Author's Pre-Print Copy

American Sign Language Video Anonymization to Support Online Participation of Deaf and Hard of Hearing Users

SOOYEON LEE, Rochester Institute of Technology, USA
ABRAHAM GLASSER, Rochester Institute of Technology, USA
BECCA DINGMAN, Rochester Institute of Technology, USA
ZHAOYANG XIA, Rutgers The State University of New Jersey, USA
DIMITRIS METAXAS, Rutgers The State University of New Jersey, USA
CAROL NEIDLE, Boston University, USA
MATT HUENERFAUTH, Rochester Institute of Technology, USA

Without a commonly accepted writing system for American Sign Language (ASL), Deaf or Hard of Hearing (DHH) ASL signers who wish to express opinions or ask questions online must post a video of their signing, if they prefer not to use written English, a language in which they may feel less proficient. Since the face conveys essential linguistic meaning, the face cannot simply be removed from the video in order to preserve anonymity. Thus, DHH ASL signers cannot easily discuss sensitive, personal, or controversial topics in their primary language, limiting engagement in online debate or inquiries about health or legal issues. We explored several recent attempts to address this problem through development of "face swap" technologies to automatically disguise the face in videos while preserving essential facial expressions and natural human appearance. We presented several prototypes to DHH ASL signers (N=16) and examined their interests in and requirements for such technology. After viewing transformed videos of other signers and of themselves, participants evaluated the understandability, naturalness of appearance, and degree of anonymity protection of these technologies. Our study revealed users' perception of key trade-offs among these three dimensions, factors that contribute to each, and their views on transformation options enabled by this technology, for use in various contexts. Our findings guide future designers of this technology and inform selection of applications and design features.

CCS Concepts: • **Human-centered computing → Accessibility design and evaluation methods**; **Empirical studies in accessibility**.

Additional Key Words and Phrases: Deaf and Hard of Hearing, Anonymization

ACM Reference Format:
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 *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3441852.3471200

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
Manuscript submitted to ACM

Author's Pre-Print Copy

1 INTRODUCTION AND MOTIVATION

Our research evaluates Deaf and Hard of Hearing (DHH) users' interest, preferences, and concerns in relation to a prototype for anonymization of sign language video communications, to provide guidance for designers of this technology. While there are many sign languages, our work focuses on American Sign Language (ASL), used by over 500,000 people in the U.S. [34]. Although often used in countries in which English is spoken, ASL is a language distinct from English, produced by movements of the face, head, hands, and torso [4, 11, 35, 47, 55].

Being able to communicate anonymously in one's preferred language is essential for participating in a variety of social, professional, and societal contexts. Some prior work [3, 33] has focused on techniques to hide the face of a user for privacy protection in circumstances where this may be important. For instance, Internet users may visit discussion boards to ask questions about sensitive topics; individuals may express dissenting political or religious views that could subject them to persecution; or essential professional activities like academic peer-review may require anonymity. While it is relatively straightforward for users of written languages to engage in anonymous written communications online, such options have not been available for users of sign languages. These languages generally lack a written form in common use among the language community, and therefore video-based communication, which reveals the face, is necessary.

While users of spoken language can hide their face on online video-sharing platforms [17, 24, 46, 52, 59], this option is not available to ASL users, as the face conveys essential linguistic information [4, 11, 27, 35, 55]. Barriers to private communication in one's primary language limit online debate or enquiries, e.g., in relation to sensitive topics, such as reproductive health, domestic abuse, or substance abuse, which prior research has revealed to have higher prevalence in the DHH community [5, 42]. Anonymizing the face, while retaining the key linguistic information it conveys, would also enable peer review of academic publications in sign language, conformity in appearance when multiple individuals contribute to a composite video or collection (e.g., entries in a video ASL dictionary), and privacy protection when users contribute videos to ASL datasets for AI research – applications discussed in [7, 32].

Over the past decade, real-time tools for face transformations have become popular among consumers, e.g., to make someone appear to be wearing makeup [25] or overlay a virtual cute animal mask [61]. More recently, AI technologies for real-time face transformation (sophisticated technologies that preserve facial expressions) have matured and become available to non-technical users for producing realistic videos in which a synthetically generated human face in a video is driven by the face of another person. As compared to earlier face-filter technologies (simplistic technologies that do not preserving facial expressions), these advancements enable new applications for DHH ASL users, as it is now possible to replace the face while preserving detailed facial expressions and head movements.

In this research, we conducted an interview study to evaluate prototype face-disguise technology (a generic term for technologies that obscure the face) applied to videos of human ASL signers, influenced by recent image-to-video technology [44, 45, 50, 60], for replacing the face in a video with a new face from a given photograph, preserving facial expressions and head movements. In one prototype variation, the torso of the human remains in the video, and in another, the torso is hidden to disguise the clothing and body for further obscuring the identity of the signer. For comparison, we also evaluated a simpler face-filter with a virtual cartoon-like Tiger mask, previously evaluated in [7]. In a 70-minute appointment, participants: (1) viewed disguised videos and attempted to identify the person in the original video from a line-up of photos, (2) viewed original and disguised videos processed by prototype variations, and provided subjective feedback about each, and (3) viewed videos of themselves transformed by this technology. In a

Author's Pre-Print Copy

semi-structured interview, participants discussed their views of the technology, preferences among appearance options, factors affecting acceptability, potential uses, and concerns.

The contributions of this work are empirical and include: (1) The first evaluation with DHH ASL users of modern face-transformation technology, capable of preserving ASL linguistic facial expressions, revealing its effectiveness at preserving anonymity; (2) Quantitative and qualitative evaluation of understandability, naturalness, and anonymity-preservation, to compare prototypes varying in their appearance transformations; (3) Evidence of users' views on the acceptability of this technology, its potential uses, and their concerns; (4) Identification of users' perceived trade-offs among understandability, naturalness, and anonymity protection, with design considerations from our analysis; (5) Evidence of ways in which preservation and transformation of identity relate to users' acceptance of this technology.

2 PRIOR WORK

2.1 Existing Methods of Conveying ASL Anonymously

While researchers have acknowledged the importance of enabling deaf signers to communicate anonymously online [15, 16], most prior efforts to address this problem have aimed to produce artificial writing systems for sign language or to create tools to allow deaf signers to create their own animations of a virtual human signing their message. Despite efforts to invent sign language writing systems, e.g., [2, 39, 48] or related technologies [8], no writing system has yet gained widespread popularity within the DHH community. Thus, written communication in ASL is not practical for enabling signers to communicate without revealing their identity.

Other work seeks to enable users to create synthetic animations of sign languages, which could, in principle, produce anonymous messages. Prior sign language animation research has largely focused on machine-translation contexts [6], but some work examines how to enable users to script the movements of virtual humans to perform sign language, e.g., [13, 21]. Unfortunately, existing tools are not yet sufficiently expressive to produce clear virtual animation, nor are the tools and techniques for building novel animated messages likely to become simple enough for use by non-experts, despite recent efforts [1, 56]. In summary, despite work on writing systems and avatar technologies, no existing approaches yet provide a satisfactory solution to the challenge of anonymous communication in sign language.

2.2 Accessibility of Written/Spoken and Sign Language Online Content Creation

Prior work has examined DHH users' interests, current practices, and barriers, in relation to producing content to share online, e.g., [10, 14, 15, 23] or in the context of social media interaction [32]. When privacy is a concern, DHH users must currently use written English to prepare online messages or content. Given the diversity in written-language literacy levels among DHH individuals [53] and the preference of many DHH users for communication in ASL, DHH users face barriers to online participation [32], if they wish to preserve their anonymity during interactions. This is an inequitable situation, as hearing individuals can express themselves online much more easily, in written or spoken form (assuming that their voice is not recognizable and their face is disguised).

