Comparing Native Signers' Perception of American Sign Language Animations and Videos via Eye Tracking

Hernisa Kacorri
The Graduate Center, CUNY
The City University of New York
Computer Science Ph.D. Program
365 Fifth Ave, New York, NY 10016
+1-212-817-8190
hkacorri@gc.cuny.edu

Allen Harper
The Graduate Center, CUNY
The City University of New York
Computer Science Ph.D. Program
365 Fifth Ave, New York, NY 10016
+1-212-817-8190
aharper@gc.cuny.edu

Matt Huenerfauth
The City University of New York
Queens College & Graduate Center
Computer Science and Linguistics
365 Fifth Ave, New York, NY 10016
+1-718-997-3264
matt@cs.qc.cuny.edu

ABSTRACT
Animations of American Sign Language (ASL) have accessibility benefits for signers with lower written-language literacy. Our lab has conducted prior evaluations of synthesized ASL animations: asking native signers to watch different versions of animations and answer comprehension and subjective questions about them. Seeking an alternative method of measuring users' reactions to animations, we are now investigating the use of eye tracking to understand how users perceive our stimuli. This study quantifies how the eye gaze of native signers varies when they view: videos of a human ASL signer or synthesized animations of ASL (of different levels of quality). We found that, when viewing videos, signers spend more time looking at the face and less frequently move their gaze between the face and body of the signer. We also found correlations between these two eye-tracking metrics and participants' responses to subjective evaluations of animation-quality. This paper provides methodological guidance for how to design user studies evaluating sign language animations that include eye tracking, and it suggests how certain eye-tracking metrics could be used as an alternative or complimentary form of measurement in evaluation studies of sign language animation.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**] User Interfaces – *evaluation/methodology*; K.4.2 [**Computers and Society**]: Social Issues – *assistive technologies for persons with disabilities*.

General Terms
Design, Experimentation, Human Factors, Measurement.

Keywords
Accessibility Technology for People who are Deaf, Eye Tracking, American Sign Language, Animation, User Study.

1. INTRODUCTION
Due to a variety of language exposure and educational factors, there is a large proportion of deaf adults in the U.S. with limited literacy in written English. In fact, a majority of deaf high school graduates (age 18-21) have English literacy skills that are at the U.S. "fourth grade" (age 10) level [32]. While there is a variety of accessibility accommodations for people who are deaf, many of them rely on the written-language reading skills of the user, e.g., captioning on television, written text on websites, etc. This information is not accessible for deaf users with low literacy.

American Sign Language (ASL) is used by over 500,000 people [26]. Because it is a distinct language from English (with a different word order and linguistic details), there are many people who are fluent in ASL, yet have difficulty reading English text. One approach to make information content more accessible for these users is to present it using ASL. While videos of humans performing ASL can be used in some contexts, the difficulty in regularly updating the information content of videos (or synthesizing novel video by splicing others) has led researchers to investigate methods for automating the synthesis of computer animations of ASL [14]. Production methods used in the entertainment industry to create high-quality animations of virtual human characters for film or television can have natural results; though the time needed to carefully control the movements of a character makes such high-effort methods impractical for quickly creating and updating ASL messages for websites or documents. Instead, researchers investigate methods for automating the synthesis of animations of a virtual human performing sign language, based on a small number of input parameters that specify the words in the message or other details. While this paper focuses on ASL, researchers have studied animation synthesis for a variety of sign languages (see survey in [14]). The methodological issues examined in this paper (eye tracking in user studies) are also relevant to researchers studying animation synthesis for other sign languages internationally.

To guide our work, our lab frequently conducts studies in which native ASL signers evaluate the naturalness and understandability of animations produced by our software [13, 15, 16]. In a typical experiment, a participant watches an ASL animation and then answers some subjective evaluation questions and some comprehension questions about the animation's content. Over the past several years, our lab has designed and evaluated new methodologies for conducting evaluation studies of sign language animation, including screening for participants with appropriate ASL skills, collecting specific ASL language samples for analysis, and designing questions for studies that are accessible to participants with low English literacy [17, 23].

This paper focuses on another important methodological issue: how eye tracking can be used in user-based experimental studies of sign language animations. There are several reasons why our laboratory has begun to focus on this topic: In recent work, we

have been investigating how to synthesize facial expressions for ASL-signing virtual humans; facial expressions are an integral part of ASL and convey important linguistic information [28]. Unfortunately, it is challenging to design experimental stimuli and questions that effectively measure whether participants have understood the information being conveyed by facial expressions. As discussed in [16, 20], signers may not consciously notice a facial expression during an ASL passage, and the subtle and complex ways in which facial expressions can affect the meaning of ASL sentences can make it difficult to invent stimuli and questions that effectively probe a participant's understanding of the information conveyed by the signer's face.

As discussed in section 2, researchers in various fields have used eye tracking to unobtrusively probe where participants are looking during an experiment (and in some cases, to infer the cognitive processes or task-strategies of those users). In fact, sections 2.1 and 2.2 discuss how researchers have successfully used these methods with participants who are deaf, to investigate perception, reading, and sign-language comprehension (of videos of humans). This paper examines whether these methods can be adapted to the evaluation of sign language animations; specifically, we ask:

- Does the eye-movement behavior of native ASL signers participating in an experiment differ depending on whether they are watching a video of a human signer or an ASL animation?
- …whether they are viewing ASL animations with some facial expressions or ASL animations with no facial expressions?
- Does the eye-movement behavior of these participants correlate to their responses to subjective evaluations questions or comprehension questions about the videos/animations?

