tools for Deaf and Impairment Hearing Children: A Survey. *Advanced Science Letters* 24, 3 (2018), 1805–1808.

[57] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2021. Sign language recognition: A deep survey. *Expert Systems with Applications* 164 (2021), 113794.

[58] Filipa M Rodrigues, Joana R Rato, Ana Mineiro, and Ingela Holmström. 2022. Unveiling teachers' beliefs on visual cognition and learning styles of deaf and hard of hearing students: A Portuguese-Swedish study. *Plos one* 17, 2 (2022), e0263216.

[59] Alfred P Rovai. 2002. Development of an instrument to measure classroom community. *The Internet and higher education* 5, 3 (2002), 197–211.

[60] Marija Sablić, Ana Mirosavljević, and Alma Škugor. 2021. Video-based learning (VBL)—past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning* 26, 4 (2021), 1061–1077.

[61] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2022. Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5141–5151.

[62] Jeff Seaman and Hester Tinti-Kane. 2013. *Social media for teaching and learning*. Pearson Learning Systems London.

[63] 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*. 68–80.

[64] Umesh Sharma and Spencer J Salend. 2016. Teaching assistants in inclusive classrooms: A systematic analysis of the international research. *Australian Journal of Teacher Education (Online)* 41, 8 (2016), 118–134.

[65] Katta Spiel and Robin Angelini. 2022. Expressive Bodies Engaging with Embodied Disability Cultures for Collaborative Design Critiques. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.

[66] Michael Stinson and Shirin Antia. 1999. Considerations in educating deaf and hard-of-hearing students in inclusive settings. *Journal of deaf studies and deaf education* 4, 3 (1999), 163–175.

[67] Shengeng Tang, Dan Guo, Richang Hong, and Meng Wang. 2021. Graph-based multimodal sequential embedding for sign language translation. *IEEE Transactions on Multimedia* 24 (2021), 4433–4445.

[68] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.

[69] Elena Tomasuolo, Tiziana Gulli, Virginia Volterra, and Sabina Fontana. 2021. The Italian Deaf community at the time of Coronavirus. *Frontiers in Sociology* 5 (2021), 125.

[70] Ehsan Toofaninejad, Esmaeil Zaraii Zavaraki, Shane Dawson, Oleksandra Poquet, and Parviz Sharifi Daramadi. 2017. Social media use for deaf and hard of hearing students in educational settings: a systematic review of literature. *Deafness & Education International* 19, 3-4 (2017), 144–161.

[71] Ashley Walker, Yaxing Yao, Christine Geeng, Roberto Hoyle, and Pamela Wisniewski. 2019. Moving beyond 'one size fits all': research considerations for working with vulnerable populations. *Interactions* 26 (10 2019), 34–39. https://doi.org/10.1145/3358904

[72] Traci Patricia Weast. 2008. *Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows*. The University of Texas at Arlington.

[73] Qunfang Wu, Yisi Sang, and Yun Huang. 2019. Danmaku: A new paradigm of social interaction via online videos. *ACM Transactions on Social Computing* 2, 2 (2019), 1–24.

[74] Qunfang Wu, Yisi Sang, Shan Zhang, and Yun Huang. 2018. Danmaku vs. forum comments: understanding user participation and knowledge sharing in online videos. In *Proceedings of the 2018 ACM International Conference on Supporting Group Work*. 209–218.

[75] Haijun Xia, Hui Xin Ng, Zhutian Chen, and James Hollan. 2022. Millions and billions of views: Understanding popular science and knowledge communication on video-sharing platforms. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 163–174.

[76] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. 2022. Vtoonify: Controllable high-resolution portrait video style transfer. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–15.

[77] Yaxing Yao, Jennifer Bort, and Yun Huang. 2017. Understanding Danmaku's potential in online video learning. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 3034–3040.

[78] Meehyun Yoon, Jungeun Lee, and Il-Hyun Jo. 2021. Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning. *The Internet and Higher Education* 50 (2021), 100806.

[79] Kexin Zhang, Elmira Deldari, Zhicong Lu, Yaxing Yao, and Yuhang Zhao. 2022. "It's Just Part of Me:" Understanding Avatar Diversity and Self-presentation of People with Disabilities in Social Virtual Reality. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.

[80] Yu Zhang, Changyang He, Huanchen Wang, and Zhicong Lu. 2023. Understanding Communication Strategies and Viewer Engagement with Science Knowledge Videos on Bilibili. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 668, 18 pages. https://doi.org/10.1145/3544548.3581476

A APPENDICES

A.1 Survey Questions

Below, we present the survey questions used in our Phase II evaluation study. Note: In the surveys, we used the term "ASL-based Danmaku" instead of "signmaku." The three types of signmakus were also referred to using different terms for easier understanding: realistic signmakus were referred to as "ASL-only comments;" cartoon signmakus were referred to as "Cartoon ASL comments;" and robotic signmakus were referred to as "Robot ASL comments."

A.1.1 Part 1: Three types of Signmakus for Video-based Learning. Danmaku is a YouTube-like video-sharing platform. Viewers' comments appear on the screen and each comment can be customized (e.g., size, color, etc.). The comments go beyond real-time, with previous and later comments being shown together.

In our tool, you experienced three types of ASL-based Danmaku: ASL-only comments, Cartoon ASL comments, and Robot ASL comments. Rank how much you enjoyed the three ASL-based Danmaku types you saw in the video. *(Tied ranking is not allowed.).*

Realistic Signmakus.

