Evaluating an AI Bi-Directional System for Communication between Deaf and Hard of Hearing Individuals and Hearing Persons: A Pilot Case Study

Zhiwei Yu*
Sign-Speak
Rochester, USA
jay@sign-speak.com

Nicholas Wilkins
Sign-Speak
Rochester, USA
nicholas@sign-speak.com

Yamillet Payano
Sign-Speak
Rochester, USA
yami@sign-speak.com

Nikolas Kelly
Sign-Speak
Rochester, USA
niko@sign-speak.com

Abstract

The Deaf and Hard of Hearing (D/HH) community faces significant communication gaps, limiting their full participation in everyday settings such as education and healthcare. This study designed an Artificial Intelligence (AI)-driven bi-directional communication system and demonstrated its efficiency via two usability tests of D/HH individuals to narrow those gaps. The bi-directional communication system comprised two major components: Sign Language Recognition (SLR) and Sign Language Production (SLP). Usability tests were conducted to survey D/HH individuals' communication preferences on 1) bi-directional system versus one-way system (only provided SLP) versus zero-way system (typing back and forth); 2) cartoon avatars versus human-like avatars. Results collected from 66 D/HH individuals showed that: 1) AI-driven communication systems should provide bi-directional support; 2) AI-generated avatars should be human-like. This work offered valuable insights for future bi-directional communication system design and SLP development for D/HH community.

CCS Concepts

• **Human-centered computing** • **Accessibility** • **Accessibility systems and tools**;

ACM Reference Format:
Zhiwei Yu, Nicholas Wilkins, Yamillet Payano, and Nikolas Kelly. 2025. Evaluating an AI Bi-Directional System for Communication between Deaf and Hard of Hearing Individuals and Hearing Persons: A Pilot Case Study. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3706678

CHI EA '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3706678

1 INTRODUCTION

Deaf and Hard of Hearing (D/HH) refers to individuals with varying degrees of hearing loss [45]. There are 48 million D/HH individuals in America and 466 million worldwide [10, 22]. Many D/HH Americans' first and primary language is American Sign Language (ASL), a language distinct from English, expressed through hands, body language, and facial expressions. Although ASL and English are two entirely different languages [36], D/HH individuals are often expected to navigate daily life using English due to a prevalent misconception that ASL is a visual representation of English. The limited access to ASL coupled with the expectation to rely on English leads to profound challenges, particularly in education. One result of this limited accessibility is the disparity in reading level between D/HH high school graduates and their hearing counterparts: 20% of Deaf high school graduates have English reading skills at or below second-grade level, while 33% read between second-grade and fourth-grade levels [9]. These educational disparities, stemming from systemic lack of ASL access, perpetuate communication barriers for D/HH individuals throughout their lives.

Communication options for D/HH individuals typically rely on English-centric methods or sign language interpretation. Despite being the gold standard for Deaf individuals, interpreter access is declining due to scarcity: although there are approximately 9 million ASL signers within the US [20], only 10,000 certified interpreters[39] exist. These barriers affect D/HH individuals' social interactions, education, and healthcare access, leading to significant disparities. There is a 16% employment gap (54% of D/HH individuals employed vs. 70% of hearing individuals), a 15% education gap (18% of D/HH individuals with bachelor's degrees vs. 33% of hearing individuals), and a 22.1% labor force participation disparity (42.9% of D/HH individuals not participating vs. 20.8% of hearing individuals) [17]. These systemic inequalities hinder millions of Americans from reaching their full potential due to addressable communication barriers.

Driven by advancements in Artificial Intelligence (AI), Sign-Speak[1] pioneered a bi-directional communication system that combines Sign Language Recognition (SLR) and Sign Language Production (SLP), inspired by our personal experiences with systemic

[1]A startup whose AI-powered language software recognizes ASL and translates it into spoken words and vice versa

barriers facing the D/HH community. This case study examined these two key technologies that emulate different interpretation functions: SLR, which translates sign language captured via camera into voice/text, and SLP, which converts spoken language into a real-time visual sign language output. Our goal is to provide functional equivalence in technology access for D/HH individuals, addressing not only the communication gap, but also the exclusion of signers from voice-activated systems such as virtual assistants, smart home devices, and automated customer services. By providing functional equivalence, we aim to be another tool in our users' accessibility toolkit. To become a valuable tool for the community, we believed that incorporating input from D/HH users is crucial in designing systems that effectively meet their communication needs. By doing so, we can enhance user satisfaction and encourage broader adoption of these technologies, especially given the challenges in balancing accuracy, user experience, and practical implementation.

To address these complexities, we crafted this case study with the following objectives: 1) described the development of automated interpretation systems; 2) analyzed optimal design strategies for these systems; 3) presented practical guidance for the 'optimal' design of these systems based on our findings. We hoped this study would highlight effective development methods and inspire continued innovation in the field. The rest of this case study is organized as follows. Section 2 reviews the background related to topics. Section 3 presents bi-directional system design. Section 4 details the case study design. Section 5 illustrates results. Section 6 summarizes contributions, concludes the study, and discusses limitations and future research directions.