Section 3 describes how we selected which eye-tracking metrics to study by considering prior research. Also, it discusses how we formulated some hypotheses (more specific than the research questions listed above) about how the eye-movements of native ASL signers relate to the quality of the ASL video/animation and to participants' responses to subjective and comprehension questions. Section 4 describes our experiments recording the eye movements of native ASL signers who view animations/videos and then answer subjective and comprehension questions. Finally, Sections 5 and 6 describe our results and conclusions.

Notably, for this paper, we are not primarily concerned with determining the level of quality of any particular ASL animation (which has traditionally been the focus of our prior experiments). Instead, we are focused on whether the eye movements of native ASL signers reveal information about the quality of the ASL video/animation being viewed. We compare videos and animations under the supposition that very high-quality ASL animations may lead to similar eye-movement patterns as videos. If a correlation can be found between eye-movement metrics and certain types of ASL videos (or participants' responses to evaluation/comprehension questions about the stimuli), then this relationship could be utilized when designing future evaluation studies of ASL animations. Those metrics could be used as an additional or alternative form of evaluation for ASL animations. In some experimental contexts, it may be desirable not to interrupt participants with questions, or asking specific questions might artificially draw attention to aspects of the animation that could lead to unnatural interactions (e.g., if we wanted to study the effect of different eye-brow movements for our animated signer, if we ask too many questions about the eye-brows, then signers may stare at them, instead of simply watching the animations for their information content). In other contexts (such as for ASL animations with facial expression), it can be difficult to engineer large numbers of stimuli and questions that effectively probe whether the animation is high-quality or well-understood.

2. EYE-TRACKING & RELATED WORK

Several authors have surveyed eye tracking in human computer interaction [6, 18, 29]. Essential information about this technology related to the current paper is summarized in this section. The bright-pupil technique used in this paper employs a near infrared light source, which illuminates the pupil (the "red-eye" effect) and creates a reflection on the cornea (first Purkinje image). Image processing software identifies: (1) the center of the pupil and (2) the corneal reflection. By comparing the relationship between these two artifacts in the eye video, the point of gaze on the stimuli can be determined. In a desktop-mounted eye-tracker, such as the Applied Science Labs D6 system used in our study [1], cameras and the illuminator are in a small device (placed directly below the computer screen that displays the visual stimuli). The participant is seated in front of the 19-inch computer screen (resolution 1440x900) at a typical viewing distance (with their eye approximately 60cm from the eye-tracker device). The participant is able to make head movements (up to 30cm) during the study and the eye-tracker software tracks the participant's head location and orientation to compensate for these movements.

The system records the horizontal and vertical coordinates on the computer screen where the eye is aimed. Human eye gaze tends to move rapidly from one location to another, during movements called "saccades." Moments when the eye is relatively stationary are called "fixations." Thus, eye-tracking data is usually processed into a list of the fixations that occur during a study, each with a: start-time, end-time, horizontal and vertical screen coordinates, and other information. To facilitate analysis, researchers typically perform one more step of processing on the fixation list. They define regions of the computer screen (during specific periods of time during the study) that are significant; such regions are called "Areas of Interest" or AOIs. For example, during a usability study, each button on a user-interface may be defined as an AOI, consisting of the shape and location of the button and the time duration when it was visible. Each fixation in the fixation list can thus be labeled as to whether it was within an AOI. In this paper, we will define AOIs for regions of the face and body of an onscreen human or animated character who is performing ASL. For each AOI, it is possible to calculate a "proportional fixation time," which is the sum of the duration of all fixations on this AOI, divided by the total time of the recording segment.

Eye tracking enables researchers to collect a detailed sequential record of how users visually interact with stimuli. While the link between visual attention and cognitive processes is not completely understood, there is a general consensus among eye-tracking researchers of the validity of the so-called "eye-mind" hypothesis, that: "eye movements and attention are assumed to serve useful purposes connected to the visual task" [21]. Modern video-based eye-tracking has been applied to diverse areas of research, including: the psychology of reading (e.g., [30]), web search (e.g., [8, 10]), user-interface usability (e.g., [9]), cognitive workload estimation (e.g., [2]), and cognitive modeling (e.g., [11, 31]).

2.1 Eye-Tracking Participants who are Deaf

Researchers have conducted eye-tracking studies with participants who are deaf, to examine reading strategies, perception, or software usability; some of these studies involve comparisons between deaf and hearing participants. For instance, in [34], deaf and hearing subjects rated static face images for ten different emotional states while their eye movements were recorded with a desktop eye tracker. Eye metrics were calculated for proportional

fixation time and average gaze duration. Interestingly, while no differences were found between the two groups in how they rated the images, there were measurable differences in eye movement patterns. In particular, deaf subjects had greater proportional fixation times as well as mean gaze duration on the eyes AOI while hearing subjects had more fixation time as well as longer gaze durations on the nose AOI.

Other researchers have studied reading, e.g., [19] compared reading strategies used by deaf and hearing participants. A desktop-mounted eye-tracker monitored eye-movement behaviors while participants read Dutch texts on websites. Chapdelaine et al. [5] compared the eye movements of deaf and hearing subjects when watching captioned videos. They recorded proportional fixation time and gaze duration for several AOIs in their videos: faces in the videos, areas of motion in the videos, and caption regions. They found that deaf users spent less proportional fixation time on the captions than the hearing group, but the deaf users scored higher on recall tests of information from the videos.

