Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research

Pengfei Lu
Department of Computer Science
Graduate Center
City University of New York (CUNY)
365 Fifth Ave, New York, NY 10016
pengfei.lu@qc.cuny.edu

Matt Huenerfauth
Department of Computer Science
Queens College and Graduate Center
City University of New York (CUNY)
65-30 Kissena Blvd, Flushing, NY 11367
matt@cs.qc.cuny.edu

Abstract

American Sign Language (ASL) generation software can improve the accessibility of information and services for deaf individuals with low English literacy. The understandability of current ASL systems is limited; they have been constructed without the benefit of annotated ASL corpora that encode detailed human movement. We discuss how linguistic challenges in ASL generation can be addressed in a data-driven manner, and we describe our current work on collecting a motion-capture corpus. To evaluate the quality of our motion-capture configuration, calibration, and recording protocol, we conducted an evaluation study with native ASL signers.

1 Introduction

American Sign Language (ASL) is the primary means of communication for about one-half million deaf people in the U.S. (Mitchell et al., 2006). ASL has a distinct word-order, syntax, and lexicon from English; it is not a representation of English using the hands. Although reading is part of the curriculum for deaf students, lack of auditory exposure to English during the language-acquisition years of childhood leads to lower literacy for many adults. In fact, the majority of deaf high school graduates in the U.S. have only a fourth-grade (age 10) English reading level (Traxler, 2000).

1.1 Applications of ASL Generation Research

Most technology used by the deaf does not address this literacy issue; many deaf people find it difficult to read the English text on a computer screen or on a television with closed-captioning. Software to present information in the form of animations of ASL could make information and services more accessible to deaf users, by displaying an animated character performing ASL, rather than English text. While writing systems for ASL have been proposed (Newkirk, 1987; Sutton, 1998), none is widely used in the Deaf community. Thus, an ASL generation system cannot produce text output; the system must produce an animation of a human character performing sign language. Coordinating the simultaneous 3D movements of parts of an animated character's body is challenging, and few researchers have attempted to build such systems.

Prior work can be divided into two areas: scripting and generation/translation. Scripting systems allow someone who knows sign language to "word process" an animation by assembling a sequence of signs from a lexicon and adding facial expressions. The eSIGN project created tools for content developers to build sign databases and assemble scripts of signing for web pages (Kennaway et al., 2007). Sign Smith Studio (Vcom3D, 2010) is a commercial tool for scripting ASL (discussed in section 4). Others study generation or machine translation (MT) of sign language (Chiu et al., 2007; Elliot & Glauert, 2008; Fotinea et al., 2008; Huenerfauth, 2006; Karpouzis et al., 2007; Marshall & Safar, 2005; Shionome et al., 2005; Sumihiro et al., 2000; van Zijl & Barker, 2003).

Experimental evaluations of the understandability of state-of-the-art ASL animation systems have shown that native signers often find animations difficult to understand (as measured by compre-

Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies, pages 89–97,
Los Angeles, California, June 2010.

hension questions) or unnatural (as measured by subjective evaluation questions) (Huenerfauth et al., 2008). Errors include a lack of smooth inter-sign transitions, lack of grammatically-required facial expressions, and inaccurate sign performances related to morphological inflection of signs.

While current ASL animation systems have limitations, there are several advantages in presenting sign language content in the form of animated virtual human characters, rather than videos:

- Generation or MT software planning ASL sentences cannot just concatenate videos of ASL. Using video clips, it is difficult to produce smooth transitions between signs, subtle motion variations in sign performances, or proper combinations of facial expressions with signs.
- If content must be frequently modified or updated, then a video performance would need to be largely re-recorded for each modification. Whereas, an animation (scripted by a human author) could be further edited or modified.
- Because the face is used to indicate important information in ASL, a human must reveal his or her identity when producing an ASL video. Instead, a virtual human character could perform sentences scripted by a human author.
- For wiki-style applications in which multiple authors are collaborating on information content, ASL videos would be distracting: the person performing each sentence may differ. A virtual human would be more uniform.
- Animations can be appealing to children for use in educational applications.
- Animations allow ASL to be viewed at different angles, at different speeds, or by different virtual humans – depending on the preferences of the user. This can enable education applications in which students learning ASL can practice their ASL comprehension skills.

