Learning a Vector-Based Model of American Sign Language Inflecting Verbs from Motion-Capture Data

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) synthesis software can improve the accessibility of information and services for deaf individuals with low English literacy. The synthesis component of current ASL animation generation and scripting systems have limited handling of the many ASL verb signs whose movement path is inflected to indicate 3D locations in the signing space associated with discourse referents. Using motion-capture data recorded from human signers, we model how the motion-paths of verb signs vary based on the location of their subject and object. This model yields a lexicon for ASL verb signs that is parameterized on the 3D locations of the verb's arguments; such a lexicon enables more realistic and understandable ASL animations. A new model presented in this paper, based on identifying the principal movement vector of the hands, shows improvement in modeling ASL verb signs, including when trained on movement data from a different human signer.

1 Introduction

American Sign Language (ASL) is a primary means of communication for over 500,000 people in the U.S. (Mitchell et al., 2006). As a natural language that is not merely an encoding of English, ASL has a distinct syntax, word order, and lexicon. Someone can be fluent in ASL yet have significant difficulty reading English; in fact, due to various educational factors, the majority of deaf high school graduates (age 18+) in the U.S. have a fourth-grade (age 10) English reading level or lower (Traxler, 2000). This leads to accessibility challenges for deaf adults when faced with English text on computers, video captions, or other sources.

Technologies for automatically generating computer animations of ASL can make information and services accessible to deaf people with lower English literacy. While videos of sign language are feasible to produce in some contexts, animated avatars are more advantageous than video when the information content is often modified, the content is generated or translated automatically, or signers scripting a message in ASL wish to preserve anonymity. This paper focuses on ASL and producing accessible sign language animations for people who are deaf in the U.S., but many of the linguistic issues, literacy rates, and animation technologies discussed within are also applicable to other sign languages used internationally.

2 Use Of Space, Inflected Verbs

ASL signers can associate entities or concepts they are discussing with arbitrary locations in space (Liddle, 2003; Lillo-Martin, 1991; McBurney, 2002; Meier, 1990). After an entity is first mentioned, a signer may point to a 3D location in space around his/her body; to refer to this entity again, the signer (or his/her conversational partner) can point to this location. Many linguists have studied this pronominal use of space (Klima et al. 1979; Liddell, 2003; McBurney, 2002; Meier, 1990). Some argue that signers tend to pick 3D locations on a semi-circular arc floating at chest height in front of their torso (McBurney, 2002; Meier, 1990); others argue that signers pick 3D locations at different heights and distances from their body (Liddell, 2003). Regardless, there are an infinite number of locations where entities may be associated for pronominal reference; as discussed below, this also means that there are a potentially infinite number of ways for some verbs to be performed: a finite fixed lexicon for ASL is not sufficient.

NAACL-HLT 2012 Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), pages 66–74,
Montréal, Canada, June 7–8, 2012.

While ASL verbs have a standard citation form, many can be inflected to indicate the 3D location in space at which their subject and/or object have been associated (Liddell, 2003; Neidle et al., 2000; Padden, 1988). Linguists refer to such verbs as "inflecting" (Padden, 1988), "indicating" (Liddell, 2003), or "agreeing" verbs (Cormier, 2002). We use the term "inflecting verbs" in this paper. When they appear in a sentence, their standard motion path may be modified such that the movement or orientation goes from the 3D location of their subject and toward the 3D location of their object (or more complex effects). The resulting performance is a synthesis of the verb's standard lexical motion path and the 3D locations associated with the subject and object. Because the verb sign indicates its subject and/or object, the names of the subject and object may not be otherwise expressed in the sentence. If the signer chooses to mention them in the sentence, it is legal to use the citation-form (uninflected) version of the verb, but the resulting sentences tend to appear less fluent. In prior studies, we have found that native ASL signers who view ASL animations report that those that include spatially inflected verbs and entities associated with locations in space are easier to understand (than those which lack spatial pronominal reference and lack verb inflection) (Huenerfauth and Lu, 2012).