Finally, several researchers have used eye-tracking technology to study how deaf students balance their attention across several sources of information during classroom lectures.

- Marschark [25] examined the eye-movements while college students were presented visual stimuli consisting of a lecturer, an interpreter, and a projection screen. Their participants included deaf students who were: skilled signers and less experienced signers. Conducted in a classroom setting, this experiment used a head mounted eye-tracker worn by the participants. Eye movement data was recorded for proportional fixation time, mean gaze duration (average length of time per gaze), and transitions between the three AOIs. They found that both groups of deaf students had similar eye metrics.
- Cavender et al. [3] conducted a usability study of a multi-modal educational user interface for deaf students that had four regions: the lecturer, the interpreter, slides, and captions. To assist students in noticing when a slide change occurred, notification schemes were implemented that altered the user interface component (e.g., color change) at that moment. A desktop eye tracking system was used to capture fixation data for the four AOIs. They found that the students spent more time looking at the interpreter and captions, as opposed to allocating their visual attention to the instructor or the slides.

2.2 Eye-Tracking with Sign-Language Video

While we are not aware of any prior studies that have used eye-tracking techniques to evaluate sign language *animations*, this section describes some examples of studies that have recorded participants viewing *videos* of sign language. For instance, Cavender et al. [4] conducted a preliminary study to evaluate the understandability of videos of sign language displayed at different sizes (based on screen sizes of mobile phones) and video-compression rates. Four participants viewed videos while eye-tracked, and they answered evaluation questions about each video. The authors found most fixations were close to the signer's mouth in the videos. They also found that the path length traced by fixations was shorter for the medium-sized video in their study, which was the video that received the highest subjective scores from participants. Finally, the authors analyzed instances when participants' gaze transitioned away from the signer's face; this occurred during some fingerspelling, when hands moved to the bottom of the screen, when the signer looked away from the camera, or when the signer pointed to locations outside the video.

Muir and Richardson [27] performed an eye tracking study to determine how native British Sign Language (BSL) signers use their central (high-resolution) vision and peripheral vision when viewing BSL videos. Their earlier work had suggested that signers tend to use their central vision on the face of a signer, and they tend to use peripheral vision for hand movements, fingerspelling, and body movements. In [27], native BSL signers watched three videos that varied in how visually challenging they were to view: (1) close-up above-the-waist camera view of the signer with no fingerspelling or body movement, (2) distant above-the-knees view of the signer with use of some fingerspelling, (3) distant above-the-knees view of the signer with use of fingerspelling and body movements. Participants' eye movements were recorded and proportional fixation time was computed over five AOIs: upper face, lower face, hands, fingers, upper body, and lower body. (The researchers had to carefully view the recordings of their eye-tracking data to determine when the participant was looking at each of these moving portions of the signer's body.)

Detailed signs and fingerspelling did not accumulate large proportional fixation time, indicating that participants used their peripheral vision to observe these aspects of sign language video. For all three videos, the face AOIs received the most proportional fixation time: 88%, 82%, 60% respectively. Video 3 included upper body movement, and participants spent more time looking at the upper body of the signer. During video 1, participants looked at the upper face 72% and lower face 16%, but during video 2 (more distant view of the signer), they looked at the upper face 47% and lower face 35%. These results are of interest to our current study because they indicate that when participants view sign language videos that have lower clarity (because the signer is more distant from the camera), their attention may shift to different areas of the video image, perhaps in an effort to search for the AOI with the most useful and visible information. This suggests that studying proportional fixation time on the face might be a useful way to analyze eye-tracking data when participants are viewing sign language videos (or animations) of different quality.

Emmorey et al. [7] conducted an eye tracking experiment to investigate differences in eye movement patterns between native and beginner ASL signers. The authors hypothesized that novice signers would have a smaller visual field from which to extract information from a signer. This in turn would lead to: less time fixating on the singer's face, more fixations on the lower mouth and upper body, and more transitions away from the face to the hands and lower body. This study was conducted with live signing performances and a mobile head-mounted eye tracker was used. Two stories were constructed which differed in the amount of fingerspelling and use of locative classifier constructions (signs that convey spatial information, investigated in our prior work [12]), with the goal of inducing more transitions in novice signers due to a restricted perceptual span. Both native and novice signers had similar proportional fixation times (89%) on the face; however, novices spent significantly more time fixating on the signer's mouth than native signers, who spent more time fixating on the signer's eyes. Also, neither novices nor native signers made transitions to the hands during fingerspelling, but did make transitions towards classifier constructions.

3. EYE-GAZE METRICS & HYPOTHESES

While our laboratory has investigated the calibration and use of motion-capture equipment (including eye-trackers) for recording sign language performances from native signers [22, 23], we had not previously used eye-tracking technology to record native signers while they viewed animations of ASL (nor did we find

prior published work in which this was done). Therefore, we consider prior work on ASL *videos* to determine the eye-tracking metrics we should examine and the hypotheses we should test.