Prior work has revealed particular challenges for users who prefer to produce content in sign language, as they must create and post a video of themselves, with their faces and physical appearance visible to whoever watches the video. Recent research [32] has highlighted challenges that DHH ASL signers face in participating in social media sites by recording and sharing ASL video. As reported in [32], the need to hold the phone with one hand (e.g., while standing) in order to record themselves leaves only one hand for signing, which is not ideal, because signing in ASL normally requires two hands. Adding text captions to videos to enable them to be understood by individuals who do not

Author's Pre-Print Copy

know ASL is also time-consuming. The authors provided potential solutions for these challenges, such as incorporating automatic captioning into social media platforms. While [32] focused on barriers to communication on social media platforms, our work focuses on preserving DHH individuals' privacy in video communication. In summary, prior work has revealed that there is strong interest among DHH users for technologies that could facilitate ASL-based communication online, especially in a manner that is privacy preserving; yet existing technologies are not providing an adequate solution to this challenge.

2.3 Video de-identification for privacy in video sharing sites

Some recent work has investigated face-disguise technology for motivating ASL signers to feel comfortable sharing videos in public ASL datasets for research [7]. This study is of particular interest, in that participants were asked questions about their interest in and impressions of face-disguise technology – albeit within this specific context of contributing to a dataset. Participants were able to see their own video transformed through some simple face-filter technology, including a filter that overlaid a cartoon tiger face on top of the signer's face without preservation of any facial expressions, aside from the degree to which the mouth opens. Participants were more willing to share their video publicly with filters mitigating privacy concerns, yet they were dissatisfied with the fact that the filters did not preserve facial expressions.

In the video/photo sharing context, trade-offs between the utility of the anonymized video/photo and privacy protection have also been investigated [18–20, 29]. Prior work has studied how the level of obfuscation from various image filtering techniques (e.g., blurring, pixelization, masking) affects the viewer's experience and the utility of the video/image for specific tasks, e.g., patient training video in a clinical setting [18]. As found in prior work [7], obfuscation from some common privacy enhancing techniques does not satisfy ASL signers because facial expressions are not preserved. Prior research suggests that providing adequate privacy protection for various contexts and uses requires careful selection of the relationship between the level (ranging from no recognition to full recognition) and the types (e.g., blurred, masking, face disguise) of anonymization. Focusing specifically on DHH signers, our study differs from prior work in two ways: (a) We investigate more advanced face-transformation technologies capable of preserving facial expressions; and (b) We investigate these technologies for preserving privacy in ASL videos for a wider variety of uses and contexts, e.g., participation on social media platforms.

In recent years, there has been tremendous progress in technologies for analyzing and synthesizing video of human faces, e.g., [3, 40, 49, 50, 60], with new applications in smart home technologies [54], health [12, 22], and other fields. Another key application of this technology has been for de-identifying videos in order to preserve privacy, e.g., [3, 33]. While most work has focused on technical details and performance of this technology, some researchers have conducted research with human participants to understand their interests in or concerns about this technology. Advances in this technology have led to recent public awareness of "deep fake" technologies for producing seemingly realistic videos of humans, in which the movement of the face is based on the performance of a human in an original video. The ease of creating videos that impersonate someone, making it appear that they are saying or doing things that they had never said or done, has raised significant ethical concerns [28, 43].

Given the complex face and head movements used in ASL for a variety of linguistic purposes, e.g., involving subtle movements of the eyebrows or head [4, 11, 35], there has been a question as to whether the resulting video would sufficiently preserve these key linguistic elements of the performance. Some researchers have begun to design face-disguise technology with a particular focus on preserving such elements of the performance [44, 45, 50, 60], necessary for applying this technology to sign language videos. However, there is a need for empirical research with DHH ASL

Author's Pre-Print Copy

signers, to understand the performance of this technology, as well as users' impressions and judgments of its suitability for the task of anonymizing ASL videos to be shared online.

3 RESEARCH GOALS AND METHODS

Emerging face-transformation technology has the potential to create realistic videos with new faces; yet prior work has revealed ethical concerns with the use of such technology. While some research has examined DHH users' interest in simple face-filter technologies for specific contexts, no prior study with DHH users has investigated state-of-the-art face-disguise technology capable of preserving facial expressions and natural human appearance for sign language video. As these new technological capabilities emerge, it is important to understand DHH users' interest in and impressions of this technology for protecting anonymity, including users' views of various dimensions of system performance, e.g., understandability and naturalness of appearance. The goal of this research is to guide the development of ASL-optimized face technology and inform designers of future applications for these users.

We conducted an interview-based study with 16 DHH individuals who reported using ASL on a daily basis; each participated in a 70-minute Zoom teleconference meeting with a DHH ASL-signing researcher. In this IRB-approved study, the participants were shown examples of videos of ASL signing processed by prototype face-transformation technology (section 3.1). Prior to transformation, some of these videos had been of the participant, submitted to us in advance of the appointment, and some were of other ASL signers from a public research dataset of ASL signing. The interview was conducted entirely in ASL, while the researchers typed notes in English. Participants were asked a mixture of open- and closed-ended questions about their subjective impression of the videos, especially in regard to how well they preserve anonymity, their understandability, and other factors, as described in section 3.2.

3.1 Anonymization Technology Prototypes

In this study, we compare multiple prototype technologies for disguising the face of an ASL signer. We refer to our first prototype as **tiger-face**, a simple video filter technology, similar to those used in SnapChat, in which a 3D mask is virtually overlaid on the face in the video. **Our rationale** for selecting this prototype is three-fold: (a) It reflects the state-of-the-art of consumer-grade face technologies popular during the 2010s; (b) The specific filter was used in a prior study that had examined DHH users' interests in using filters to hide their identity [7], the open-source tiger-face filter from Jeeliz [26]; and (c) It also provides a baseline point-of-comparison for participants, to determine whether the more computationally intensive facial-expression-preserving transformations were useful. The filter detects the human's face and overlays an animated tiger avatar head, which emits blue bubbles from its mouth, triggered whenever the human's mouth opens. Participants in that prior study commented on the limitations of this filter, which does not preserve any other facial expression details, e.g., eyebrow movements, despite this being linguistically important in ASL. We included tiger-face in our study as a baseline for comparison, reflective of the prior state-of-the-art for available face-disguise technologies.

Our prototype, **with-torso**, is based on recent work on image-to-video transformation and video editing, to enable the replacement of the underlying facial geometry, while preserving the linguistically significant facial expressions [44, 45, 50, 60]. **The rationale** for including this transformation in our study was that it reflects a state-of-the-art facial image animation and transformation technology. This specific technology was selected because of its ability to animate face images based on image-to-video transformation, to enable the replacement of the underlying facial geometry by editing the latent facial representations [51, 57]. The torso and background of the signer are not touched or modified in any way. Colloquially, we may refer to the face of the signer being "swapped" with a different human face, based on

Author's Pre-Print Copy

an input photograph of the desired "target face." However, the resulting output video actually appears as a blend of the facial structure of the original signer and those of the individual pictured in the "target face," resulting in a novel composite face that mimics the head movements and facial expressions of the original signer. Sample images of the output of this transformation are shown in Figure 2, and the electronic supplementary files provided with this paper include video samples.

The third prototype, **without-torso**, is identical to the with-torso prototype, except that the signer's torso and the background are both replaced by a flat gray color, as shown in Figure 2. **The rationale** for including this transformation is that identity may be revealed not only by the face, but also by body appearance, clothing, or background, especially if the person viewing the video is familiar with the person in the video.

For both with-torso and without-torso, the resulting output can be varied, by selection of different "target faces," and throughout our study we displayed videos based on a variety of target faces, selected from the Chicago Faces Dataset [30, 31]. We took into account the gender and race/ethnicity of the person in the original video, and we selected target faces of other people with corresponding demographic characteristics – with variation in age, hair style, and hair color. **The rationale** for selecting these variations was that they reflect common options for the selection of video-game avatars or personalized emojis on social media, and several pilot interviews with DHH ASL signers prior to our study revealed their interest in such options. More details about the transformations used in the separate phases of this study are described below. Figure 2 shows screenshots from a few videos and their transformations.