1.2 ASL is Challenging for NLP Research

Natural Language Processing (NLP) researchers often apply techniques originally designed for one language to another, but research is not commonly ported to sign languages. One reason is that without a written form for ASL, NLP researchers must produce animation and thus address several issues:

- *Timing:* An ASL performance's speed consists of: the speed of individual sign performances, the transitional time between signs, and the insertion of pauses during signing – all of which are based on linguistic factors such as syntactic boundaries, repetition of signs in a discourse, and the part-of-speech of signs (Grosjean et al., 1979). ASL animations whose speed and pausing are incorrect are significantly less understandable to ASL signers (Huenerfauth, 2009).
- *Spatial Reference:* Signers arrange invisible placeholders in the space around their body to represent objects or persons under discussion (Meier, 1990). To perform personal, possessive, or reflexive pronouns that refer to these entities, signers later point to these locations. Signers may not repeat the identity of these entities again; so, their conversational partner must remember where they have been placed. An ASL generator must select which entities should be assigned 3D locations (and where).
- *Inflection:* Many verbs change their motion paths to indicate the 3D location where a spatial reference point has been established for their subject, object, or both (Padden, 1988). Generally, the motion paths of these inflecting verbs change so that their direction goes from the subject to the object (Figure 1); however, their paths are more complex than this. Each verb has a standard motion path that is affected by the subject's and the object's 3D locations. When a verb is inflected in this way, the signer does not need to overtly state the subject/object of a sentence. An ASL generator must produce appropriately inflected verb paths based on the layout of the spatial reference points.

(a.) (b.)

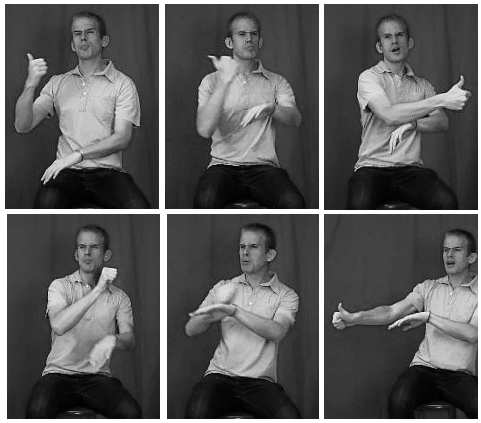


Figure 1: An ASL inflecting verb "BLAME": (a.) (person on left) blames (person on right), (b.) (person on right) blames (person on left).

- *Coarticulation:* As in speech production, the surrounding signs in a sentence affect finger, hand, and body movements. ASL generators that use overly simple interpolation rules to produce these coarticulation effects yield unnatural and non-fluent ASL animation output.
- *Non-Manuals:* Head-tilt and eye-gaze indicate the 3D location of a verb's subject and object (or other information); facial expressions also indicate negation, questions, topicalization, and other essential syntactic phenomena not conveyed by the hands (Neidle et al., 2000). Animations without proper facial expressions (and proper timing relative to manual signs) cannot convey the proper meaning of ASL sentences in a fluent and understandable manner.
- *Evaluation:* With no standard written form for ASL, string-based metrics cannot be used to evaluate ASL generation output automatically. User-based experiments are necessary, but it is difficult to accurately: screen for *native* signers, prevent English environmental influences (that affect signer's linguistic judgments), and design questions that measure comprehension of ASL animations (Huenerfauth et al., 2008).

1.3 Need for Data-Driven ASL Generation

Due to these challenges, most prior sign language generation or MT projects have been short-lived, producing few example outputs (Zhao et al., 2000; Veale et al., 1998). Further developed systems also have limited coverage; e.g., Marshall and Safar (2005) hand-built translation transfer rules from English to British Sign Language. Huenerfauth (2006) surveys several rule-based systems and discusses how they generally: have limited coverage; often merely concatenate signs; and do not address the *Coarticulation, Spatial Reference, Timing*, *Non-Manuals,* or *Inflection* issues (section 1.2).