While Muir and Richardson [27] did not study sign language *animation*, they observed changes in proportional fixation time on the face of signers when the visual difficulty of videos varied. Thus, we decided to examine the proportional fixation time on the signer's face. Since there is some imprecision in the coordinates recorded from a desktop-mounted eye-tracker, we decided not to track the precise location of the signer's face at each moment in time during the videos. Instead, we decided to define an AOI that consists of a box that contains the entire face of the signer in approximately 95% of the signing stories. (We never observed the signer's nose leaving this box during the stories.) Details of the AOIs in our study can be found in section 4.

The problem with examining only the proportional fixation time metric is that it does not elucidate whether the participant: (a) stared at the face for a long time and then stared at the hands for a long time or (b) often switched their gaze between the face and the hands during the entire story. Both types of behaviors could produce the same proportional fixation time value. Thus, we also decided to define a second AOI over the region of the screen where the signer's hands may be located, and we record the number of "transitions" between the face AOI and the hands AOI during the sign language videos and animations.

Since prior researchers have recorded that native signers viewing understandable videos of ASL focus their eye-gaze almost exclusively on the face, we make the supposition that if a participant spends time gazing at the hands (or transitioning between the face and hands), then this might be evidence of non-fluency in our animations. It could indicate that the signer's face is not giving the participant useful information (so there is no value in looking at it), or it could indicate that the participant is having some difficulty in recognizing the hand shape/movement for a sign (so participants need to direct their gaze at the hands). In [7], less skilled signers were more likely to transition their gaze to the hands of the signer. If we make the supposition that this is a behavior that occurs when a participant is having greater difficulty understanding a message, then we would expect more transitions in our lower-quality or hard-to-understand animations or videos. While [7] also noted eye-gaze at locative classifier constructions by both skilled and unskilled signers, the stimuli in our study do not contain classifier constructions (complex signs that convey 3D motion paths or spatial arrangements).

Based on these prior studies, we hypothesize the following:

- H1: There is a significant difference in native signers' eye-movement behavior between when they view *videos* of ASL and when they view *animations* of ASL.
- H2: There is a significant difference in native signers' eye-movement behavior when they view animations of ASL *with* some facial expressions and when they view animations of ASL *without* any facial expressions.
- H3: There is a significant correlation between a native signer's eye movement behavior and the scalar *subjective scores* (grammatical, understandable, natural) that the signer assigns to an animation or video.
- H4: There is a significant correlation between a native signer's eye movement behavior and the signer *reporting having noticed* a facial expression in a video or animation.
- H5: There is a significant correlation between a native signer's eye movement behavior and the signer correctly answering *comprehension questions* about a video or animation.

Each hypothesis above will be examined in terms of the following two eye-tracking metrics: proportional fixation time on the face and transition frequency between the face and body/hands. Based on the results of H1, we will determine whether to consider video separately from animations for H3 to H5. Similarly, results from H2 will determine if animations with facial expressions are considered separately from animations without, for H3 to H5.

4. USER STUDY

To evaluate hypotheses H1-H5, we conducted a user study, where participants viewed short stories in ASL performed by either a human signer or an animated character. In particular, each story was one of three types: a "video" recording of a native ASL signer, an animation with facial expressions based on a "model," and an animation with a static face (no facial expressions) as shown in Fig. 1. Each "model" animation contained a single ASL facial expression (yes/no question, wh-word question, rhetorical question, negation, topic, or an emotion), based on a simple rule: apply one facial expression over an entire sentence, e.g. use a rhetorical-question facial expression during a sentence asking a question that doesn't require an answer. Additional details of the facial expressions in our stimuli appear in [20, 24].

Fig. 1: Screenshots from the three types of stimuli: i) video of human signer, ii) animation with facial expressions, and iii) animation without facial expressions.

A native ASL signer wrote a script for each of the 21 stories, including one of six types of facial expressions. To produce the video stimuli, we recorded a second native signer performing these scripts in an ASL-focused lab environment, as illustrated in [24]. Then another native signer created both the model and no facial expressions animated stimuli by consulting the recorded videos and using some animation software [33]. The video size, resolution, and frame-rate for all stimuli were identical.

During the study, after viewing a story, each participant responded to three types of questions. All questions were presented onscreen (embedded in the stimuli interface) as HTML forms, as shown in Fig. 2, to minimize possible loss of tracking accuracy due to head movements of participants between the screen and a paper questionnaire. On one screen, they answered 1-to-10 Likert-scale questions: three subjective evaluation questions (of how grammatically correct, easy to understand, and naturally moving the signer appeared) and a "notice" question (1-to-10 from "yes" to "no" in relation to how much they noticed an emotional, negative, questions, and topic facial expression during the story). On the next screen, they answered four comprehension questions on a 7-point Likert scale from "definitely no" to "definitely yes." Given that facial expressions in ASL can differentiate the meaning of identical sequences of hand movements [28], both stories and comprehension questions were engineered in such a way that the wrong answers to the comprehension questions would indicate that the participants had misunderstood the facial expression displayed [20]. E.g. the comprehension-question responses would indicate whether a participant had noticed a "yes/no question" facial expression or instead had considered the story to be a declarative statement.

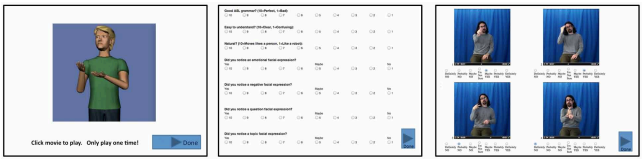

Fig. 2: An example of a stimulus used in the study: story, subjective and notice questions, and comprehension questions.