3.2 Study Design

The 70-minute appointment was temporally partitioned into three phases, for participation in three different activities. During each phase, the participant viewed the videos and then answered semi-structured interview questions. In the first phase, we evaluated face disguise technology from the perspective of participants' seeing a disguised video of other people. In the next phase, the understandability, naturalness, and anonymity protection of the transformed videos were assessed, with participants viewing a variety of face-disguise options. (Prior to the main study, we had conducted pilot interview studies with DHH participants to ask them about their interest in technologies for disguising the face, and this had suggested that understandability, naturalness, and anonymity may be key issues for users, which helped us in finalizing the design of our interview questions for this phase.) In the final phase, participants saw themselves disguised, and they commented on the acceptability of the transformed videos and shared other concerns.

3.3 Phase 1 of the Study

The first phase focused on evaluating **how effectively videos had been disguised** by the with-torso and without-torso software; participants were asked to attempt to identify the original person in the video. The source videos used in this phase of the study were from the Boston University American Sign Language Linguistic Research Project [36–38]. To produce a variety of videos, we selected two videos of a male signer and two of a female signer from this dataset; in each video, the signer produces 1-2 ASL sentences. Next, we processed the videos using each of the two prototypes, with-torso and without-torso, using two different "target faces" for each (two male target faces for the male signer, and two target faces for the female signer). Overall, this yielded 16 disguised output videos.

Each participant viewed one disguised video of the male signer, and one disguised video of the female signer. One video was processed using the with-torso prototype, and the other, using the without-torso prototype. The order in which these stimuli were shown to participants, and the assignment of prototype-condition to each gender, were counterbalanced via Latin square. After viewing each video, participants were shown a line-up of six different faces,

Author's Pre-Print Copy

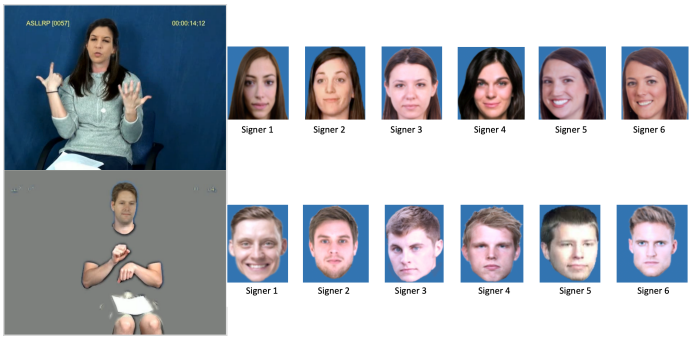

Fig. 1. Disguised videos shown in phase 1, along with line-up photos including the actual signer and other face images selected with similar hair and skin color; to measure the effectiveness of the disguise, participants were asked to guess the correct face.

one of which was the true face of the ASL signer in the anonymized videos. The order in which these line-up faces were shown to the participants was also counterbalanced via Latin square. Figure 1 shows example line-up photos for both the male and female faces. After participants guessed which face was the original person in the video, they indicated their agreement with the Likert item: "It was very difficult to guess the original signer." Phase 1 concluded with questions about participants' opinions of the videos and their difficulty in guessing the signer, including whether seeing the original signer's body and background made it easier to guess the original signer's face.

3.4 Phase 2 of the Study

The second phase focused on the **understandability, naturalness, and anonymity-protection** of videos from all three prototypes, including with-torso and without-torso videos based on a variety of target faces, as well as the tiger-face prototype. In this phase, each participant viewed a total of 34 videos, half based on a source video from a male signer from [36–38], and half from a female signer from the same dataset. For each signer, participants were shown an original, unmodified video, followed by 16 transformed videos associated with that source video. The 16 transformed videos consisted of several sets, each of which focused on one appearance characteristic that varied within each set:

- age (3 videos; based on a young, middle, and older-aged target face),
- artificially colored hair (3 videos; blue, pink, and green colored hair),
- natural-colored hair (3 videos; light, medium, and dark shades),
- with-torso (2 videos with the torso visible – all the others had the torso removed), and
- tiger-face (1 video shown with an animated cartoon tiger face, as used in [7]).

The order of these sets was counterbalanced between participants, and whether male or female videos were shown first was also counterbalanced. After the first with-torso video was shown, the researcher on the video call interrupted the participant to ask the participant to indicate agreement with each of three Likert items, "This video was completely understandable," "This video was very natural in appearance," and "This video disguised the identity of the original signer completely." Similarly, as soon as the first without-torso video was shown, and immediately after the tiger-face was shown, the participant was asked these same three questions. After the participant viewed all videos in this phase, semi-structured open-ended interview questions were asked about the overall understandability, naturalness, and anonymity-protection of the transformations.

Author's Pre-Print Copy

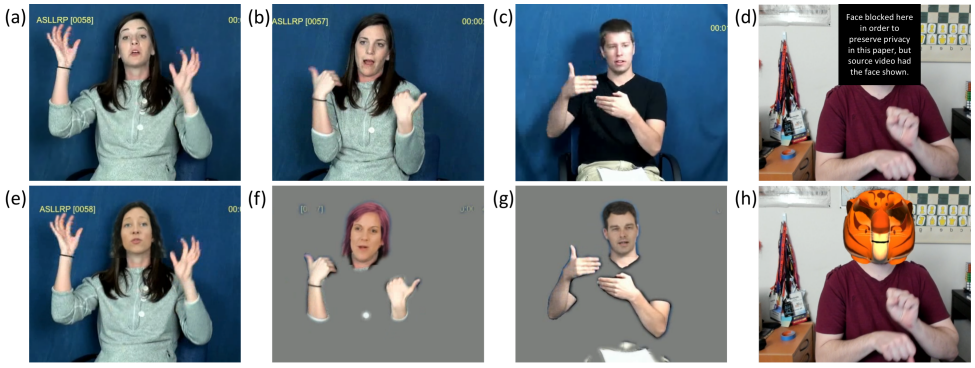

Fig. 2. Sample of videos shown in phase 2: (a-d) source videos and (e-h) transformed videos below corresponding source, e.g., (a) transformed to (e). Samples include: (e) with-torso, (f-g) without-torso, and (h) tiger-face. Source videos (a-c) from [36, 38] and (d) illustrates the type of videos participants submitted (blocked here for anonymity). Videos are in electronic supplementary files.

3.5 Phase 3 of the Study

In the third phase, participants saw a video of themselves transformed using all three prototypes so that we could evaluate their view of **how acceptable this technology is for disguising their own videos**. While the tiger-face prototype could run in real time, the with-torso and without-torso prototypes required additional processing time. Thus, prior to the appointment, we asked participants to submit a video of themselves signing a short ASL passage. Because of limitations in the anonymization prototype and in order to ensure good-quality output, participants were instructed to make sure they had good lighting and a plain background, and they were asked to pull shoulder-length or longer hair back in a ponytail. Participants were also asked to remove any glasses, headgear, and hand jewelry. Lastly, participants were asked to sign in a manner that avoids having their hands obstruct their face, as the prototype system is not robust to face occlusions. For this reason, signers were given an ASL gloss script for a specific passage to perform that excluded signs in which the hands would come close to the face, while also requiring the grammatical use of several facial expressions in ASL: "BOOK, I BUY. TODAY, YOU BORROW. BOOK, READ YOU? BOOK WHERE?"

During phase 3 of the appointment, participants viewed 13 transformed videos, based on the video they had submitted. Six were with-torso, with another six without-torso, using the same set of target faces. The target-face set was matched to the participants' self-reported gender and apparent race/ethnicity in their submitted videos. The 13th video was a live demo website with the tiger-face effect, which participants were instructed on how to use.

After viewing all videos, participants responded to open-ended questions about their perception of and preference among the videos, whether they thought the quality of these videos was good enough for them to consider using software like this, and whether it would be helpful for them to have software that could anonymize videos. Finally, participants were asked what situations they would or would not use this software for, whether they thought it would be acceptable for other people to use software like this, and whether they had concerns about software like this.