2 RELATED WORK

This section examined prior research relevant to the central themes of the current study and outlines the motivations behind our key innovations. Section 2.1 discusses how SLR are researched and used in the community. Section 2.2 introduces previous works on SLP. Finally, Section 2.3 reviews the background of system usability in this domain.

2.1 Sign Language Recognition Systems

SLR refers to the use of computer vision and natural language processing techniques to interpret and understand sign language automatically [30]. It is a growing field within Human-computer Interaction, aiming to narrow the communication gap between D/HH community and hearing individuals for various contexts, such as facilitating communication with providers [15], learning mathematics [1], and enhancing accessibility in public services (e,g., bank) [25]. Additionally, this technology can help achieve functional equivalence by being integrated into platforms with voice recognition capabilities.

Recent advancements in machine learning and natural language processing have fostered significant interest in developing automated SLR systems that can facilitate real-time translation between sign language and spoken or written languages for the D/HH community. For example, Rastgoo et al. [29] proposed a real-time isolated hand sign language recognition model that included a single shot detector, 2-dimensional CNN, singular value decomposition, and Long Short-Term Memory (LSTM) to extract and process discriminative features from 3-dimensional hand key-points. They confirmed that the model achieved competitive results in both accuracy and recognition time on four benchmark datasets (e.g., RKS-PERSIANSIGN (99.5 ± 0.04) [28]), demonstrating its efficiency. Furthermore, Lee et al. [18] designed an application that included a LEAP motion controller for real-time ASL recognition in a whack-a-mole game format to improve the effectiveness of ASL learning. An LSTM combined with k-Nearest-Neighbor was used to classify static and dynamic ASL signs based on extracted features such as finger angles, distances, and sphere radius. Results showed that the model achieved an average recognition accuracy of 91.82%. Sharma et al. [32] employed a 3D Convolutional Neural Network-based model to recognize dynamic signs in ASL from volumetric video data. They demonstrated that the proposed approach outperformed the existing novel models (i.e., 3.7% improvement in precision).

However, it is important to note that these successful approaches have all either been **constrained** (limited vocabulary size (< 100)) or **isolated** (the model classifies individual signs rather than translating complete signed phrases into English). In addition, such unconstrained continuous SLR systems trained on datasets like OpenASL [34] reported BLEU scores not exceeding 10 [19, 34], indicating insufficient accuracy for practical use. Most studies have limited their evaluation to validation sets, often neglecting real-world testing with D/HH signers and thorough analysis of human factors. Few attempts have assessed automated SLR systems' feasibility in real-life settings, heavily relying on Wizard of Oz techniques [38] for human-computer interaction components. Therefore, we developed an unconstrained and continuous SLR system that enables D/HH individuals to sign freely using any device and tested the system's practicality with D/HH community in real-world scenarios.

2.2 Sign Language Production: Avatar Rendering

In bidirectional systems, SLP is crucial as D/HH individuals have varied English proficiency [5].In SLP, avatar rendering is a graphical representation of human figures to visually depict sign language [2]. While most common approaches use 3D rendering software to generate models replicating hand movements [6], facial expressions [12, 44], and body language [16, 33], our system directly generates image sequences. We focused on systems capable of generating signed content from text, ensuring broad applicability beyond simple motion sequence masking.

Advancements in AI have made avatar rendering widely applied and have shown promising acceptance results in ASL. These approaches can be divided into models driven by 3D rendering systems, which we term cartoon-based avatars, and systems driven by generative image modeling (directly regressing to images from some condition set), which we term human-like [41]. For example, Quandt et al. [26] introduced the embodied learning-based Signing Avatars and Immersive Learning (SAIL) system that rendering a high-quality cartoon signer by tracking gestures using the LEAP Motion system, which aims to teach ASL in the virtual reality environment. Their usability test disclosed that users reacted positively to the overall experiences and reported positive feedback about the potential for learning ASL through the SAIL system. Xu et al.

[43] applied a transformer-based Conditional Variational Autoencoder to generate ASL fingerspelling alphabets and evaluated it on three different mainstream video-based human representations: two-stream inflated 3D ConvNet, 3D landmarks of body joints, and rotation matrices of body joints. Results showed that the best ASL alphabet signing generation was achieved using rotation matrices of the upper body joints and signing hand. Baltatzis et al. [3] proposed a diffusion-based SLP model which generated motion sequences and rendered them via a 3D rendering system. Their model was trained on a large-scale dataset of 3D dynamic ASL sign sequences with associated text transcripts. They argued that the proposed method considerably outperformed other methods of SLP in generating dynamic sequences of 3D avatars from an unconstrained domain of discourse using a diffusion process on an anatomically informed graph neural network-based on the SMPL-X skeleton [24].

Previous research have had most commonly advanced SLP for D/HH community using cartoon or 3D-rendering based techniques rather than image-based human-like avatars. It is currently unknown if cartoon-based models fully capture the nuances and expressiveness of natural signing. The acceptability and usability of avatar rendering in practical applications for D/HH community remain understudied, leaving questions about expressiveness and user satisfaction unanswered. Our study addressed these gaps by developing a human-like sign language avatar and investigating user preferences to improve SLP technology for DHH community. We used this SLP system to complement the SLR technology.