An initial sample story familiarized the participants with the experiment and the eye tracking system. All of the instructions and interactions were conducted in ASL; Likert scale questions were explained in ASL. Part of the introduction, included in the beginning of the experiment, and the comprehension questions were presented by a video recording of a native ASL signer.

Fig. 3 illustrates how we defined the "Face" and "Hands" areas of interest (AOIs) for the videos of the human signer and the animations of the virtual character. Identical AOIs were used for the animations with or without facial expressions. Note that the region of the screen where the hands may be located could potentially overlap with where the face is located (signers may move their hands in front of their face when signing), but our AOIs are defined so that they do not overlap. We made the simplifying assumption that the face should take precedence, and that is why the Hands AOI has an irregular shape to accommodate the Face AOI. Thus, if a participant were looking at our signer's hands when they moved in front of the signer's face, we would count that moment of time as a "face" fixation. This is a limitation of our study, but it simplified the eye-tracking analysis, and we believe that it had a minimal effect on the results obtained, given that the signer's hands do not overlap with the face during the vast majority of signing. While the Face AOIs have different horizontal/vertical ratios to accommodate the different head shapes and movements of the signers, the area (length x width) of the Face AOI for the animated character is identical to the area of the Face AOI for the human. The human signer performed some torso movements when signing, such as bending forward slightly, therefore the region of the screen where his hands tend to occupy is a little lower compared to the animated character. So, we set the borders of the Hands AOI lower for the human signer; to preserve fairness, we kept the area of the two Hands AOIs as similar as possible. The area of the animation Hands AOI is 99.3% of the area of the video Hands AOI.

Fig. 3: Screen regions for the face and hands AOIs of the animated character and the human signer.

5. RESULTS AND COMPARISONS

Ads were posted on New York City Deaf community websites asking potential participants if they had grown up using ASL at home or had attended an ASL-based school as a young child. Of the 11 participants recruited for the study: 7 learned ASL since birth, 3 learned ASL prior to age 4, and 1 learned ASL at age 8. This final participant attended schools for the deaf with instruction in ASL from age 8 to 18, and she uses ASL daily at home and at work. There were 4 men and 7 women of ages 24-44 (average age 33.4). We recorded eye-tracking data once for each story that was shown to participants (prior to the participant being asked Likert-scale or comprehension questions about the story). Because the eye-tracker could occasionally "lose" the pupil of the participant's eye during tracking (e.g., if the participant rubbed their face with their hand during the experiment), we needed to filter out any eye-tracking data in which there was a loss of tracking accuracy. Therefore, we decided to analyze only those recordings that meet both of these criteria:

- The eye-tracker was able to identify the participant's head and pupil location for at least 50% of the story time.
- The eye-tracker recorded that participant was looking at the video/animation for at least 50% of the time. (This criterion may fail if the participant looked away from the screen or there was a tracking calibration problem for that story. Eye-trackers must be calibrated periodically during use so that they know how the observed eye angles correspond to screen coordinates.)

The threshold values of 50% in these two criteria were selected after consulting a histogram of the relevant eye-tracking values to determine a natural boundary in the data. Applying these filtering criteria reduced the number of recordings from 231 to 181.

Section 3 discussed how we considered two eye-tracking metrics during our analysis:

- FacePFT: Proportional fixation time on the face AOI (i.e., total time of all fixations on the face AOI divided by story duration).
- TransFH: Number of transitions between the face AOI and hands/body AOI, divided by the story duration (in seconds).

Proportional fixation time and transition frequencies are typically not normally distributed, and Shapiro-Wilk tests confirmed this on the data collected in our study. For this reason, we used non-parametric tests of statistical significance (Kruskal-Wallis) and correlation (Spearman's Rho) during our analysis.

Fig. 4 shows a box plot of the FacePFT values for the video, animation with facial expressions ("Model"), and animation without facial expression ("Non"). Box edges indicate quartiles, whiskers indicate minimum/maximum values (all with values of 0 or 1), and the centerline indicates the median. Stars indicate significant pairwise differences (Kruskal-Wallis, $p<0.05$).

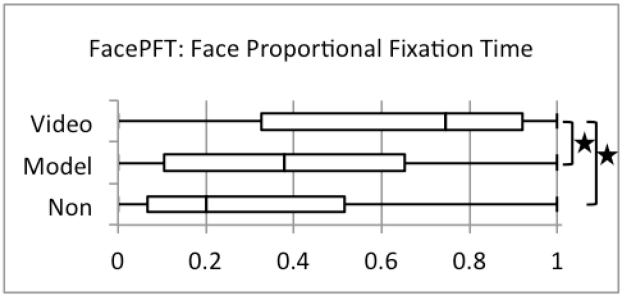

Fig. 4: Proportional Fixation Time on the Face

In Fig. 4, Hypothesis H1 was supported: participants spent significantly more time looking at the face AOI of the videos. We did not observe any support for H2: no significant difference was observed between the animations with and without facial expression. However, the median of the FacePFT values for Non, was (not significantly) lower than the median of the FacePFT values for Model. In a future study, we may record a larger

number of participants to determine if the lack of significance here was due to an insufficient number of participants.

Fig. 5 shows TransFH values. Note that due to the preponderance of zero values in the TransFH data, the boxes and whiskers of the plots are against the zero axis. These results show a similar (but inverse) pattern as the FacePFT values. When watching videos, participants moved their eyes between the face AOI and the body/hands AOI less frequently, than when watching animations. H1 was again supported: TransFH was lower for Video. H2 was not supported: no significant difference was observed between the animations with and without facial expressions (Model vs. Non).