Unfortunately, most prior work is not "data-driven," i.e. not based on statistical modeling of corpora, the dominant successful modern NLP approach. The sign language generation research that has thus far been *the most data-driven* includes:

- Some researchers have used motion-capture (see section 3) to build lexicons of animations of individual signs, e.g. (Cox et al., 2002). However, their focus is recording a single citation form of each sign, not creating annotated corpora of full sentences or discourse. Single-sign recordings do not enable researchers to examine the *Timing, Coarticulation, Spatial Reference, Non-Manuals*, or *Inflection* phenomena (section 1.2), which operate over multiple signs or sentences in an ASL discourse.
- Other researchers have examined how statistical MT techniques could be used to translate from a written language to a sign language. Morrissey and Way (2005) discuss an example-based MT architecture for Irish Sign Language, and Stein et al. (2006) apply simple statistical MT approaches to German Sign Language. Unfortunately, the sign language "corpora" used in these studies consist of transcriptions of the sequence of signs performed, not recordings of actual human performances. A transcription does not capture subtleties in the 3D movements of the hands, facial movements, or speed of an ASL performance. Such information is needed in order to address the *Spatial Reference, Inflection, Coarticulation, Timing*, or *Non-Manuals* issues (section 1.2).
- Seguoat and Braffort (2009) derive models of coarticulation for French Sign Language based on a semi-automated "rotoscoping" annotation of hand location from videos of signing.

1.4 Prior Sign Language Corpora Resources

The reason why most prior ASL generation research has not been data-driven is that sufficiently detailed and annotated sign language corpora are in short supply and are time-consuming to construct. Without a writing system in common use, it is not possible to harvest some naturally arising source of ASL "text"; instead, it is necessary to record the performance of a signer (through video or a motion-capture suit). Human signers must then transcribe and annotate this data by adding time-stamped linguistic details. For ASL (Neidle et al., 2000) and European sign languages (Bungeroth et al., 2006; Crasborn et al., 2004, 2006; Efthimiou & Fotinea, 2007), signers have been videotaped and experts marked time spans when events occur – e.g. the right hand is performing the sign "CAT" during time index 250-300 milliseconds, and the eyebrows are raised during time index 270-300. Such annotation is time-consuming to add; the largest ASL corpus has a few thousand sentences.

In order to learn how to control the movements of an animated virtual human based on a corpus,

we need precise hand locations and joint angles of the human signer's body throughout the performance. Asking humans to write down 3D angles and coordinates is time-consuming and inexact; some researchers have used computer vision techniques to model the signers' movements (see survey in (Loeding et al., 2004)). Unfortunately, the complex shape of hands/face, rapid speed, and frequent occlusion of parts of the body during ASL limit the accuracy of vision-based recognition; it is not yet a reliable way to build a 3D model of a signer for a corpus. Motion-capture technology (discussed in section 3) is required for this level of detail.

2 Research Goals & Focus of This Paper

To address the lack of sufficiently detailed and linguistically annotated ASL corpora, we have begun a multi-year project to collect and annotate a motion-capture corpus of ASL (section 3). Digital 3D body movement and handshape data collected from native signers will become a permanent research resource for study by NLP researchers and ASL linguists. This corpus will allow us to create new ASL generation technologies in a data-driven manner by analyzing the subtleties in the motion data and its relationship to the linguistic structure. Specifically, we plan to model where signers tend to place spatial reference points around them in space. We also plan to uncover patterns in the motion paths of inflecting verbs and model how they relate to layout of spatial references points. These models could be used in ASL generation software or could be used to partially automate with work of humans using ASL-scripting systems. To evaluate our ASL models, native signers will be asked to judge ASL animations produced using them. There are several unique aspects of our research:

- We use a novel combination of hand, body, head, and eye motion-tracking technologies and simultaneous video recordings (section 3).
- We collect multi-sentence single-signer ASL discourse, and we annotate novel linguistic information (relevant to spatial reference points).
- We involve ASL signers in the research in several ways: as evaluators of our generation software, as research assistants conducting evaluation studies, and as corpus annotators.

This paper will focus on the first of these aspects of our project. Specifically, section 4 will examine the following research question: *Have we successfully configured and calibrated our motion-capture equipment so that we are recording good-quality data that will be useful for NLP research?*

Since the particular combination of motion-capture equipment we are using is novel and because there have not been prior motion-capture-based ASL corpora projects, section 4 will evaluate whether the data we are collecting is of sufficient quality to drive ASL animations of a virtual human character. In corpus-creation projects for traditional written/spoken languages, researchers typically gather text, audio, or (sometimes) video of human performances. The quality of the gathered recordings is typically easier to verify and evaluate; for motion-capture data collected with a complex configuration of equipment, a more complex experimental design is necessary (section 4).