2.3 Sign Language Recognition and Sign Language Production: System Usability

System Usability refers to the measurement of how easy and efficient a system is to use [40] by users. Usability is crucial for D/HH adoption of SLR and SLP systems to communicate with hearing individuals. Three key factors contribute to usability success:

- **Ease of Use.** SLR and SLP systems require intuitive interfaces with minimal learning curves for effective adoption. Key features include simple navigation, customizable settings, and accessible feedback mechanisms [31]. Systems should accommodate D/HH users' visual and signing communication preferences. Most SLR and SLP assistive technologies for D/HH communication remain in prototype phases [8], with HandTalk being a notable exception, offering text/audio to ASL/Brazilian Sign Language translation [11].
- **Real-time Performance.** Real-time performance is crucial for SLR and SLP systems in D/HH daily communication. These systems must balance minimal latency with maximal accuracy to maintain natural conversation flow and user satisfaction [23]. For example, Jolly et al. [37] found that real-time captioning, despite initial barriers like lag, effectively aided D/HH college students in accessing information and facilitating classroom communication.
- **System Reliability.** Reliable SLR and SLP systems must accurately interpret both from and into ASL. Expert human interpreters achieve similar accuracy in English-to-ASL (72.7%) and ASL-to-English (75.7%) translations [21]. Automated systems should aim for comparable accuracy to ensure user satisfaction and trust. For example, Boudreault et al. [4] demonstrated that feature customizability and placement are crucial for successful closed-interpreting implementations, emphasizing the importance of reliable system performance across diverse communication scenarios.

Few studies have explored the real-world usability of combined SLR and SLP systems, limiting our understanding of their usability in daily contexts. This knowledge gap impedes the development of truly accessible and user-friendly technologies for D/HH individuals. Our study addressed this limitation by conducting comprehensive usability tests of both systems in real-life settings, aiming to enhance the design of SLR and SLP systems for effective daily communication between D/HH and hearing individuals.

3 BI-DIRECTIONAL SYSTEM DESIGN

3.1 Bi-directional System Framework

To address the aforementioned research gaps, we developed a bi-directional system (i.e., combined SLR and SLP) to model real-time communication between D/HH and hearing individuals. Fig. 1 illustrates the system architecture with featuring bi-directional information flow: sign language from D/HH users was recognized and translated for hearing individuals as spoken language, while spoken language was rendered into sign language via an image-based digital human-like avatar. This system enables clear and accurate communication between D/HH and hearing individuals.

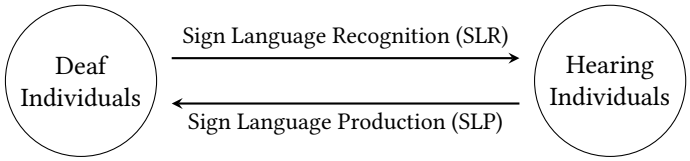

Figure 1: An overview of bi-directional system architecture.

3.2 Bi-directional System Development

3.2.1 Data collection and preparation. We have collected a dataset of ASL signing comprised of sentences and signs paired with the associated English. From this, we proceed with three feature extraction steps:

- **Annotation of Gloss**: We annotated each sequence with the gloss it contains.
- **Annotation of Linguistic Information**: We annotated each gloss and each sequence with linguistic information.
- **Extraction of Low-Dimension Data Representation** We extracted a low-dimensional feature representation from each data point containing pose, resenet features, and cropped areas of interest (face, hands).

3.2.2 Bi-directional system model design. We obtained dataset $D = \{(l, x, g, e, f)_i\}_{i=1}^{N}$ for linguistic, low dimensional representation, pose, gloss, and English features respectively. Examining each space [2]: $l \in L \subset \mathcal{L}^*$ represents a sequence of linguistic information.

[2]for brevity, we abused the Kleeny star operator $*$ to also operate over continuous spaces. $X^* = \bigcup_{i \geq 0} X^i$

$x \in X \subset \mathbb{R}^{K^*}$ represents the low-dimensional data representation, containing a concatenation of extracted pose, extracted resenet, and dimensionality-reduced per-frame cropped regions of interest. The per-frame cropped region of interest is important as motion blur caused by the high-velocity motion in signing frequently prevents the pose prediction model from functioning[3]; $g \in G \subset \mathcal{V}_G^*$ represents the strings of gloss. Note that we primarily relied on L to capture the broad morphological[4] variations of root-signs; $e \in E \subset \mathcal{V}_E^*$ represents strings in the vocabulary of English. Note that these are not words, but rather tokens; and, $f \in F \subset \mathbb{R}^{K \times K^*}$ represents K by K frames.