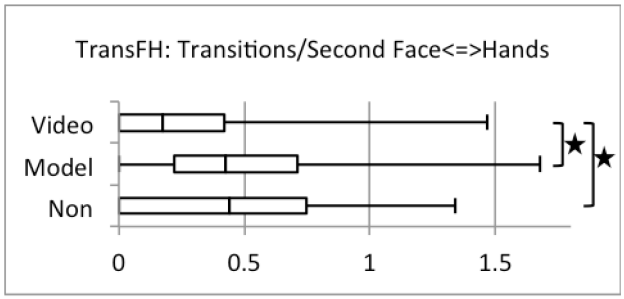

Fig. 5: Transition Frequency Between the Face and Hands

Given that H1 was supported, when we are examining the results for hypotheses H3-H5, it is logical to consider the results for videos separately from animations. However, we will group the Model and Non videos together during the correlation analysis, since H2 was not supported. Table 1 displays the Spearman's *Rho* correlation values for FacePFT and TransFH. Values for which the p(uncorrelated) value is below 0.05 have been marked with an asterisk; the *Rho* is shown between the eye-metric and each of the responses recorded during the experiment:

- Likert-scale subjective responses for whether the story was: grammatical, understandable, and had natural movement;
- Likert-scale response as to whether the participant noticed the particular facial expression in that story; and
- participant's accuracy on the comprehension questions.

Table 1: Correlations between Eye Metrics and Responses

Spearman's Rho (* if p < 0.05)	FacePFT Video	FacePFT Anim.	TransFH Video	TransFH Anim.
Grammatical	0.008	* -0.271	0.235	* -0.295
Understandable	-0.033	* -0.336	0.260	-0.160
Natural Movement	0.083	* -0.473	* 0.329	* -0.227
Notice Face Expr.	0.019	-0.103	0.218	-0.094
Comprehension	0.023	-0.063	0.003	-0.084

Hypothesis H3 was supported for animations: there was a correlation between the subjective evaluation questions and the eye metrics. Specifically, FacePFT was significantly correlated with all three subjective scores, and TransFH, with grammaticality and naturalness of movement. The H3 results for videos are inconclusive: while TransFH was significantly correlated with naturalness of movement, it was not significant for the other two.

H4 was not supported by our results: none of the correlations were significant between eye metrics and the responses to the question about whether participants noticed the facial expression.

H5 was not supported by our results: There was no significant correlation between the eye metrics we examined and the accuracy of participants on comprehension questions.

6. CONCLUSIONS AND FUTURE WORK

This paper provides accessibility researchers with methodological guidance on the use of eye tracking in user-based experimental studies of sign language animations. By conducting a study in which native signers viewed short stories in ASL performed by either a human signer or an animated character, we quantified how participants' eye gaze was affected by the quality of the ASL video or animation that was displayed. We also quantified how particular eye metrics correlated with participants' responses to evaluation questions about ASL videos and animations.

The main contributions for future sign language animation researchers designing a user-based evaluation study are: (1) They can consider eye tracking as a complimentary form of measurement in their study. (2) They can use the results presented in this paper for comparison purposes to understand how to characterize eye metrics they obtain in their studies. (3) They can further investigate other eye-tracking metrics that might better capture native signers' eye-gazing for a different sign language.

Specifically, we examined five hypotheses in this study:

- **H1: supported.** When viewing videos, signers spend more time looking at the face and less frequently move their gaze between the face and body of the signer. This indicates that our two metrics (proportional fixation time on the face and transition frequency between face and body/hands) can be used to capture the difference between the model-synthesized ASL animation being evaluated in a study and a video recording of human signer (shown as an upper-baseline for comparison).
- **H2: not supported.** No significant difference was observed between the animations with and without facial expression in either the time signers spent looking at the face or the frequency of gaze transitions between the face and body. We speculate that the following might have had some effect:
  - The animations with facial expressions were the result of an overly simplistic synthesis model. If the face of the signer had better-quality facial expressions, then perhaps it would have been more useful for participants to look at it. Thus, perhaps the difference in quality between our Non and Model animations was too subtle to detect with these eye metrics.
  - There was no difference in the appearance of the signer in the animations with and without facial expressions. Perhaps participants mentally grouped these animations as being the "same" because an identical virtual human was used in both. Thus, if a participant saw a Non story with the virtual human, then they may conclude that he never moves his face. During a subsequent Model animation, the participant may not look at the face of this "same" virtual human, even though some facial expressions appear in those versions of the animation.
- **H3: supported for animations.** There was a significant correlation between the *subjective scores* (grammatical, understandable, natural) that native signers assigned to an animation and the time they spent looking at the face of the virtual human character. Further, there was a significant correlation between their grammaticality and naturalness subjective scores and the frequency of eye-gaze transitions between face and hands during the animation. These animation results for H3 may be the most useful finding in this paper for future researchers; this is the first published result that indicates a relationship between eye-tracker metrics and participants' subjective judgments of sign language animation quality.
- **H3: partially supported for videos.** There was a significant correlation only between the naturalness subjective scores that

native signers assign to a video recording of a human signer and the frequency of eye-gaze transitions between the face and the hands of the signer. While prior researchers observed some eye metric differences for different types of sign language video [27], none had correlated subjective ratings of those videos with eye-metrics. It could be the case that no relationship exists, or the set of videos shown in this study may have been too homogenous in their quality (all of them contained facial expressions, with an identical signer, and identical camera angle). This homogeneity in video quality may have limited our ability to detect a correlation in this study.