3 Our Motion-Capture Configuration

The first stage of our research is to accurately and efficiently record 3D motion-capture data from ASL signers. Assuming an ASL signer's pelvis bone is stationary in 3D space, we want to record movement data for the upper body. We are interested in the shapes of each hand; the 3D location of the hands; the 3D orientation of the palms; joint angles for the wrists, elbows, shoulders, clavicle, neck, and waist; and a vector representing the eye-gaze aim. We are using a customized configuration of several commercial motion-capture devices (as shown in Figure 2, worn by a human signer):

- Two Immersion CyberGloves®: The 22 flexible sensor strips sewn into each of these spandex gloves record finger joint angles so that we can record the signer's handshapes. These gloves are ideal for recording ASL because they are flexible and lightweight. Humans viewing a subject wearing the gloves are able to discern ASL fingerspelling and signing.
- Applied Science Labs H6 eye-tracker: This lightweight head-mounted eye-tracker with a near-eye camera records a signer's eye gaze direction. A camera on the headband aims down, and a small clear plastic panel in front of the cheek reflects the image of the subject's eye. When combined with the head tracking information from the IS-900 system below, the H6 identifies a 3D vector of eye-gaze in a room.

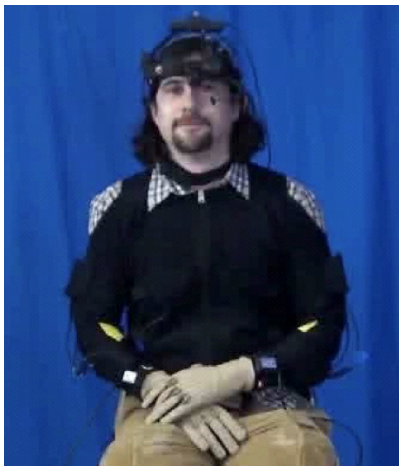
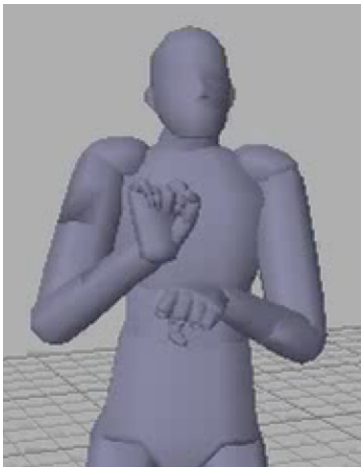
a.  b.  c. 

Figure 2: (a) Motion-capture equipment configuration, (b) animation produced from motion-capture data (shown in evaluation study), and (c) animation produced using Sign Smith (shown in evaluation study).

- Intersense IS-900: This acoustical/intertial motion-capture system uses a ceiling-mounted ultrasonic speaker array and a set of directional microphones on a small sensor to record the location and orientation of the signer's head. A sensor sits atop the helmet shown in Figure 2a. IS-900 data is used to compensate for head movement when calculating eye-gaze direction with the Applied Science Labs H6 eye-tracker.
- Animazoo IGS-190: This spandex bodysuit is covered with soft Velcro to which small sensors attach. A sensor placed on each segment of the human's body records inertial and magnetic information. Subjects wearing the suit stand facing north with their arms down at their sides at the beginning of the recording session; given this known starting pose and direction, the system calculates joint angles for the wrists, elbows, shoulders, clavicle, neck, and waist. We do not record leg/foot information in our corpus. Prior to recording data, we photograph subjects standing in a cube-shaped rig of known size; this allows us to identify bone lengths of the human subject, which are needed for the IGS-190 system to accurately calculate joint angles from the sensor data.

Motion-capture recording sessions are videotaped to facilitate later linguistic analysis and annotation. Videotaping the session also facilitates the "clean up" of the motion-capture data in post-processing, during which algorithms are applied to adjust synchronization of different sensors or remove "jitter" or other noise artifacts from the recording. Three digital high-speed video cameras film front view, facial close-up, and side views of the signer – a setup that has been used in video-based ASL-corpora-building projects (Neidle et al., 2000). The front view is similar to Figure 2a (but wider). The facial close-up view is useful when later identifying specific non-manual facial expressions during ASL performances, which are essential to correctly understanding and annotating the collected data. To facilitate synchronizing the three video files during post-processing, a strobe is flashed once at the start of the recording session.

A "blue screen" curtain hangs on the back and side walls of the motion-capture studio. If future computer vision researchers wish to use this corpus to study ASL recognition from video, it is useful to have solid color walls for "chroma key" background removal. Photographic studio lighting with spectra compatible with the eye-tracking system is used to support high-quality video recording.