3.6 Participants

Via social-media postings, we recruited 16 DHH adults who use ASL on a daily basis; 12 indicated that ASL was their primary language. Four participants had used ASL since birth, 6 learned ASL by age 5, and 6 learned ASL during their late teens (with all in this latter group having used ASL for at least 8 years). Participants' ages ranged from 19 to 47 years

Author's Pre-Print Copy

old (median 27.5). Eight self-identified as male, 1 as non-binary, and 7 as female. Participants' education levels varied: 1 had some undergraduate education, 1 had an associate's degree, 10 had a bachelor's degree, and 4 had a master's degree. Eight self-identified as Caucasian, 1 as Black, 3 as Asian, 1 as Vietnamese, 1 as Latino, 1 as Asian & Hispanic, and 1 as Spanish & Native American. A demographics table appears in electronic supplementary files.

3.7 Data Analysis

All data collected from the three phases of studies were analyzed with both quantitative and qualitative approaches. We conducted statistical analysis with Friedman tests on the quantitative data, and we performed an iterative thematic analysis [9] on our qualitative data, employing both deductive and inductive approaches. We manually developed a deductive coding framework with the main topics of our interview questions. In the framework, we aggregated all the data and iteratively performed open coding using colors. Then codes were generated with the color-coded data and organized with categorization. Finally, main and sub-themes were identified and developed using a bottom-up approach. We went through the same process with the data from all three phases of the study.

4 FINDINGS

To investigate the usefulness of the anonymized ASL video, we compared three prototypes (with-torso, without-torso, tiger-face) along three evaluation dimensions: understandability, naturalness, and anonymity. During the study we had collected some quantitative data, e.g., participants' Likert response to questions in phase 2 about each of these dimensions. Our quantitative analysis consisted of conducting Friedman tests, which indicated a statistical significance in understandability and naturalness among the three types of transformations, but no statistical significance for anonymity-protection. Following up with pairwise Wilcoxon signed-rank tests, significant differences among the types of transformations were identified. In our qualitative analysis, we found that the participants overall perceive the video transformation as interesting and useful. However we observed differing perspectives among participants in regard to how they compare these three prototypes along the three dimensions, as well as how this affects their overall views on the ASL video anonymization and its value. We present the details of the findings in the following sections.

4.1 Understandability

4.1.1 Quantitative Analysis. Figure 3 displays participants' responses during phase 2 of the study to the Likert item "This video was completely understandable." Analysis with a Friedman test revealed that the type of video transformation had a significant main effect on understandability ($p < .05$). Overall, 81% of respondents strongly agreed with this statement in regard to the with-torso videos, 62% of respondents strongly agreed in regard to without-torso videos, but only 25% agreed in regard to tiger-face videos. Post hoc pairwise analysis with a Wilcoxon signed-rank test and Bonferroni correction revealed that **participants believed the with-torso videos were more understandable than the tiger-face videos** ($p<0.01$). However, no significant difference was observed between with-torso and without-torso, nor between without-torso and tiger-face. Overall, these quantitative findings indicate that ASL signers believed that the modern 3D face transformation videos with a torso displayed (with-torso) were more understandable than the simple mask-overlay videos (tiger-face), when viewing videos of ASL.

4.1.2 Qualitative Analysis. The overall feedback in regard to the understandability of the anonymized videos of all three prototypes was **generally positive**, which aligned with the quantitative findings presented above. Most participants indicated that the transformed videos were clear and conveyed the same information as the original videos. Among the

Author's Pre-Print Copy

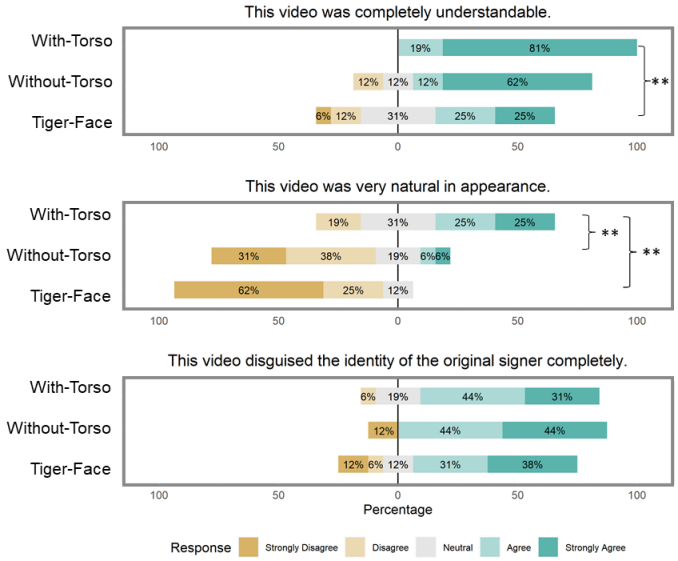

Fig. 3. Participants' agreement with Likert items in phase 2 of the study, for each of the 3 prototypes.

three transformations, participants agreed that the with-torso version was easiest to understand, but they believed the without-torso version was still relatively understandable, although somewhat less so.

Participants commented that even though the hand movements were intact in both the with-torso and without-torso videos, they felt that **understandability was reduced when the torso was cut out**. P9 said *"When you take out the torso, it was harder to understand."* P13 provided the reason: *"Without the body it was hard to detect the body language."* P14 implied that the removal of the torso may have interfered with their ability to focus on the message, and consequently to understand: *"It was distracting to have no body."*

Although participants said that seeing the body was useful, they emphasized that **facial expression was most important for understandability**. While both the tiger-face and with-torso videos retained the signer's original body appearance, all but one participant indicated that the tiger-face was the least understandable, because of the absence of the facial expression. P6 and P8 said respectively: *"Tiger face did not really bring the same information, the facial expressions were lost in that video"* and *"For tiger face, there are no facial expressions, feels weird."* Some participants described how the tiger-face itself was distracting. P13 said *"with tiger face, it was not very clear because it kind of blocked the signing because the face was big."* P5 expressed that while for a short video, the tiger-face animation could be understood with great effort, that *"if the tiger face was longer video then I might not understand as much."* Beyond the ASL linguistic information on the face, P5 described how the face conveys other information: *"The message is the same but because I see different faces - one person looks stoned and other person looks like messy hair - so I interpret things differently - same messenger but different feeling."*

4.2 Naturalness

4.2.1 Quantitative Analysis. Participants' responses during phase 2 of the study to the Likert item "This video was very natural in appearance" are shown in Figure 3. Quantitative analysis with a Friedman test revealed a significant ($p<0.01$)

Author's Pre-Print Copy

main effect of the type of video transformation on participants' rating of naturalness. Post hoc pairwise testing with a Wilcoxon signed-rank test, with Bonferroni correction, revealed that participants believed that the with-torso videos were significantly more natural ($p < .01$) than both the without-torso or tiger-face videos. Post hoc testing did not reveal any pairwise significant difference between without-torso and tiger-face videos. Overall, these findings indicate that **participants believed the with-torso videos were the most natural in appearance**.

4.2.2 Qualitative Analysis. Participants commented that **none of the three prototypes was completely natural**. In alignment with the quantitative findings above, most agreed that without-torso videos were less natural than with-torso videos, e.g., P14 said *"With torso was better because it looks more natural."* P13 agreed: *"With body was better because you can see the whole body as natural."* Participants indicated that they disliked the gray background color of the without-torso videos, and they also commented on there being visual "noise" at times, e.g., a flickering effect due to video-transformation artifacts. A few participants expressed concerns about insufficient facial expression in the disguised video, inadequate skin color match (between the face and the neck/arms), or unnatural hair color.