- How **mentally demanding** was **viewing** the pop-up ASL-only comments? Mentally demanding refers to actions (e.g., thinking, deciding, remembering, looking, searching, etc.) that were hard to do WHILE viewing the ASL-only comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How **physically demanding** was **viewing** the ASL-only comments? Physically demanding refers to actions (e.g., clicking, pausing, moving body forward to read, switching eye gaze between features, etc.) that were hard to do WHILE viewing the ASL-only comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How much **time pressure** did you feel **viewing** the ASL-only comments and the video at the same time? Was the pace slow & relaxed OR rapid & rushed?
  Choose on a 10-point Likert scale, where 1 is Slow & Relaxed and 10 is Rapid & Rushed.

Cartoon Signmakus.

- How **mentally demanding** was **viewing** the pop-up Cartoon ASL comments? Mentally demanding refers to actions (e.g., thinking, deciding, remembering, looking, searching, etc.) that were hard to do WHILE viewing the Cartoon ASL comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How **physically demanding** was **viewing** the Cartoon ASL comments? Physically demanding refers to actions (e.g., clicking, pausing, moving body forward to read, switching eye gaze between features, etc.) that were hard to do WHILE viewing the Cartoon ASL comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How much **time pressure** did you feel **viewing** the Cartoon ASL comments and the video at the same time? Was the pace slow & relaxed OR rapid & rushed?
  Choose on a 10-point Likert scale, where 1 is Slow & Relaxed and 10 is Rapid & Rushed.

Robotic Signmakus.

- How **mentally demanding** was **viewing** the pop-up Robot ASL comments? Mentally demanding refers to actions (e.g., thinking, deciding, remembering, looking, searching, etc.) that were hard to do WHILE viewing the Robot ASL comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How **physically demanding** was **viewing** the Robot ASL comments? Physically demanding refers to actions (e.g., clicking, pausing, moving body forward to read, switching eye gaze between features, etc.) that were hard to do WHILE viewing the Robot ASL comments.
  Choose on a 10-point Likert scale, where 1 is Easy (low demand) and 10 is Hard (high demand).
- How much **time pressure** did you feel **viewing** the Robot ASL comments and the video at the same time? Was the pace slow & relaxed OR rapid & rushed?
  Choose on a 10-point Likert scale, where 1 is Slow & Relaxed and 10 is Rapid & Rushed.

A.1.2 Part 2: Perceived Sense of Community for Learning. (20 Questions). These 20 questions were developed based on Rovai's Classroom Community Scales [26, 59]. The questions were divided into 2 sets of 10 questions measuring two factors: connectedness and learning. Below, we present these two sets of questions separately. The number before each question represents the order of appearance. For each question, participants may choose between the following options: *Strongly disagree, Disagree, Neutral, Agree, Strongly agree*, which is converted to 1-5 in the analysis, and certain statements' answers are converted to 5-1 according to [26, 59].

First, rank how much you enjoyed providing three ASL-based Danmaku to future learners. *(Tied ranking is not allowed.)*

Then, answer the following questions. When we refer to "tool", please consider you creating and sharing your own ASL comments.

Sense of Connectedness: Compared to using video to learn by myself,

1. I feel students using this tool care about each others.
3. I feel connected to others using this tool.

5. I do NOT feel a sense of community using this tool.
7. I feel using this tool is like family.
9. I feel isolated using this tool.
11. I trust others using this tool.
13. I feel I can rely on others using this tool.
15. I feel like others using this tool depend on me.
17. I feel uncertain about others using this tool.
19. I feel confident others will support me using this tool.

Sense of Learning: Compared to using video to learn by myself,

2. I feel I am encouraged to ask questions using this tool.
4. I feel it hard to get help when I have a question using this tool.
6. I feel I receive timely feedback using this tool.
8. I feel UNeasy that this tool exposes gaps in my understanding.
10. I feel reluctant to sign/speak openly using this tool.
12. I feel using this tool results in only modest learning.
14. I feel others do NOT help me learn using this tool.
16. I feel I am given plenty opportunities to learn using this tool.
18. I feel my educational needs are NOT being met using this tool.
20. I feel that using this tool does NOT promote my desire to learn.

A.1.3   Demographics Information.
- What is your status as undergraduate student?
  Choose one of the following: Freshman; Sophomore; Junior; Senior; Other: [textbox].
- What is your gender identity?
  [textbox]
- What is your ethnicity?
  [textbox]
- What is your age?
  [textbox]
- How do you identify (within the Deaf community)?
  Choose one of the following: Deaf; Hard of Hearing; Hearing.
- Which of the following statements is true about your immediate family (parents and siblings)?
  Choose one of the following: One or more of my family members are deaf or hard of hearing; None of my immediate family members is deaf or hard of hearing.
- What language did you use growing up in face-to-face communication?
  Rank the following options: English; Other SPOKEN language; ASL; Other SIGN language. Participants may choose "Does not apply" for any of the options.
- At what age did you learn ASL?
  Choose "From birth (0-12 months)" or any number from dropdown box.
- What is your country of origin?
  [textbox]
- Is English your 1st written language?
  Choose one of the following: Yes; No.
- Which are you most comfortable with as a classroom instructional language?
  Choose one of the following: ASL; English; Both ASL and English.
- What is your major (or planned major)?
  [textbox]