During data collection, a native ASL signer (called the "prompter") sits directly behind the front-view camera to engage the participant wearing the suit (the "performer") in natural conversation. While the corpus we are collecting consists of unscripted *single-signer* discourse, prior ASL corpora projects have identified the importance of surrounding signers with an ASL-centric environment during data collection (Neidle et al., 2000). English influence in the studio must be minimized to prevent signers from inadvertently code-switching to an English-like form of signing. Thus, it is important that a native signer acts as the prompter, who conversationally suggests topics for the performer to discuss (to be recorded as part of the corpus).

In our first year, we have collected and annotated 58 passages from 6 signers (40 minutes). We prefer to collect multi-sentence passages discussing varied numbers of topics and with few "classifier predicates," phenomena that aren't our current research focus. In (Huenerfauth & Lu, 2010), we discuss details of: the genre of discourse we record, our target linguistic phenomena to capture (spatial reference points and inflected verbs), the types of linguistic annotation added to the corpus, and the effectiveness of different "prompts" used to elicit the desired type of spontaneous discourse.

This paper focuses on verifying the quality of the motion-capture data we can record using our current equipment configuration and protocols. We want to measure how well we have compensated for several possible sources of error in recordings:

- If a sensor connection is temporarily lost, then data gaps occur. We have selected equipment that does not require line-of-sight connections and tried to arrange the studio to avoid frequent dropping of any wireless connections.
- We ask subjects to perform a quick head movement and distinctive eye blink pattern at the beginning of the recording session to facilitate "synchronization" of the various motion-capture data streams during post-processing.
- Electronic and physical properties of sensors can lead to "noise" in the data, which we attempt to remove with smoothing algorithms.
- Differences between the bone lengths of the human and the "virtual skeleton" of the animated character being recorded could lead to "retargeting" errors, in which the body poses of the human do not match the recording. We must be careful in the measurement of the bone lengths of the human participant and in the design of the virtual animation skeleton.
- To compensate for differences in how equipment sits on the body on different occasions or on different humans, we must set "calibration" values; e.g., we designed a novel protocol for efficiently and accurately calibrating gloves for ASL signers (Lu & Huenerfauth, 2009).

4 Evaluating Our Collected Motion Data

If a speech synthesis researcher were using a novel microphone technology to record audio performances from human speakers to build a corpus, that researcher would want to experimentally confirm that the audio recordings were of high enough quality for research. Even when perfectly clear audio recordings of human speech are recorded in a corpus, the automatic speech synthesis models trained on this data are not perfect. Degradations in the quality of the corpus would yield even lower quality speech synthesis systems. In the same way, it is essential that we evaluate the quality of the ASL motion-capture data we are collecting.

In an earlier study, we sought to collect motion-data from humans and directly produce animations from them as an "upper baseline" for an experimental study (Huenerfauth, 2006). We were not analyzing the collected data or using it for data-driven generation, we merely wanted the data to directly drive an animation of a virtual human character as a "virtual puppet." This earlier project used a different configuration of motion-capture equipment, including an earlier version of CyberGloves® and an optical motion-capture system that required line-of-sight connections between infrared emitters on the signer's body and cameras around the room. Unfortunately, the data collected was so poor that the animations produced from the motion-capture were not an "upper" baseline – in fact, they were barely understandable to native signers. Errors arose from dropped connections, poor calibration, and insufficient removal of data noise.

We have selected different equipment and have designed better protocols for recording high quality ASL data since that earlier study – to compensate for the "noise," "retargeting," "synchronization," and "calibration" issues mentioned in section 3. However, we know that under some recording conditions, the quality of collected motion-capture data is so poor that "virtual puppet" animations synthesized from it are not understandable. We expect that an even higher level of data quality is needed for a motion-capture *corpus*, which will be analyzed and manipulated in order to synthesize novel ASL animations from it. Therefore, we conducted a study (discussed below) to evaluate the quality of our current motion-capture configuration. As in our past study, we use the motion-capture data to directly control the body movements of a virtual human "puppet." We then ask native ASL signers to evaluate the understandability and naturalness of the resulting animations (and compare them to some baseline animations produced using ASL-animation scripting software).