All but two participants commented that the **tiger-face videos were most unnatural** and explained that this was because no human face was visible. Among those with the minority opinion that tiger-face videos were more natural than without-torso videos, P9 explained that *"with the tiger face, you still see the body, body language, you can see the body shape. The face does not look natural but the rest of body was natural."* P1 was alone in believing that the tiger-face was the most natural of the three prototypes, saying *"because it was slightly believable while the others weren't."*

4.3 Anonymity

4.3.1 Quantitative Analysis. An analysis of the Likert response data in phase 2 in regard to the statement "This video disguised the identity of the original signer completely" did not reveal any significant effect of the type of video transformation on participants' response to this item, as shown in Figure 3. However, participants' high level of agreement on this item for all three prototypes suggests that they were **all seen as effective at disguising identity**. The vast majority of responses to this question were strongly agree or agree: 75% of respondents for the with-torso videos, 88% of respondents for the without-torso videos, and 69% of participants for the tiger-face videos.

After viewing a disguised video in phase 1, participants were asked to guess the original human face from a photo line-up, including other faces we had selected with similar race, gender, and hair style. From the male line-up, 9 of 16 participants were able to guess the original signer correctly. From the female line-up, 4 of 16 participants guessed the original signer correctly. This finding provides some evidence for **the potential of face swapping for anonymizing ASL videos**.

4.3.2 Qualitative Analysis. In addition to being asked in phase 1 to guess the identity of the signer in the line-up of photos, participants were also asked to comment on how difficult it was for them to do so. Almost all participants agreed that it was hard to guess the original signer from the transformed videos of both with-torso and without-torso versions. Participants explained that they could not use the face as a clue, but **they made use of used other appearance details, such as head shapes or skin color**. P14 described how he tried to approach this task, explaining that he tried to *"remember the skin color based on the arms."* P13 used a similar strategy, explaining that they made their selection *"based on the color of skin except for face."*

There was consensus among participants that the with-torso version would make it easier to identify someone from their clothes, background, or body shape, **especially if the signer was a friend or family member**. Overall, participants believed that the without-torso transformation would be most effective for anonymization. As P4 explained,

Author's Pre-Print Copy

"Knowing the person and seeing their torso and background would make it easier to identify them because the more you hang out with the person you know their body language and how they sign." P14 agreed that the without-torso videos had the greatest anonymity protection: *"Without torso is the best, sometimes you can identify people by the body shape, etc, but without seeing the body it is very difficult to guess despite that it might be harder to understand or not natural."*

Some participants believed that the tiger-face videos were most effective at disguising the face, **which is simply blocked, without any facial expressions revealed**. However, P13 explained that there are trade-offs between the ability of some prototypes to disguise the face or to disguise the body. As P13 explained, *"Without torso is the best. It covers the face and also hides the body language. You can't look at the body shape, size, etc. For tiger face, it hides the face the best but it doesn't hide the body at all. Without-torso has the best balance at hiding body but keeping facial expression."*

While participants agreed that without-torso videos were most effective at preserving anonymity, all participants commented that they would prefer to view a video with a torso – because of naturalness and understandability, as discussed previously. Several commented that it would be useful if this technology could make **modifications to the body of the signer instead of removing it**, e.g., suggesting that the tool could change the signer's clothing.

4.4 Preferences for Transforming Specific Characteristics

Throughout the study, participants viewed disguised videos of both themselves and other people, with a variety of characteristics transformed, e.g., age, hair color. Many participants indicated their preferences for video transformations **that closely matched their own traits**, such as race, age, hair, and skin color. P6's comment conveys this clearly: *"What I liked was all the faces using the same race or traits as me...I liked the faces that looked similar to my face."* Participants also emphasized the importance of having the transformed video match their own age. For instance, in response to a question about which of their own transformed videos of was their favorite, P7 answered: *"4th video. It was similar age, and looked the most natural to me."* P7 went on to emphasize the importance of the age feature for the natural appearance of the video: *"I would use a similar age, but I don't care about the other features as long as it doesn't look way off or too unnatural."* In the same vein, P1 expressed unhappiness with a transformed video with an older looking face, indicating that it was the least favorite video, and commenting, *"I didn't like that you made me old, I didn't like the age change."*

What mattered most to participants was whether **changing specific characteristics reduced the naturalness** of the resulting video; participants generally disliked transformations that resulted in artificial-looking hair color or the tiger face. In fact, all but one participant disliked having their hair transformed into bright colors. For instance, P2 commented, *"I didn't like using the different hair colors like purple hair was strange.*, and P9 added, *"it was funny to see my hair color look different."* Similarly, all but one participant disliked having their video transformed into the tiger face – with participants commenting on its artificial appearance and oversized head.

To a lesser degree, participants preferred transformations of characteristics that **supported understandability**. For instance, P12 noticed that some transformations preserved facial expression more clearly than others: *"it was easier to understand the younger faces than the older faces because I could see their mouth move."* Participants also mentioned that some transformations led to a **distracting result**, which interfered with their visual focus and thus their understanding. For example, P13 said, *"with tiger face, it was not very clear because it kind of blocked the signing, the face was big, and was distracting."* For the without-torso version, P14 commented, *"It was distracting to have no body".* P4 disliked brightly colored hair, explaining: *"it seemed distracting for me."*

Participants' preference among most transformations **did not depend upon whether it was applied to videos of themselves or of others**, with one exception: the removal of the torso from a video. Before participants viewed

Author's Pre-Print Copy

their own transformed videos, all of participants favored the with-torso videos, commenting on the natural appearance. Participants tended to retain this preference until they viewed their own transformed videos during phase 3, at which time half of the participants switched their preference from with-torso to without-torso, because they were worried that identifiable characteristics were visible in their with-torso videos.

In fact, upon seeing transformed videos of themselves, some participants not only became interested in the without-torso feature, but they also wondered how they could strategically transform **as many demographic and appearance characteristics as possible**, to protect their identity. For instance, P13 suggested: *"for improving anonymization, I would use a neutral color skin on the arms, neck, etc. And doesn't have to match gender, you could use neutral gender or opposite gender instead of having to match."* However, some participants noted that using this technology to change the skin color of one's face could produce offensive or insensitive results, with P9 musing, *"There could be a few issues with race..."*

Finally, participants believed that the appropriateness of specific appearance transformations would **depend upon the context of use**, as some situations required more anonymity or seriousness. P8 said, *"Doesn't matter to me which appearance, it's more about how serious I want to be when hiding my anonymity. If I wanted to hide, as is, I would pick without torso, doesn't really matter what hair color/age."* P5 indicated that *"If its formal, then it needs to look real/natural. Suppose Biden was presenting with a funny tiger face then I would be more resistant to watching while if it was comedian using it then I would understand. I think context is important."*

4.5 Potential Uses

Participants identified a variety of possible use cases for ASL video anonymization technology. In particular, nearly all participants agreed that the technology would be useful for safely expressing **personal views on sensitive or confidential topics**. P3 was interested in using it *"to avoid being targeted, want it to be anonymous. Some people might want to share important information but don't want to tie it to their identity."* P7 wanted to us this technology to *"post videos where I say things that I don't want associated with my identity. For example, political, abuse reports, protests, etc."* P10 was interested in using this tool to *"share my personal experience or feelings and I didn't want people to know who I am."* P2 explained they would use if for a "sticky question. If I was telling a powerful, heavy topic but wanted my identity hidden then I would use this. Mostly for sensitive topics." The ASL sign STICKY, used by P2 in their response, translates to the English concepts of awkward or embarrassing.