- **H4: not supported for video or animations.** No significant correlation was observed between the participants reporting having noticed a facial expression and their eye-gazing behavior (as measured by the proportional fixation time on the face or the eye-gaze transition frequency between the face and hands). Perhaps there is no relationship between these two eye metrics and whether a participant consciously notices a facial expression; alternatively, asking participants to respond to a Likert-scale question about how confident they are that they noticed a particular facial expression is an ineffective way to measure this response. In future work, we may explore alternative approaches to asking about this information.
- **H5: not supported for video or animations.** No significant correlation was observed between the native signers' correctly answering comprehension questions and their eye movement behavior (as measured by the proportional fixation time on the face or the eye-gaze transition frequency between the face and hands). This suggests that these eye-tracking metrics cannot be used as an alternative form of measurement in evaluating the comprehensibility of synthesized ASL animations. As with H4, this may be due to a lack of relationship for these particular eye metrics, or our comprehension questions may be a poor measure of participant's understanding of the animations and videos. In [16, 20], we discuss the difficulty in designing comprehension questions to evaluate facial expressions.

In short, the results presented in this paper indicate that eye tracking analysis is valid for use as a complimentary form of measurement in a user-study to evaluate animations of sign language. Researchers who are studying computer graphics issues relating to the appearance of a virtual human for sign language animations and who are interested in obtaining participants' responses to subjective evaluations of the animation-quality may use eye-gazing metrics as an alternative form of measurement. This may be useful in experimental contexts in which the researchers cannot (or prefer not) to interrupt participants with questions while they are viewing a sequence of ASL animations. Additionally, researchers could directly compare eye movement of the participants between videos (that would serve as an "ideal" of photorealism) and their animations. If researchers obtain eye metric results that are similar for both videos of human signers and for their ASL animations, this may serve as evidence that their ASL animations are high-quality.

Sign language animation researchers who are considering using eye-tracking approaches with deaf users should consider some practical issues: First, they should minimize the need for the participants to look away from the screen during the experiment, to reduce eye tracking data loss and promote accuracy. Unlike hearing subjects who may ask questions and receive answers without taking their eyes off the computer screen, deaf participants would need to look away from the stimulus to communicate with the researcher. We recommend: (i) embedding the instructions and the questionnaires in the stimuli application, (ii) familiarizing the participants with a sample case initially, and (iii) positioning the ASL-signing researcher giving instructions to the participant opposite to the participant and behind the screen. If the researcher is at the participant's side, the participant may tend to shift their head towards the researcher occasionally during the study, to monitor for communication or confirmation. Second, a delicate balance is needed when selecting the size of the video/animation on the computer screen. While a bigger video/animation permits for fine-grained (distinct) AOIs, the participant should be able to see the human/animated character in full, without the need for head movements. Also, when stimuli are so large that they approach the edges of the computer screen, there can be a loss in eye-tracker accuracy: when a participant's eye is rotated farther from its neutral position, some eye-trackers "lose" the pupil or see reflection artifacts on the white sclera of the eye.

In future work, we want to further examine why there were no significant difference in the eye movements for animation with and without facial expressions (H2). Was it because (i) the model was not sophisticated enough to provide good facial expressions that would capture the eye gaze of the participants, (ii) there was no obvious difference in the appearance of the animations with and without facial expressions and they could have been mentally grouped as one character or because (iii) the stories were two short to allow for a significantly distinct eye movement behavior? A follow-up study could disambiguate this.

We are also interested in determining whether the very act of asking certain types of questions during a study could have an effect on the eye tracking results (e.g. asking participants about the facial expressions might cause them to look to the face more). While we did display each story before asking questions about it in this study, our participants would have noticed that they were always asked the same subjective and notice questions (the comprehension questions differed). So, unfortunately, we did not address this issue in the current paper. To be more confident that the participants are looking at the videos naturally and without prompting, in future work, we would need to display all the stimuli first and then allow participants to re-play them in order to respond to the questions. By replicating such a study, we could determine if the use of questions has an effect on the eye metrics.

7. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under award number 0746556 and 1065009. We acknowledge support from PSC CUNY Research Awards, Siemens A&D UGS PLM Software and Visage Technologies AB. Pengfei Lu helped to prepare stimuli used in this study; Jonathan Lamberton and Miriam Morrow helped to recruit participants.

8. REFERENCES

[1] Applied Science Labs. 2013. Homepage. http://www.asleyetracking.com

[2] Bartels, M., Marshall, S.P., 2006. Eye tracking insights into cognitive modeling. In *Proceedings of the Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (San Diego, California), ACM.

[3] Cavender, A.C, Bigham, J.P., Ladner, R.E. 2009. ClassInFocus: enabling improved visual attention strategies for deaf and hard of hearing students. In *Proc. ASSETS'09.* ACM, New York, NY, USA, 67-74.