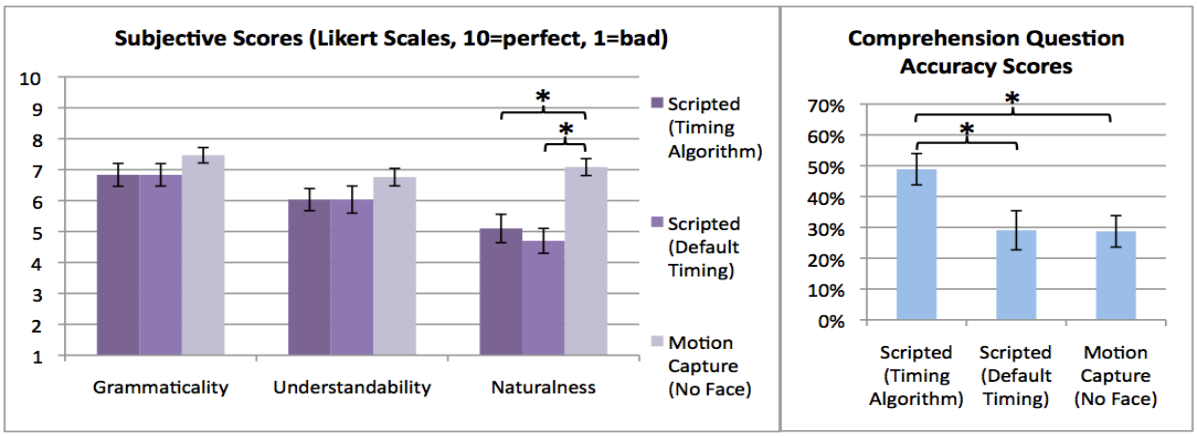


Figure 3: Evaluation and comprehension scores (asterisks mark significant pairwise differences).

In our prior work, a native ASL signer designed a set of ASL stories and corresponding comprehension questions for use in evaluation studies (Huenerfauth, 2009). The stories' average length is approximately 70 signs, and they consist of news stories, encyclopedia articles, and short narratives. We produced animations of each using Sign Smith Studio (SSS), commercial ASL-animation scripting software (Vcom3D, 2010). Signs from SSS's lexicon are placed on a timeline, and linguistically appropriate facial expressions are added. The software synthesizes an animation of a virtual human performing the story (Figure 2c). In earlier work, we designed algorithms for determining sign-speed and pause-insertion in ASL animations based on linguistic features of the sentence. We conducted a study to compare animations with default timing settings (uniform pauses and speed) and animations governed by our timing algorithm – at various speeds. The use of our timing algorithm yielded ASL animations that native signers found more understandable (Huenerfauth, 2009). We are reusing these stories and animations as baselines for comparison in a new evaluation study (below).

While we are collecting *unscripted* passages in our corpus, it is easier to compare the quality of different versions of animations when using a common set of *scripted* stories. Thus, we used the script from 10 of the stories above, and each was performed by a native signer, a 22-year-old male who learned ASL prior to age 2. He wore the full set of motion-capture equipment, and we followed the same calibration process and protocols as we do when recording ASL passages for our corpus. The signer rehearsed and memorized each story; "cue cards" were also available when recording.

Autodesk MotionBuilder software was used to produce a virtual human whose movements were driven by the motion-capture data (see Figure 2b). While our corpus contains *video* of facial expression, our motion-capture equipment does not digitize it; so, the virtual human character has no facial movements. The recorded signer moved at an average speed of 1.12 signs/second, and so for comparison, we selected the version of the scripted ASL animations with the closest speed from our earlier study: 1.2 signs/second. (Since the scripted animations are slightly slower and include linguistic facial expressions, we expected them to receive higher understandability scores than our motion-capture animations.) In our earlier work, we produced two versions of each scripted story: one with default timing and one with our novel timing algorithm. Both versions are used as baselines for comparison in this new study; thus, we compare three versions of the same set of 10 ASL stories.

Using questions designed to screen for native ASL signers developed in prior work (Huenerfauth et al., 2008), we recruited 12 participants to evaluate the ASL animations. A native ASL signer conducted the studies, in which participants viewed an animation and were then asked two types of questions after each: (1) ten-point Likert-scale questions about the ASL animation's grammatical correctness, understandability, and naturalness of movement and (2) multiple-choice comprehension questions about basic facts from the story. The comprehension questions were presented in the form of scripted ASL animations (produced in SSS), and answer choices were presented in the form of clip-art images (so that strong English literacy was not necessary). Identical questions were

used to evaluate the motion-capture animations and the scripted animations. Examples of the questions are included in (Huenerfauth, 2009).