Participants also identified uses of this technology on social media, especially when they needed to share information that may be re-shared **beyond their own immediate personal network**, especially when ASL video would be more effective than text. For instance, P13 discussed sharing ASL lessons anonymously: *"I would use it for posting videos that strangers have access to, teaching ASL without revealing my face."* P11 discussed social media contexts in which protection of privacy is especially important, e.g., *"social media, OnlyFans, anonymous groups, etc.."* Participants also discussed uses for this technology on personal social media contexts during **fun or casual interaction with people they know**. P8 was interested in *"entertainment with friends and family, like the gaming community."* Similarly, P9 indicated that the transformed videos themselves may be entertaining or fun to share, explaining, *"I would also use it for entertainment...with friends, assuming they would not share it publicly."*

Finally, participants described contexts in which they would **not use this technology**, at times disagreeing with uses suggested by other participants. For example, several participants saw no use for this technology when interacting online with family or friends. As P13 explained, *"I would not want to use this if I was just talking with friends or people I know and trust."* P14 agreed and extended this to fellow students: *"If I was signing on my social media or with friends or*

Author's Pre-Print Copy

schoolwork, I wouldn't use it." Most broadly, P2 felt strongly that they *"wouldn't use this software for any other purpose that I am not trying to hide my identity like having fun, etc."*

4.6 Concerns

Participants indicated that they would find it acceptable to watch a video from someone that has been disguised by this technology, **as long as there was an ethical purpose**, e.g., if anonymity was needed in order to share important information or ask sensitive questions. However, participants expressed concerns about use of this technology for unethical purposes, such as harassment, trolling, or degrading someone online – as well as someone using this technology to steal information or to impersonate someone for fraudulent purposes. P5 expressed this trade-off: *"That's ethics. I wish it was safe for everyone to express their thoughts and concerns without being identified; however, this could be misused [or] abused so there needs to be a set of rules."*

Participants were especially concerned about the potential for this technology to enable someone to **impersonate another real person**. Although participants liked the concept of having a natural face, different from their own, appear in a disguised video, a majority of participants shared concerns about using another real person's face. P8 was worried about this technology producing videos using the faces of their family or friends, explaining, *"I don't want another person to intentionally disguise themselves as another person that I know. It could also be used for a scam or something. The faces should be fake and not from real people."* P1 suggested that a computer-generated virtual avatar face could be used instead of a photo of a real person, explaining *"also [I] don't understand why not using avatar - I would use my own avatar."* P1 wondered whose faces had been used as the basis of transformation, and she was worried about someone impersonating her, e.g., asking *"Who are the faces they are using? ... Are you using my face to hide other?"* The researcher explained that no face images of participants in the study were used to disguise the face of others. The faces were from a public research dataset [30, 31].

Finally, participants indicated that seeing someone use this technology may lead to **feelings of distrust**, as they may wonder about someone's motives for hiding their identity. For instance P1 indicated that upon seeing a video of someone that has been disguised, she would want to ask that person *"Why you feel the need to hide you face that much?"*

5 DISCUSSION

Our findings revealed trade-offs between key dimensions of importance to users, including: understandability, naturalness, and anonymity. In this section, we discuss pairs of these dimensions to inform the design of ASL video de-identification technology, and we examine factors that affect the acceptability of this technology for users.

5.1 Understandability vs. Anonymity: Design Considerations

Participants mentioned that facial expression and the movement and location of the body were important for preserving meaning in ASL, and transformations in which the signer's original torso was retained were rated as most understandable. However, there was a **tension** between greater understandability and participants' perceived degree of anonymity protection. Participants noted that with-torso videos revealed visual clues about the identity of the signer: their individual style of body movement, their clothing appearance, and other physiological traits, such as body size. Given that participants liked the understandability of with-torso videos, they were interested in improving anonymity protection without removing the torso completely. For instance, participants suggested that it would be valuable to extend this transformation so that, rather than hiding the body of the signer, the technology could apply some disguise to

Author's Pre-Print Copy

the body, while preserving its location and movement. As previously mentioned, some participants suggested virtually changing the body appearance or the clothes of the signer.

Our findings revealed that participants viewed the without-torso and tiger-face prototypes as being relatively similar in their degree of anonymity protection, which was striking given that these two tended to occlude or omit opposite portions of the signer's body. That is, the tiger-face blocked the signer's face—whereas, the without-torso videos omitted the signer's torso while conveying the facial expression information on a transformed face. Our qualitative findings revealed that participants judged the without-torso videos as more understandable; thus, **occlusion of the face led to a relatively greater reduction in understandability, for a similar anonymity improvement**. Recent work by Bragg et al. [7] had investigated ASL video anonymization within the context of motivating users to contribute videos to public research datasets; their participants had used the same tiger-face filter and had similar concerns about the negative effect on understandability of the absence of facial expressions.

For designers creating face-transformation applications, sensitivity to this understandability vs. anonymity trade-off is essential. While it would be ideal for the underlying transformation technology to achieve both high understandability and high anonymity (perhaps as further advances in face and body modification technology are created), in the meantime designers might consider **offering users choices** in transformation options that vary along this trade-off axis. For evaluation of these applications in studies, it is important for **both properties to be measured**, in relation to intended use cases, to avoid optimizing for one at the expense of the other.

5.2 Naturalness vs. Anonymity: Design Considerations

Participants indicated that it was important for videos to appear natural; however, our analysis revealed that there was a **trade-off** between naturalness and anonymity protection. Unanimously, our participants indicated that the with-torso videos were the most natural, yet these videos had weaker anonymity protection, as details of the signer's body and background were visible. In contrast, our qualitative analysis revealed that participants believed the without-torso and tiger-face videos were better at protecting anonymity, yet both of these had much lower levels of naturalness, due to the unfamiliar appearance of the torso being cut out of the video or the artificial animal face.

For individuals interested in disguising themselves, a decision must be made about where on this naturalness vs. anonymity trade-off the user would prefer for their video to be. This decision may depend upon the context of use, and designers creating face-disguise applications may wish to **provide users with options** that vary along this axis.

5.3 Understandability vs. Naturalness: Design Considerations

Whereas the discussion above identified trade-offs between naturalness vs. anonymity and understandability vs. anonymity, our findings revealed a **complementary relationship** between understandability and naturalness. Participants discussed how improvements in naturalness led to increased understandability, explaining that unnatural appearance could be distracting, which would draw attention away from the message. For designers of transformation technologies for face disguise applications, this relationship is important to consider when making improvements to the technology. In efforts to achieve increases in the understandability of the resulting video, it is important to ensure a baseline level of naturalness, to avoid interfering with the viewer's ability to focus on the message.

While there are relationships among these factors, the signer's intended usage of this technology is likely to influence how these factors are prioritized. Before seeing transformed videos of themselves, participants focused on the **perspective of people viewing videos** of other people who have been disguised, and understandability was seen as being of greater importance so that the message could to be understood. After seeing videos of themselves transformed,

Author's Pre-Print Copy

they shifted to the **perspective of someone who is transforming their own video** and became concerned with how they present themselves online, prioritizing naturalness to a greater degree. For instance, we reported that P1 disliked having been transformed into an older face, and this participant later explained: *"I would use this for situations when I want to look nice... In Snapchat you have the filters and you look better than normal, while this technology makes you look worse than normal. If it helps people look better than normal then it would be accepted."*

Our findings also inform how future designers or researchers investigating this technology should **design studies** to gather requirements from both perspectives. We found that simply asking participants to imagine using this technology to transform their own video was ineffective. Actually seeing their own videos transformed was what had sparked participants to re-prioritize their preferences and requirements for this technology.

5.4 Role of Identity in Acceptability: Design Consideration

Given the degree to which face appearance is considered a unique identifying characteristic of individuals, when participants first saw transformed videos of themselves, many expressed **mixed feelings**. Participants were struck with how well the technology had preserved their anonymity, to the point that many did not realize that they were viewing a transformed video of themselves. As P5 said, *"That was me? I didn't realize it was me. It was really interesting because I was watching I was looking for something... I recognizing the shirt...Now that I know it was me, I don't like it."* In addition, many participants expressed discomfort when first seeing another person's face on their body, as P16 explained, *"I was a little shocked to see the faces changed, huge difference."*

Our findings also revealed that the acceptability of specific transformations was dependent on a participant's concept of the **individual characteristics of their own identity**—and whether the technology had preserved, hidden, or transformed each. For instance, we mentioned earlier that half of our participants changed their preference from with-torso to without-torso videos upon seeing the first transformed video of themselves. They noticed personal traits on their body that were visible, e.g., ring on finger, clothes, nail polish. Beyond this risk to their anonymity, participants disliked partially transformed videos, as the unnatural appearance made the result appear fake. They were more comfortable when **all or none** of their identity characteristics were disguised or hidden.