3.3 Sign Language Recognition System Design

Due to the high dimensional nature of both input space (raw videos and analog), and target space, paired with limited data, we opted to add inductive bias to our models enabling them to converge prior to overfitting. Therefore, we utilized a cascaded series of models which were trained jointly. In particular, we aimed to map low-level frame information to English through a cascade of sub-models. Following our feature extractor, we applied a sign boundary detection model trained to recognize annotated sign boundaries: $s_\zeta : X \to [0,1]^*$. These boundaries were applied softly through an attention map to allow an isolated sign recognition model: $r_\eta : X \times A \to L \times G$ (where A is the set of all cross-attention maps) to independently attend to each sign. Critically, this isolated sign recognition model not only predicted the gloss, but also the raw linguistic phonemes (as this carries inflectional information), which was trained on cropped dataset samples to ensure the model was robust to sign boundary errors and ASL assimilation. The isolated signs were passed through a conditional random field to smooth the output. Finally, the resultant gloss and linguistic information was passed through a Transformer model which translated the results into English. By training within this paradigm, our model was able to achieve a SacreBLEU of 55.1 ± 6.67 on unseen test datapoints (for context, the BLEU of human interpreters were measured to be 27.6 ± 5.26).[5]

3.4 Sign Language Production System Design

We utilized a multi-step approach to generate the human-like avatar for the SLP representation.

3.4.1 Generation of pose information from text. We initially regressed from English to a written sign language analog. This step preceded pose generation, as it reduced output variance, thereby simplifying the subsequent production of motion sequences.

We learned $f_\phi : E \to G \times L$ via f_ϕ (the gloss and linguistic features). Note that we not only regressed to the gloss, as gloss alone typically omits many crucial morphological features. This model can be done via a Seq2Seq model such as a Transformer [47], and trained via cross entropy as both the gloss space and linguistic phonetic space were discrete.

3.4.2 Generation of pose information from written sign analog. Following this, We mapped from our gloss-linguistic sign analog space to pose data using a function $g_\psi : G \times L \to X$, parameterized by a Transformer with ψ. The model outputted pose p_i and a terminator signal T, with generation terminating when $\hat{T} > 0.5$. We employed a binary cross-entropy loss for T and mean squared error for X.

3.4.3 Generation of frames from pose information. The human-like avatar employed a deepfake technique using a conditional GAN [35] that conditioned on pose information. We unrolled frames and extracted features (f, x), then trained the GAN with generator G_ρ and discriminator D_θ, learning the inverse problem to low-level feature extraction. GANs were chosen over diffusion models for their real-time generation capability, as diffusion models' iterative process makes them significantly slower despite superior image fidelity.

We trained SLR and SLP models using an Adam optimizer with a learning rate of 1e-5 (batch size = 128, epochs = 100) on NVIDIA H100 GPUs.

4 DESIGN OF CASE STUDY

This case study comprised two Institutional Review Board (IRB)-exempted hybrid experiments usability tests evaluating the practicality of the combined SLR and SLP system as daily tools for the D/HH community. We aimed to assess the overall systems' effectiveness and user experiences in real-world applications.

Recruitment flyers for Deaf individuals were distributed through various channels (e.g., social media). The inclusion criteria were as follows: D/HH participants should be at least 18 years old, use ASL as their primary language with sufficient proficiency, and should not have any conditions which would impact their ability to interact with our system (e.g. Cerebral Palsy). Interested D/HH individuals contacted the research team for voluntary participation. Participants completed a screening process to determine eligibility. Eligible participants provided informed consent. The Deaf researcher thoroughly explained the study, potential risks, and benefits. Each participant also provided demographic (e.g., age) information and was later compensated with a $25 gift card.

4.1 Usability Test 1: Bi-directional Communication System

An ideal AI-powered automated interpretation system would likely be bi-directional [42], facilitating seamless communication between D/HH and hearing individuals. Until our study, however, limited research had been done to demonstrate this.

The bi-directional system was hypothesized to be ideal for supporting natural, interactive communication with inclusive participation, surpassing open-loop SLR or SLP systems [13]. However, most bi-directional systems have remained in prototype phases with their feasibility for daily closed-loop communication between D/HH and hearing communities largely unexplored. Therefore, Test 1 investigated the technical performance and practical feasibility of our bi-directional system and compared it with: **zero-way System** (text-based (e.g., using phone) communication for both

[3] linguistically, as handshapes are frequently changed during Movement portions (see Movement-Hold model), it is crucial to capture the handshape changes

[4] while this set only captures the phonologic details, as morphemes are representable as sets of sequences of phonemes, it serves dual role to carry morphological data through the system.

[5] our test datapoints were structured so no signer in the test dataset was trained on. In addition, all entries required to have no more than a 4-gram overlap with any sentence in our training dataset.

D/HH and hearing individuals); **one-way system** (avatar rendering for hearing-to-D/HH communication; text-based input for D/HH-to-hearing communication). Given that many D/HH individuals mainly used zero-way system for communication purposes, we used it as the baseline to compare with one-way and bi-directional systems. This helped us determine which communication system could provide a systemic benefit over the current status quo. Therefore, we conducted a usability test to assess communication effectiveness across systems. Participants tested three systems in random order. For each system, participants communicated with a hearing participant on randomly selected topics for 5-8 minutes, followed by a custom e-survey based on System Usability Scale (SUS)[14]. The e-survey comprised six 10-point[6] Likert-scale questions (such as 'This system was efficient.' and 'This system met your communication needs.') assessing efficiency and communication effectiveness, a ranking question for overall preference, and included an open-ended question for qualitative feedback.