Fig. 1 shows the ASL verb EMAIL, which inflects for its subject and object locations. Some ASL verbs do not inflect or inflect for their object's location only (Liddell, 2003; Padden, 1988). There are other categories of ASL verbs (e.g., "depicting," "locative," or "classifier") whose movements convey complex spatial information and other forms of verb inflection (e.g., for temporal aspect); these are not the focus of this paper.

Fig. 1. Two inflected versions of the ASL verb EMAIL: (top) subject associated with location on left and object on right, (bottom) subject on right and object on left.

3 Related Work on Sign Animation

Given how the association of entities with locations in space affects how signs are performed, it is not possible to pre-store all possible combinations of all the signs the system may need. For pointing signs, inflecting verbs, and other space-affected signs, successful ASL systems must synthesize a specific instance of the sign as needed. Few sign language animation researchers have studied spatial inflection of verbs. There are two major types of ASL animation research: scripting software (Elliott et al., 2008; Traxler, 2000) or generation software (e.g., Fotinea et al., 2008; Huenerfauth, 2006; Marshall and Safar, 2005; VCom3D, 2012) as surveyed previously by (Huenerfauth and Hanson, 2009). Unfortunately, current generation and scripting systems for sign language animations typically do not make extensive use of spatial locations to represent entities under discussion, the output of these systems looks much like the animations without space use and without verb inflection that we evaluated in (Huenerfauth and Lu, 2012).

For instance, Sign Smith Studio (VCom3D, 2012), a commercially available scripting system for ASL, contains a single uninflected version of most ASL verbs in its dictionary. To produce an inflected form of a verb, a user must use an accompanying piece of software to precisely pose a character's hands to produce a verb sign; this significantly slows down the process of scripting an ASL animation. One British Sign Language animation generator (Marshall and Safar, 2005) can associate entities under discussion with a finite number of locations in space (approximately 6). Its repertoire also includes a few verbs whose subject/object are positioned at these locations. However, most of the verbs handled by their system involved relatively simple motion paths for the hands from subject to object locations, and the system did not allow for the arrangement of pronominal reference points at arbitrary locations in space.

Toro (Toro, 2004; 2005) focused on ASL inflected verbs; they analyzed the videos of human signers to note the 2D hand locations in the image for different verbs. Next, they wrote animation code for planning motion paths for the hands based on their observations. A limitation of this work is that asking humans to look for hand locations in a video and write down angles and coordinates is

inexact; further, a human looked for patterns in the data – machine learning approaches were not used.

There are some sign language animation researchers who have used modeling techniques applied to human motion data. Researchers studying coarticulation for French Sign Language (LSF) animations (Segouat & Braffort, 2009) digitally analyzed the movements of human signers in video and trained mathematical models of the movements between signs, which could be used during animation synthesis. Because collecting data from human via video requires researchers to estimate movements from a 2D image, it is more accurate and efficient to use motion-capture sensors. Duarte et al. collected data via motion capture in their SignCom project for LSF (Duarte and Gibet, 2011), and they reassembled elements of the recordings to synthesize novel animations.

4 Our Prior Modeling Research

The goal of our research is to construct computational models of ASL verbs that can automate the work of human users of scripting software or be used within generation. Given the name of the verb, the location in space associated with verb's subject, and the location associated with the object, our software should access its parameterized lexicon of ASL verb signs to synthesize the specific inflected form needed for a sentence. Our technique for building these parameterized lexicon entries for each verb is data-driven: based on samples of sign language motion from human signers. Specifically, we record a list of examples of each verb for a variety of arrangements of the verb's subject and object around the signer's body. Fig. 2. shows how we identified 7 locations on an arc around the signer; we then collected examples of each verb for all possible combinations of these seven locations for subject and object. Table 1 lists the ASL verbs modeled in our prior work (Huenerfauth and Lu, 2010; Lu and Huenerfauth, 2011).