When viewing a transformed video of another person, participants preferred for the video to retain as many characteristics of the original signer as possible. However, when considering how to transform a video of themselves, participants saw two sides to this issue: If the characteristics of the disguised face were similar to their real appearance, then they could **convey individual elements of their identity** when transmitting their message. Knowing the gender, age, or race of the person who had produced a message may be important context. On the other hand, selecting characteristics that differ from their real identity could provide a better disguise, thereby protecting anonymity better.

While several participants expressed interest in being able to transform their face into that of someone of a different **race or gender**, we did not enable this option in our study. Our rationale was that the current version of our prototype was limited to changing the face of the signer—not the skin color on the neck, arms, or body. Because of ethical concerns, we did not display videos in this study in which a face was overlaid on a body of someone of a different race, to avoid producing videos that may be insensitive or offensive. Future designers of face-transformation technology may need to address this desire for users to be able to replace their face with characteristics unlike their own, while providing guidance for users about ethical use of this technology.

The concept of identity was at the heart of many participants' ethical concerns over potential misuse of this technology described in section 4.6. Using real people's faces could lead to identity theft or impersonation that damages someone's reputation. Participants were concerned about someone using their face in this manner. To avoid such misuse, future

Author's Pre-Print Copy

designers of this technology could display a a disclaimer on the video to indicate that it has been transformed—or rather than using real faces as the target result, videos could be a composite/hybrid of the source and target face, and through this combination thereby producing a **novel identity** for the individual appearing in the resulting video.

6 CONCLUSION

DHH ASL signers who disprefer communication through written English must use video of ASL to communicate online, thereby revealing their face, which conveys essential linguistic information. These users currently lack effective options for communicating anonymously online, which prevents them from discussing sensitive topics or other activities. New advances in face transformation technologies enable replacing faces in video at a level of quality that preserves linguistic facial expressions and head movements essential for ASL. We conducted an **interview study with 16 DHH ASL signers**, who viewed ASL videos of themselves and others transformed by prototypes for disguising the face.

Our study evaluated three key dimensions of acceptability (understandability, naturalness, and anonymity protection), and quantitative and qualitative analysis of our findings revealed relationships among these dimensions. Our findings revealed that a prototype based on modern face-transformation technology was effective for preserving anonymity, and we **contribute empirical knowledge** about participants' assessment of this technology, preferences among various appearance transformation options, factors affecting the acceptability of this technology, uses of interest, and potential ethical concerns with this technology. Our study provides guidance for both designers of face-disguise applications and creators of anonymization technology for providing DHH ASL signers with new options for participating online.

7 LIMITATIONS AND FUTURE WORK

Future users of this technology may apply it to videos intended for sharing online, which they may record under various camera set-ups or environments. A limitation of our study is that participants generally produced videos at home in front of a computer, in a setting typical of a video-conference, because of the ongoing COVID-19 pandemic. Future research should investigate a **more diverse range of videos**, produced under a variety of real use-cases, to determine both the performance (understandability, naturalness, and anonymity) of this technology, and whether users' requirements or preferences are influenced by these factors. Further, while participants in our study had the experience of seeing their own video transformed, they did not have the experience of **actually posting that video online** to share with various audiences. A future study could investigate this technology as used in a more realistic context, which could reveal social factors that affect users' acceptance of such technology or preferences for it should be designed.

Another limitation is that the participants we recruited do not reflect the **full diversity of potential users** of this technology, which may include ASL signers who vary in age, technology experience, cultural background, or ASL fluency—as compared to the specific participants in our study, which disproportionately included recent university graduates, with a narrower range of demographic characteristics and life experience. Future work should investigate a wider range of potential users' interests and requirements in relation to this technology.

Section 3.1 explained our rationale for selecting the prototypes and transformations examined in our study, but future research is needed to explore a **wider range of design alternatives**, to understand more of this design space. In addition, we had selected the specific set of transformations applied to each participant's video in phase 3 of our study, but future work could investigate which transformation options participants would choose for themselves, e.g., if they were provided with an interface that enabled them to select among such options. Such a study may provide further insight as to how DHH signers may balance trade-offs among anonymity, understandability, and naturalness.

Author's Pre-Print Copy

Recent work has investigated applications of body-swap illusions in virtual reality [41], with users' new appearance leading to **changes in behavior** [58]. Our study did not examine whether signers might change their signing content or style if they were to see their own face transformed in real time; future work is needed to investigate this.

Finally, the with-torso and without-torso prototypes in our study were based on modern face transformation technologies, of which the state of the art is rapidly advancing. Future research is needed to understand users' perspectives of these technologies as they **improve over time**. In fact, our work should inform the work of future designers of such technology and of researchers creating the underlying disguise technologies, as we discussed in section 5. In particular, our research has motivated future work on technology for disguising not only the face of a signer but also their body—to better protect anonymity—while also preserving body location and movement, which contribute to the understandability and naturalness of the resulting video.

8 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under award No. 2040638.

REFERENCES

[1] Nicoletta Adamo-Villani and Ronnie B. Wilbur. 2015. ASL-Pro: American Sign Language Animation with Prosodic Elements. In *Universal Access in Human-Computer Interaction. Access to Interaction*. Springer International Publishing, 307–318. https://doi.org/10.1007/978-3-319-20681-3_29

[2] Robert W. Arnold. 2007. A proposal for a written system of American Sign Language.

[3] Werner Bailer and Martin Winter. 2019. On Improving Face Generation for Privacy Preservation. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*. 1–6. https://doi.org/10.1109/CBMI.2019.8877442

[4] 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.

[5] Steven Barnett, Jonathan D. Klein, Robert Q. Pollard, Vincent Samar, Deirdre Schlehofer, Matthew Starr, Erika Sutter, Hongmei Yang, and Thomas A. Pearson. 2011. Community Participatory Research With Deaf Sign Language Users to Identify Health Inequities. *American Journal of Public Health* 101, 12 (2011). https://doi.org/10.2105/ajph.2011.300247

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

[7] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring Collection of Sign Language Datasets: Privacy, Participation, and Model Performance. *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Oct 2020). https://doi.org/10.1145/3373625.3417024

[8] Danielle Bragg, Raja Kushalnagar, and Richard Ladner. 2018. Designing an Animated Character System for American Sign Language. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM. https://doi.org/10.1145/3234695.3236338

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

[10] Anna Cavender, Richard E Ladner, and Eve A Riskin. 2006. MobileASL: Intelligibility of sign language video as constrained by mobile phone technology. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. 71–78.

[11] G.R. Coulter. 1979. American Sign Language Typology.

[12] Xianghua Ding, Yanqi Jiang, Xiankang Qin, Yunan Chen, Wenqiang Zhang, and Lizhe Qi. 2019. Reading Face, Reading Health: Exploring Face Reading Technologies for Everyday Health. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300435

[13] Eleni Efthimiou, Stavroula-Evita Fotinea, Theodore Goulas, and Panos Kakoulidis. 2015. User Friendly Interfaces for Sign Retrieval and Sign Synthesis. In *Universal Access in Human-Computer Interaction. Access to Interaction*. Springer International Publishing, 351–361. https://doi.org/10.1007/978-3-319-20681-3_33

[14] Eleni Efthimiou, Stavroula-Evita Fotinea, Theodore Goulas, Anna Vacalopoulou, Kiki Vasilaki, and Athanasia-Lida Dimou. 2018. Sign Language technologies in view of Future Internet accessibility services. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 495–501.