To mitigate locality bias, as opinions from the epicenter of the D/HH community may differ from regions with less accessibility we designed a web-based testing interface. The interface was based on Randomized Complete Block Design [46] shown in Fig. 3, containing two distinct components to ensure smooth communication. Toggles on the left side were used to allow participants to swap between systems (B: bi-directional system; O: one-way system; Z: zero-way system) and the e-survey ('S'). For online participants, a video conferencing platform was integrated with the interface to ensure participants and researchers could see each other.

This interface ensured smooth communication between SLR and SLP. For SLR, the interface included a user self-view for positioning adjustment and a multi-state button ("Start", "Stop", "Redo") controlling the recognition engine. A loading icon indicates processing, with results displayed in a text box after 1-2 seconds. Once the text shows up, the user could submit the text by pressing the check mark, edit the text by typing into the text box, or redo their signing by pressing the "redo" button. For SLP, the system automatically triggered the avatar to sign when a hearing person speaks, displaying a "Translating..." icon during the rendering. Unlike D/HH users, hearing individuals have limited control over output, as the system was designed for the interface to face the D/HH users. This design, leveraged well-established speech recognition technology. Both hearing and D/HH users can mute the microphone using the mic button. Some examples of the UI functioning can be seen in supplemental materials.

4.2 Usability Test 2: Human-like Avatar vs. Cartoon Avatar

Previous studies on signing avatar preferences used interpreter videos as proxies for human-like avatars [27], potentially overlooking unique factors such as artifacting and unnatural signing. We aimed to re-examine these findings using an actual human-like avatar. We hypothesized that human-like avatars will be more comprehensible due to their resemblance to everyday D/HH communication. Therefore, we conducted a survey assessing preference scores for different avatar types.

[6] 1: strongly disagree; 10: strongly agree

To explore D/HH individuals' avatar preference, we showed them a cartoon avatar[7] (Fig. 3 (a)) and a Sign-Speak[8] human-like avatar (Fig. 3 (b)). Long (~4 sentences) and short (1 sentence) excerpts were randomly selected from a dataset of 10 each and rendered into signing videos using two distinct avatars. Participants viewed each of the two randomly selected sentences rendered onto each avatar to mitigate order effects. Users provided feedback via a custom questionnaire adapted from the System usability scale (SUS) [14]. For each video, four 6-point[9] Likert-scale questions assessed aspects such as comprehensibility (e.g., The signing in the video was easy to understand) and understandability (i.e., if users could understand the avatar). Participants also ranked their overall signing avatar preference and offered qualitative feedback.

5 RESULTS

Eighteen (Age (years): 30.85 ± 12.54) and forty-eight (Age (years): 30.85 ± 12.54) D/HH individuals participated in Test 1 and 2, respectively. Based on these case studies, we have identified essential guidelines for developing and implementing this technology. Significance level of $p < 0.05$ was applied.

5.1 System Preference

Table 1 shows results of system preference analysis using pairwise one-way binomial tests revealed no significant preference for one-way over zero-way systems, but significant preference for bi-directional system over both zero-way ($p = 0.02$) and one-way ($p < 0.01$) systems. These findings rejected the null hypothesis that favors zero-way over bi-directional systems. Bayesian analysis (Bayes factor with Jeffery's prior) yielded an odds factor of 4.63, indicating moderate evidence that one way systems offer no significant advantage over text-based communication. The proposed bi-directional approach was preferred by 78% of users over the zero-way system, with either one-way system or the bi-directional system being preferred over the zero-way system by 88% of the users. Participants rated the bi-directional system highly for ease of learning (8.83±1.83), meeting communication needs (8.61±1.77), and willingness to use if offered (8.33±2.47)[10].

In summary, we found that 1) **AI communication systems should provide bi-directional support in ASL**; 2) **Merely providing one direction is no better than requiring individuals to write back and forth**.

5.2 Avatar Preference

Participants reported SLP accuracy of 78% ± 1.7% for our system and 61% ± 2.4% for HandTalk, indicating that our SLP is comparable to expert human interpreters ([21] reported an accuracy of 72.7 for expert human interpreters working into ASL). Avatar preference scores were binary, with users choosing exactly one avatar. Each user's rankings were independent and identically distributed.

[7] We used Handtalk's avatar as they are commercially used
[8] A startup whose AI-powered language software recognizes ASL and translates it into spoken words and vice versa
[9] 1: strongly disagree; 6: strongly agree
[10] 1: strongly disagree; 10: strongly agree