Figure 3 displays results of the Likert-scale subjective questions and comprehension-question success scores for the three types of animations evaluated in this study. The scripted animations using our timing algorithm have higher comprehension scores, but the motion-capture animations have higher naturalness scores. All of the other scores for the animations are quite similar. Statistically significant differences are marked with an asterisk ($p<0.05$, Mann-Whitney pairwise comparisons with Bonferroni-corrected p-values). Non-parameteric tests were selected because the Likert-scale responses were not normally distributed.

5 Conclusion and Future Research Goals

The research question addressed by this paper was whether our motion-capture configuration and recording protocols enabled us to collect motion-data of sufficient quality for data-driven ASL generation research. In our study, the evaluation scores of the animations driven by the motion-capture data were similar to those of animations produced using state-of-the-art ASL animation scripting software. This is a promising result, especially considering the slightly faster speed and lack of facial expression information in the motion-capture animations. While this suggests that the data we are collecting is of good quality, the *real* test will be when this corpus is used in future research. If we can build useful ASL-animation generation software based on analysis of this corpus, then we will know that we have sufficient quality of motion-capture data.

5.1 Our Long-Term Research Goal: Making ASL Accessible to More NLP Researchers

It is our goal to produce high-quality broad-coverage ASL generation software, which would benefit many deaf individuals with low English literacy. However, this ambition is too large for any one team; for this technology to become reality, ASL must become a language commonly studied by NLP researchers. For this reason, we seek to build ASL software, models, and experimental techniques to serve as a resource for other NLP researchers. Our goal is to make ASL "accessible" to the NLP community. By developing tools to address some of the modality-specific and spatial aspects of ASL, we can make it easier for other researchers to transfer their new NLP techniques to ASL. The goal is to "normalize" ASL in the eyes of the NLP community. Bridging NLP and ASL research will not only benefit deaf users: ASL will push the limits of current NLP techniques and will thus benefit other work in the field of NLP. Section 1.2 listed six challenges for ASL NLP research; we address several of these in our research:

We have conducted many experimental studies in which signers evaluate the understandability and naturalness of ASL animations (Huenerfauth et al., 2008; Huenerfauth, 2009). To begin to address the *Evaluation* issue (section 1.2), we have published best-practices, survey materials, and experimental protocols for effectively evaluating ASL animation systems through the participation of native signers. We have also published baseline comprehension scores for ASL animations. We will continue to produce such resources in future work.

Our earlier work on timing algorithms for ASL animations (mentioned in section 4) was based on data reported in the linguistics literature (Grosjean et al., 1979). In future work, we want to learn timing models directly from our collected corpus – to further address the *Timing* issue (section 1.2).

To address the issues of *Spatial Reference* and *Inflection* (section 1.2), we plan on analyzing our ASL corpus to build models that can predict where in 3D space signers establish spatial reference points. Further, we will analyze our corpus to analyze how certain ASL verbs are inflected based on the 3D location of their subject and object. We want to build a *parameterized* lexicon of ASL verbs: given a 3D location for subject and object, we want to predict a 3D motion-path for the character's hands for a specific performance of a verb.

While addressing the issues of *Coarticulation* and *Non-Manuals* (section 1.2) are not immediate research priorities, we believe our ASL corpus may also be useful in building computational models of these phenomena for data-driven ASL generation.

Acknowledgments

This material is based upon work supported by the National Science Foundation (Award #0746556), Siemens (Go PLM Academic Grant), and Visage Technologies AB (free academic license for software). Jonathan Lamberton, Wesley Clarke, Kelsey Gallagher, Amanda Krieger, and Aaron Pagan assisted with ASL data collection and experiments.

References

J. Bungeroth, D. Stein, P. Dreuw, M. Zahedi, H. Ney. 2006. A German sign language corpus of the domain weather report. *Proc. LREC 2006 workshop on representation & processing of sign languages.*

Y.H. Chiu, C.H. Wu, H.Y. Su, C.J. Cheng. 2007. Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Trans Pattern Anal Mach Intell* 29(1):28-39.

S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, S. Abbott. 2002. Tessa, a system to aid communication with deaf people. *Proc. ASSETS.*

O. Crasborn, E. van der Kooij, D. Broeder, H. Brugman. 2004. Sharing sign language corpora online: proposals for transcription and metadata categories. *Proc. LREC 2004 workshop on representation & processing of sign languages*, pp. 20-23.