[4] Cavender, A.C., Rice, E.A., Wilamowska, K.M. 2005. SignWave: Human perception of sign language video quality as constrained by mobile phone technology. Retrieved from:

http://www.cs.washington.edu/education/courses/cse510/05sp/project-reports/cse510-signwave.pdf

[5] Chapdelaine, C., Gouaillier, V., Beaulieu, M., Gagnon, L., 2007. Improving video captioning for deaf and hearing-impaired people based on eye movement and attention overload, In *Proc. SPIE* 6492, Human Vision and Electronic Imaging XII, 64921K.

[6] Duchowski A., 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4 (November 1, 2002), 455-470.

[7] Emmorey, K., Thompson, R., Colvin, R. 2009. Eye gaze during comprehension of American Sign Language by native and beginning signers. *J Deaf Stud Deaf Educ* 14, 2, 237-43.

[8] Goldberg, J. H., Stimson, M.J., Lewenstein, M., Scott, N., Wichansky, A.M. 2002. Eye tracking in web search tasks: design implications. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (New Orleans, Louisiana)*, ACM, 507082, 51-58.

[9] Goldberg, J.H., Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. *Int J Ind Ergonom* 24, 6, 631-645.

[10] Guan, Z., Cutrell, E. 2007. An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. (San Jose, California, USA), ACM.

[11] Halverson, T., Hornof, A. J. 2007. A minimal model for predicting visual search in human-computer interaction. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI '07).* ACM, New York, NY, USA, 431-434.

[12] Huenerfauth, M. 2004. Spatial and planning models of ASL classifier predicates for machine translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2004).*

[13] Huenerfauth, M. 2008. Evaluation of a psycholinguistically motivated timing model for animations of American Sign Language. *In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08). ACM, New York, NY, USA, 129-136.*

[14] Huenerfauth, M., Hanson, V. 2009. Sign language in the interface: access for deaf signers. In C. Stephanidis (ed.), *Universal Access Handbook*. NJ: Erlbaum. 38.1-38.18.

[15] Huenerfauth, M., Lu, P. 2012. Effect of spatial reference and verb inflection on the usability of American sign language animation. In *Univ Access Inf Soc*. Berlin: Springer.

[16] Huenerfauth, M., Lu, P., and Rosenberg, A. 2011. Evaluating Importance of Facial Expression in American Sign Language and Pidgin Signed English Animations. In *Proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '11)*. ACM, New York, NY, USA, 99-106.

[17] Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J. 2008. Evaluation of American sign language generation by native ASL signers. *ACM Trans Access Comput* 1(1):1-27.

[18] Jacob, R. J. K., Karn, K. S. 2003. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. The Mind's Eye (First Edition). J. Hyönä, R. Radach and H. Deubel. Amsterdam: 573-605.

[19] Jensen, S.S., Pedersen, T. 2011. Eye tracking deaf people's metatcognitive comprehension strategies on the internet. (Master's thesis). Retrieved from http://projekter.aau.dk/projekter/files/52662579/SamletProjekt.pdf

[20] Kacorri, H., Lu, P., Huenerfauth, M. 2013. Evaluating Facial Expressions in American Sign Language Animations for Accessible Online Information. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction (UAHCI).* Las Vegas, NV, USA.

[21] Kowler, E. 2011. Eye movements: The past 25years. *Vision Research*, 51(13), 1457-1483.

[22] Lu, P., Huenerfauth, M. 2009. Accessible Motion-Capture Glove Calibration Protocol for Recording Sign Language Data from Deaf Subjects. In *Proceedings of The 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009),* Pittsburgh, PA, USA. New York: ACM Press, 83-90.

[23] Lu, P., Huenerfauth, M. 2010. Collecting a motion-capture corpus of American Sign Language for data-driven generation research. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 89-97.

[24] Lu, P., Kacorri, H. 2012. Effect of Presenting Video as a Baseline During an American Sign Language Animation User Study. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '12).* ACM, New York, NY, USA, 183-190.

[25] Marschark M, Pelz J, Convertino C, Sapere P, Arndt ME, Seewagen R. 2005. Classroom interpreting and visual information processing in mainstream education for deaf students: Live or Memorex®? *Am Educ Res J* 42:727–762.

[26] Mitchell, R., Young, T., Bachleda, B., & Karchmer, M. 2006. How many people use ASL in the United States? Why estimates need updating. *Sign Lang Studies*, 6(3):306-335.

[27] Muir, L. J., & Richardson, I. E. 2005. Perception of sign language and its application to visual communications for deaf people. *J Deaf Stud Deaf Educ* 10, 4, 390-401.

[28] Neidle, C., D. Kegl, D. MacLaughlin, B. Bahan, R.G. Lee. 2000. *The syntax of ASL: functional categories and hierarchical structure.* Cambridge: MIT Press.

[29] Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124(3): 372-422.

[30] Rayner, K. 2009. Eye Movements in Reading: Models and Data. *Journal of Eye Movement Research*, 2(5):2, 1-10.

[31] Salvucci, D.D., Anderson, J.R. 2001. Automated eye-movement protocol analysis. *Hum-Comput Interact* 16(1), 39-86.

[32] Traxler, C. 2000. The Stanford achievement test, 9th edition: national norming and performance standards for deaf & hard-of-hearing students. *J Deaf Stud & Deaf Edu* 5(4):337-348.

[33] VCOM3D. 2013. Homepage. http://www.vcom3d.com/

[34] Watanabe K, Matsuda T, Nishioka T, Namatame M. 2011. Eye Gaze during Observation of Static Faces in Deaf People. *PLoS ONE* 6(2): e16919.