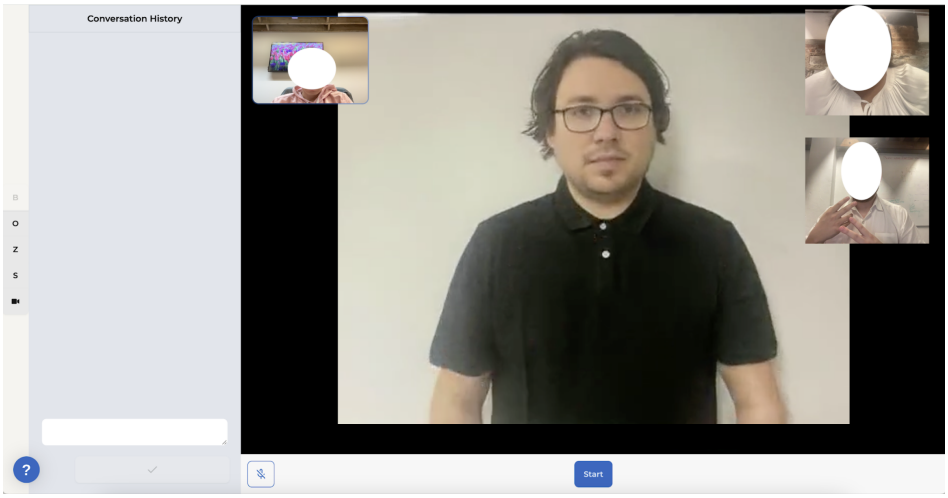

Figure 2: An overview of interface design

Table 1: Results of binomial tests with the alternative hypothesis that the row entry is preferred to the column entry. Statistically significant results are bolded. These results indicate that the two-way system was preferred over both the zero-way and one-way system. Additionally, the zero-way and one-way system were not preferable to any other system.

Row > Column	Zero-way system	One-way system	Bi-directional system
Zero-way system	1.00	0.88	1.00
One-way system	0.24	1.00	1.00
Bi-directional system	**0.02**	**<.01**	1.00

(a) (b)

Figure 3: Demonstration of avatar style: (a) cartoon, (b) human. The human in (b) consented and was face-swapped with a non-existent person.

Consequently, we performed a binomial test with the alternative hypothesis that users favored the human-like avatar. The test resulted in $p < 0.01$, allowing us to reject the null hypothesis. We conclude that human-like avatars are statistically significantly preferred over cartoon avatars. Therefore, we assert that **AI avatars SHOULD be human-like to be broadly accepted by D/HH community**.

6 CONCLUSIONS AND DISCUSSION

This pilot case study designed and evaluated a bi-directional communication system, demonstrating its superiority over zero-way and one-way systems through two usability tests. Results revealed D/HH participants' preference for human-like avatars over cartoon avatars. Our findings have significant implications for AI-powered interpretation systems bridging communication gaps between D/HH and hearing individuals. For example, we found it is significant for prioritizing bi-directional communication capabilities and enhancing SLR technologies for natural ASL interpretation. Integrating and improving human-like avatars in SLP systems is crucial for better replicating ASL intricacies. Adopting user-centered design approaches, involving D/HH individuals throughout development, ensures technologies meet specific needs and preferences. This understanding matches the U.S. Disability Advisory Committee's opinions: "without the ability to have other participants' audio converted to sign language and to have their own sign language converted to speech, a person who is Deaf or Hard of Hearing [...] may not be able to effectively participate in video conferences [or conversations]" [7]. Their statement encourages innovative efforts to include sign language services to meet the needs and requirements of daily life. Though some preexisting approaches to close the communication gap between D/HH and hearing people hold promise, it is clear that SLR or SLP alone will fail to do so effectively. To fully bridge the communication gap, a bi-directional system with a human-like avatar is essential.

We found feedback from D/HH individuals throughout the usability test to be invaluable, and highly encouraged such consultations to become industry standard. Deaf leadership is additionally paramount when conducting these studies as collecting raw feedback

obtained in this study was only possible due to involvement from our Deaf researcher.

To the best of our knowledge, this case study was the first work to answer usability questions for D/HH community using a bi-directional communication system. Our contributions included the development of an automated sign language interpretation system, evaluating and comparing design strategies to determine optimal approaches, and providing evidence-based recommendations for the design and implementation of effective automated interpretation systems. These insights aim to guide future D/HH accessibility innovation efforts, emphasizing the importance of designing with, rather than for, the D/HH community.

While our study offers valuable findings, it has limitations. The studies focused on short-term interactions and immediate user preferences rather than long-term usability or sustained impact. Long-term engagement with the system might reveal different usability challenges, learning curves, evolving preferences, and the extent to which latency affects practical use in real-world scenarios. In addition, system latency (< 5s) may also have impacted user satisfaction, causing deployment challenges such as handling large user bases and cross-platform compatibility.

Our advancements showed promise in enhancing service access, workforce participation, and educational opportunities for the D/HH community, marking a significant step towards increased autonomy across all societal aspects. For example, 1) **Service Access**: enable private communication with healthcare providers or customer service without interpreters; 2) **Education**: facilitate equal access to lectures through real-time ASL interpretations and responses; 3) **Emergency Communication**: provide life-saving communication with first responders during crises.

Acknowledgments

We would like to extend our gratitude to Dr. Scot Atkins (associate professor at National Technical Institute for the Deaf) and Dennis Murphy (graduate student research fellow at Georgia Institute of Technology) for their contributions to this case study. Their insights and support were helpful throughout the development of this study.