O. Crasborn, H. Sloetjes, E. Auer, and P. Wittenburg. 2006. Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. *Proc. LREC 2006 workshop on representation & processing of sign languages,* 82-87.

E. Efthimiou, S.E. Fotinea. 2007. GSLC: creation and annotation of a Greek sign language corpus for HCI. *Proc. HCI International.*

R. Elliot, J. Glauert. 2008. Linguistic modeling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society* 6(4):375-391.

S.E. Fotinea, E. Efthimiou, G. Caridakis, K. Karpouzis. 2008. A knowledge-based sign synthesis architecture. *Univ. Access in Information Society* 6(4):405-418.

F. Grosjean, L. Grosjean, H. Lane. 1979. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology* 11:58-81.

M. Huenerfauth. 2006. Generating American sign language classifier predicates for English-to-ASL machine translation, dissertation, U. of Pennsylvania.

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

M. Huenerfauth. 2009. A linguistically motivated model for speed and pausing in animations of American sign language. *ACM Trans Access Comput* 2(2):1-31.

M. Huenerfauth, P. Lu. 2010. Annotating spatial reference in a motion-capture corpus of American sign language discourse. *Proc. LREC 2010 workshop on representation & processing of sign languages.*

K. Karpouzis, G. Caridakis, S.E. Fotinea, E. Efthimiou. 2007. Educational resources and implementation of a Greek sign language synthesis architecture. *Computers & Education* 49(1):54-74.

J. Kennaway, J. Glauert, I. Zwitserlood. 2007. Providing signed content on Internet by synthesized animation. *ACM Trans Comput-Hum Interact* 14(3):15.

B. Loeding, S. Sarkar, A. Parashar, A. Karshmer. 2004. Progress in automated computer recognition of sign language, *Proc. ICCHP*, 1079-1087.

P. Lu, M. Huenerfauth. 2009. Accessible motion-capture glove calibration protocol for recording sign language data from deaf subjects. *Proc. ASSETS.*

I. Marshall, E. Safar. 2005. Grammar development for sign language avatar-based synthesis. *Proc. UAHCI.*

R. Meier. 1990. Person deixis in American sign language. In: S. Fischer & P. Siple (eds.), *Theoretical issues in sign language research, vol. 1: Linguistics*. Chicago: University of Chicago Press, 175-190.

R. Mitchell, T. Young, B. Bachleda, M. Karchmer. 2006. How many people use ASL in the United States? *Sign Language Studies* 6(3):306-335.

S. Morrissey, A. Way. 2005. An example-based approach to translating sign language. *Proc. Workshop on Example-Based Machine Translation*, 109-116.

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

D. Newkirk. 1987. *SignFont Handbook*. San Diego: Emerson and Associates.

C. Padden. 1988. Interaction of morphology & syntax in American sign language. *Outstanding dissertations in linguistics, series IV*. New York: Garland Press.

J. Segouat, A. Braffort. 2009. Toward the study of sign language coarticulation: methodology proposal. *Proc Advances in Comput.-Human Interactions,* 369-374.

T. Shionome, K. Kamata, H. Yamamoto, S. Fischer. 2005. Effects of display size on perception of Japanese sign language---Mobile access in signed language. *Proc. Human-Computer Interaction*, 22-27.

D. Stein, J. Bungeroth, H. Ney. 2006. Morpho-syntax based statistical methods for sign language translation. *Proc. European Association for MT*, 169-177.

K. Sumihiro, S. Yoshihisa, K. Takao. 2000. Synthesis of sign animation with facial expression and its effects on understanding of sign language. *IEIC Technical Report* 100(331):31-36.

V. Sutton. 1998. The Signwriting Literacy Project. In *Impact of Deafness on Cognition AERA Conference*.

C. Traxler. 2000. The Stanford achievement test, ninth edition: national norming and performance standards for deaf and hard-of-hearing students. *J. Deaf Studies and Deaf Education* 5(4):337-348.

L. van Zijl, D. Barker. 2003. South African sign language MT system. *Proc. AFRIGRAPH*, 49-52.

VCom3D. 2010. Sign Smith Studio. http://www.vcom3d.com/signsmith.php

T. Veale, A. Conway, B. Collins. 1998. Challenges of cross-modal translation: English to sign translation in ZARDOZ system. *Machine Translation* 13:81-106.

L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, M. Palmer. 2000. A machine translation system from English to American sign language. *Proc. AMTA.*