[15] Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Lefebvre-Albaret. 2012. The Dicta-Sign Wiki: Enabling Web Communication for the Deaf. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 205–212. https://doi.org/10.1007/978-3-642-31534-3_32

Author's Pre-Print Copy

[16] Eleni Efthimiou, Stavroula-Evita Fotinea, Christian Vogler, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and Jérémie Segouat. 2009. Sign Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication. In *Universal Access in Human-Computer Interaction. Addressing Diversity*. Springer Berlin Heidelberg, 21–30. https://doi.org/10.1007/978-3-642-02707-9_3

[17] Facebook. 2021. Facebook Homepage. https://www.facebook.com/

[18] Jianping Fan, Hangzai Luo, Mohand-Said Hacid, and Elisa Bertino. 2005. A novel approach for privacy-preserving video sharing. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. 609–616.

[19] Ralph Gross, Edoardo Airoldi, Bradley Malin, and Latanya Sweeney. 2005. Integrating utility into face de-identification. In *International Workshop on Privacy Enhancing Technologies*. Springer, 227–242.

[20] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[21] Alexis Heloir and Fabrizio Nunnari. 2015. Toward an intuitive sign language animation authoring system for the deaf. *Universal Access in the Information Society* 15, 4 (May 2015), 513–523. https://doi.org/10.1007/s10209-015-0409-0

[22] M. Shamim Hossain and Ghulam Muhammad. 2015. Cloud-Assisted Speech and Face Recognition Framework for Health Monitoring. *Mob. Netw. Appl.* 20, 3 (June 2015), 391–399. https://doi.org/10.1007/s11036-015-0586-3

[23] Matt Huenerfauth and Vicki Hanson. 2009. Sign language in the interface: access for deaf signers. *Universal Access Handbook. NJ: Erlbaum* 38 (2009).

[24] Instagram. 2021. Instagram website. https://www.instagram.com/

[25] Brielle Jaeke. 2016. Sephora boosts augmented reality shopping with real-time facial recognition. http://t.cn/EGjuPZK

[26] Jeeliz. 2021. Jeeliz website. https://jeeliz.com/demos/faceFilter/demos/threejs/tiger/

[27] Hernisa Kacorri and Matt Huenerfauth. 2016. Continuous profile models in asl syntactic facial expression synthesis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2084–2093.

[28] Pavel Korshunov and Sébastien Marcel. 2018. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685* (2018).

[29] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[30] Debbie S. Ma, Joshua Correll, and Bernd Wittenbrink. 2015. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods* 47, 4 (Jan. 2015), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

[31] Debbie S. Ma, Justin Kantner, and Bernd Wittenbrink. 2020. Chicago Face Database: Multiracial expansion. *Behavior Research Methods* (Oct. 2020). https://doi.org/10.3758/s13428-020-01482-5

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

[33] Sachit Mahajan, Ling-Jyh Chen, and Tzu-Chieh Tsai. 2017. SwapItUp: A Face Swap Application for Privacy Protection. In *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. 46–50. https://doi.org/10.1109/AINA.2017.53

[34] Ross E. Mitchell. 2005. How Many Deaf People Are There in the United States? Estimates From the Survey of Income and Program Participation. *The Journal of Deaf Studies and Deaf Education* 11, 1 (09 2005), 112–119. https://doi.org/10.1093/deafed/enj004

[35] Carol Neidle, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. 2000. In *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, MA: MIT Press.

[36] Carol Neidle and Augustine Opoku. 2020. A User's Guide to the American Sign Language Linguistic Research Project (ASLLRP) Data Access Interface (DAI) 2 — Version 2. American Sign Language Linguistic Research Project Report No. 18, Boston University.. http://www.bu.edu/asllrp/rpt18/asllrpr18.pdf

[37] Carol Neidle and Augustine Opoku. 2021. Update on Linguistically Annotated ASL Video Data Available through the American Sign Language Linguistic Research Project (ASLLRP). American Sign Language Linguistic Research Project Report No. 19, Boston University.. http://www.bu.edu/asllrp/rpt18/asllrpr18.pdf

[38] Carol Neidle, Augustine Opoku, Gregory Dimitriadis, and Dimitris Metaxas. 2018. NEW Shared & Interconnected ASL Resources: SignStream® 3 Software; DAI 2 for Web Access to Linguistically Annotated Video Corpora; and a Sign Bank. In *8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Miyagawa, Japan) *(LREC 2018)*. 147–154.

[39] Don Newkirk. 1987. SignFont Handbook. (1987).

[40] Yuval Nirkin, Iacopo Masi, Anh Tran Tuan, Tal Hassner, and Gerard Medioni. 2018. On face segmentation, face swapping, and face perception. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 98–105.

[41] Tabitha C. Peck, Jessica J. Good, and Kimberly A. Bourne. 2020. Inducing and Mitigating Stereotype Threat Through Gendered Virtual Body-Swap Illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376419

[42] Robert Q Pollard, Erika Sutter, and Catherine Cerulli. 2013. Intimate Partner Violence Reported by Two Samples of Deaf Adults via a Computerized American Sign Language Survey. *Journal of Interpersonal Violence* 29, 5 (2013), 948–965. https://doi.org/10.1177/0886260513505703

[43] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. 2019. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1–11.

Author's Pre-Print Copy

[44] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. 2019. First Order Motion Model for Image Animation. In *Conference on Neural Information Processing Systems (NeurIPS)*.

[45] Aliaksandr Siarohin, Subhankar Roy, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. 2020. Motion Supervised co-part Segmentation. *arXiv preprint* (2020).

[46] Snapchat. 2021. Snapchat website. https://www.snapchat.com/

[47] 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.

[48] Valerie Sutton. 1998. The Signwriting Literacy Project. *Impact of Deafness on Cognition AERA Conference* (1998).

[49] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2016. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2387–2395.

[50] Yu Tian, Xi Peng, Long Zhao, Shaoting Zhang, and Dimitris N. Metaxas. 2018. CR-GAN: Learning Complete Representations for Multi-view Generation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 942–948. https://doi.org/10.24963/ijcai.2018/131

[51] Yu Tian, Xi Peng, Long Zhao, Shaoting Zhang, and Dimitris N. Metaxas. 2018. CR-GAN: Learning Complete Representations for Multi-view Generation. *International joint conference on artificial intelligence (IJCAI)* (2018), 942–948.

[52] TikTok. 2021. TikTok website. https://www.tiktok.com/

[53] Carol Bloomquist Traxler. 2000. The Stanford Achievement Test, 9th Edition: National Norming and Performance Standards for Deaf and Hard-of-Hearing Students. *Journal of Deaf Studies and Deaf Education* 5, 4 (Jan 2000), 337––348. https://doi.org/10.1093/deafed/5.4.337

[54] Hitomi Tsujita and Jun Rekimoto. 2011. Smiling Makes Us Happier: Enhancing Positive Mood and Communication with Smile-Encouraging Digital Appliances. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2030112.2030114

[55] Clayton Valli and Ceil Lucas. 2000. *Linguistics of American sign language: An introduction.* Gallaudet University Press.

[56] Vcom3D. 2015. Sign Smith Studio website. http://www.vcom3d.com/signsmith.php

[57] Lezi Wang, Chongyang Bai, Maksim Bolonkin, Judee Burgoon, Norah Dunbar, V. S. Subrahmanian, and Dimitris N. Metaxas. 2019. Attention-based facial behavior analytics in social communication. *30th British Machine Vision Conference (BMVC'19)*.

[58] Nick Yee and Jeremy Bailenson. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research* 33, 3 (July 2007), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

[59] YouTube. 2021. YouTube website website. https://www.youtube.com/

[60] Long Zhao, Xi Peng, Yu Tian, Mubbasir Kapadia, and Dimitris N. Metaxas. 2020. Towards Image-to-Video Translation: A Structure-Aware Approach via Multi-stage Generative Adversarial Networks. *International Journal of Computer Vision* 128, 10-11 (April 2020), 2514–2533. https://doi.org/10.1007/s11263-020-01328-9

[61] New Zoogle. 2021. Best Animal Face Changer Apps for Android. https://newzoogle.com/best-animal-face-changer-apps-android/