References

[1] Wenda Alifulloh, Dadang Juandi, Aan Hasanah, Dian Usdiyana, and Humam Nuralam. 2024. Research Trends of Mathematics Learning for Deaf Junior High School Students in Indonesia: A Systematic Literature Review. *The Eurasia Proceedings of Educational and Social Sciences* (2024), 24–33.

[2] Maryam Aziz and Achraf Othman. 2023. Evolution and Trends in Sign Language Avatar Systems: Unveiling a 40-Year Journey via Systematic Review. *Multimodal Technologies and Interaction* 7, 10 (2023), 97.

[3] Vasileios Baltatzis, Rolandos Alexandros Potamias, Evangelos Ververas, Guanxiong Sun, Jiankang Deng, and Stefanos Zafeiriou. 2024. Neural Sign Actors: A diffusion model for 3D sign language production from text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1985–1995.

[4] Patrick Boudreault, Muhammad Abubakar, Andrew Duran, Bridget Lam, Zehui Liu, Christian Vogler, and Raja Kushalnagar. 2024. Closed Sign Language Interpreting: A Usability Study. In *International Conference on Computers Helping People with Special Needs*. Springer, 42–49.

[5] National Deaf Center. 2024. *Deaf Awareness.* https://nationaldeafcenter.org/resources/deaf-awareness/

[6] Xingyu Chen, Baoyuan Wang, and Heung-Yeung Shum. 2023. Hand avatar: Free-pose hand animation and rendering from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8683–8693.

[7] Federal Communication Commission. 2014. *Disability Advisory Committee.* https://www.fcc.gov/disability-advisory-committee

[8] Vernandi Dyzel, Rony Oosterom-Calo, Mijkje Worm, and Paula S Sterkenburg. 2020. Assistive technology to promote communication and social interaction for people with deafblindness: a systematic review. In *Frontiers in Education*, Vol. 5. Frontiers Media SA, 578389.

[9] Susan R Easterbrooks, Amy R Lederberg, Shirin Antia, Brenda Schick, Poorna Kushalnagar, Mi-Young Webb, Lee Branum-Martin, and Carol McDonald Connor. 2015. Reading among diverse DHH learners: What, how, and for whom? *American Annals of the Deaf* 159, 5 (2015), 419–432.

[10] Center for Research on Disability. 2022. *Civilians Living in the Community for the United States and States - Hearing Disability: 2022.* https://www.researchondisability.org/ADSC/build-your-own-statistics

[11] Handtalk. 2012. *Discover the largest Sign Language translation platform in the world.* https://www.handtalk.me/en/

[12] Liwen Hu, Shunsuke Saito, Lingyu Wei, Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. 2017. Avatar digitization from a single image for real-time rendering. *ACM Transactions on Graphics (ToG)* 36, 6 (2017), 1–14.

[13] Anjali Kanvinde, Abhishek Revadekar, Mahesh Tamse, Dhananjay R Kalbande, and Nida Bakereywala. 2021. Bidirectional sign language translation. In *2021 International Conference on Communication Information and Computing Technology (ICCICT)*. IEEE, 1–5.

[14] Aycan Kaya, Reha Ozturk, and Cigdem Altin Gumussoy. 2019. Usability measurement of mobile applications with system usability scale (SUS). In *Industrial Engineering in the Big Data Era: Selected Papers from the Global Joint Conference on Industrial Engineering and Its Application Areas, GJCIE 2018, June 21–22, 2018, Nevsehir, Turkey*. Springer, 389–400.

[15] Janis Kritzinger, Marguerite Schneider, Leslie Swartz, and Stine Hellum Braathen. 2014. "I just answer 'yes' to everything they say": Access to health care for deaf people in Worcester, South Africa and the politics of exclusion. *Patient Education and Counseling* 94, 3 (2014), 379–383. https://doi.org/10.1016/j.pec.2013.12.006

[16] Kit Yung Lam, Liang Yang, Ahmad Alhilal, Lik-Hang Lee, Gareth Tyson, and Pan Hui. 2022. Human-avatar interaction in metaverse: Framework for full-body interaction. In *Proceedings of the 4th ACM International Conference on Multimedia in Asia*. 1–7.

[17] Amy Lederberg. 2022. Special Education Research and Development Center on Reading Instruction for Deaf and Hard of Hearing Students. *https://ies.ed.gov/ncser/RandD/details.asp?ID=1325* (2022).

[18] Carman KM Lee, Kam KH Ng, Chun-Hsien Chen, Henry CW Lau, Sui Ying Chung, and Tiffany Tsoi. 2021. American sign language recognition and training method with recurrent neural network. *Expert Systems with Applications* 167 (2021), 114403.

[19] Kezhou Lin, Xiaohan Wang, Linchao Zhu, Ke Sun, Bang Zhang, and Yi Yang. 2023. Gloss-free end-to-end sign language translation. *arXiv preprint arXiv:2305.12876* (2023).

[20] Ross E Mitchell and Travas A Young. 2023. How many people use sign language? A national health survey-based estimate. *Journal of Deaf Studies and Deaf Education* 28, 1 (2023), 1–6.

[21] Brenda Nicodemus and Karen Emmorey. 2015. Directionality in ASL-English interpreting: Accuracy and articulation quality in L1 and L2. *Interpreting* 17, 2 (2015), 145–166.

[22] World Health Organization. 2024. *Deafness and hearing loss.* https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[23] Maria Papatsimouli, Panos Sarigiannidis, and George F Fragulis. 2023. A survey of advancements in real-time sign language translators: integration with IoT technology. *Technologies* 11, 4 (2023), 83.

[24] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. 2019. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10975–10985.

[25] Varshini Prakash and BK Tripathy. 2020. Recent advancements in automatic sign language recognition (SLR). In *Computational Intelligence for Human Action Recognition*. Chapman and Hall/CRC, 1–24.

[26] Lorna Quandt. 2020. Teaching ASL signs using signing avatars and immersive learning in virtual reality. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

[27] Lorna C. Quandt, Athena Willis, Melody Schwenk, Kaitlyn Weeks, and Ruthie Ferster. 2022. Attitudes Toward Signing Avatars Vary Depending on Hearing Status, Age of Signed Language Acquisition, and Avatar Type. *Frontiers in Psychology* 13 (2022). https://doi.org/10.3389/fpsyg.2022.730917

[28] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2020. Hand sign language recognition using multi-view hand skeleton. *Expert Systems with Applications* 150 (2020), 113336.

[29] Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2022. Real-time isolated hand sign language recognition using deep networks and SVD. *Journal of Ambient Intelligence and Humanized Computing* 13, 1 (2022), 591–611.

[30] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera, and Mohammad Sabokrou. 2021. Sign language production: A review. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3451–3461.

[31] Amit Shankar and Biplab Datta. 2020. Measuring e-service quality: a review of literature. *International Journal of Services Technology and Management* 26, 1

(2020), 77–100.

[32] Shikhar Sharma and Krishan Kumar. 2021. ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks. *Multimedia Tools and Applications* 80, 17 (2021), 26319–26331.

[33] Kaiyue Shen, Chen Guo, Manuel Kaufmann, Juan Jose Zarate, Julien Valentin, Jie Song, and Otmar Hilliges. 2023. X-avatar: Expressive human avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16911–16921.

[34] Bowen Shi, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2022. Open-domain sign language translation learned from online video. *arXiv preprint arXiv:2205.12870* (2022).

[35] Kaleb E Smith and Anthony O Smith. 2020. Conditional GAN for timeseries generation. *arXiv preprint arXiv:2006.16477* (2020).

[36] William C Stokoe Jr. 2005. Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of deaf studies and deaf education* 10, 1 (2005), 3–37.

[37] Fatma M Talaat, Walid El-Shafai, Naglaa F Soliman, Abeer D Algarni, Fathi E Abd El-Samie, and Ali I Siam. 2024. Real-time Arabic avatar for deaf-mute communication enabled by deep learning sign language translation. *Computers and Electrical Engineering* 119 (2024), 109475.

[38] Nina Tran, Paige S DeVries, Matthew Seita, Raja Kushalnagar, Abraham Glasser, and Christian Vogler. 2024. Assessment of Sign Language-Based versus Touch-Based Input for Deaf Users Interacting with Intelligent Personal Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[39] DEAF SERVICES UNLIMITED. 2023. *ASL Interpreter Shortage and Accessibility in Higher Education.* https://deafservicesunlimited.com/asl-interpreter-shortage-and-accessibility-in-higher-education/

[40] Prokopia Vlachogianni and Nikolaos Tselios. 2022. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education* 54, 3 (2022), 392–409.

[41] Rosalee Wolfe, John C McDonald, Thomas Hanke, Sarah Ebling, Davy Van Landuyt, Frankie Picron, Verena Krausneker, Eleni Efthimiou, Evita Fotinea, and Annelies Braffort. 2022. Sign language avatars: a question of representation. *Information* 13, 4 (2022), 206.

[42] Qinkun Xiao, Lu Li, and Yilin Zhu. [n. d.]. Bidirectional Sign Language Communication Based on Variational and Adversarial Learning. *Available at SSRN 4383371* ([n. d.]).

[43] Fei Xu, Lipisha Chaudhary, Lu Dong, Srirangaraj Setlur, Venu Govindaraju, and Ifeoma Nwogu. 2024. A Comparative Study of Video-Based Human Representations for American Sign Language Alphabet Generation. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–6.

[44] Sicheng Xu, Guojun Chen, Yu-Xiao Guo, Jiaolong Yang, Chong Li, Zhenyu Zang, Yizhong Zhang, Xin Tong, and Baining Guo. 2024. Vasa-1: Lifelike audio-driven talking faces generated in real time. *arXiv preprint arXiv:2404.10667* (2024).

[45] Christine Yoshinaga-Itano. 2014. Principles and guidelines for early intervention after confirmation that a child is deaf or hard of hearing. *Journal of deaf studies and deaf education* 19, 2 (2014), 143–175.

[46] Emily C Zabor, Alexander M Kaizer, and Brian P Hobbs. 2020. Randomized controlled trials. *Chest* 158, 1 (2020), S79–S87.

[47] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. 2021. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 6881–6890.