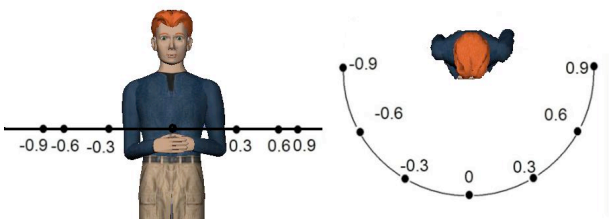

Fig. 2. Front & top view of arc positions around the signer.

Table 1: Five ASL Verbs We Have Modeled

Verb	*Inflection Type*	*Description*
ASK	Subject & Object	The signer moves an extended index finger from the "asker" (subject) to the "person being asked" (object). During the movement, the finger bends into a hooked shape. (ASL "1" to "X" handshape.)
GIVE	Subject & Object	In this two-handed version of the sign, the signer moves two hands as a pair from the "giver" (subject) toward the "recipient" (object). (Both hands have an ASL "flat-O" handshape.)
MEET	Subject & Object	Signer moves two index fingers towards each other (pointing upward) to "meet" at some point in the middle. (ASL "1" handshape.)
SCOLD	Object Only	The signer "wags" (bounces up and down while pointing) an extended index finger at the "person being scolded" (object). (ASL "1" handshape.)
TELL	Object Only	The signer moves an extended index finger from the mouth/chin toward the "person being told" (object). (ASL "1" handshape.)

For verbs inflected for both subject and object location (MEET, GIVE), our training data contained 42 examples for all non-reflexive combinations of the 7 arc positions. For verbs inflected for object location only (TELL, SCOLD, ASK), 7 examples were collected. While we focused on these five verbs as examples, we intend for our lexicon building methodology to be generalizable to other verbs and other sign languages. In our early work (Huenerfauth and Lu, 2010), we collected samples of inflected verbs by asking a native ASL signer with animation experience to produce these verbs using the Gesture Builder sign creation software (VCom3D, 2012). In later work, we collected more natural/accurate data by using motion-capture equipment to record a human signer performing a verb for various arrangements of subject/object in space (Lu and Huenerfauth, 2011).

Regardless of the data source, we extracted the hand position for each keyframe for each verb. (A keyframe is an important moment for a movement; a straight-line path can be represented merely by its beginning and end.) Thus, for a two-handed verb (e.g., GIVE) that is inflected for both subject and object, we collected 504 location values: 42 examples x 2 keyframes x 2 hands x 3 (x, y, z) values. Next, we fit third-order polynomial models for each dimension (x, y, z) of the hand position at each keyframe – parameterized on the arc locations of the verb's subject and object for that instance in the training data (Huenerfauth and Lu, 2010).

At this point, we could use the model to synthesize novel ASL verb sign instances (properly inflected for different locations of subject and object,

including combinations not present in the training data) by predicting the location of the hand for each of the keyframes of a verb, given the location of the verb's subject and object on the arc. Our animation software is keyframe based, and it uses inverse kinematics and motion interpolation to synthesize a full animation from a list of hand location targets for specific keyframe times during the animation. Additional details appear in (Huenerfauth and Lu, 2010; Lu and Huenerfauth, 2011).

To evaluate our models in prior work, we conducted a variety of user-based and distance-metric-based evaluations. For instance, we showed native ASL signer participants animations of short ASL stories that contained verbs (some versions produced by our model, and some produced by a human animator) to measure whether the stories containing our modeled verbs were easily understood, as measured on comprehension questions or side-by-side subjective evaluations (Huenerfauth and Lu, 2010). No significant differences in comprehension or evaluation scores were observed in these prior studies, indicating that the ASL animations synthesized from our model had similar quality to verb signs produced by a human animator.

5 Collecting More Verb Examples

In prior work, we used motion-capture data from only a single human signer performing many inflected forms of five ASL verbs. For this paper, we asked two additional signers to perform examples of each inflected form of the five verbs. This section summarizes the collection methodology, described in detail in (Lu and Huenerfauth, 2011). During a videotaped 90-minute recording session, each native ASL signer wore a set of motion-capture sensors while performing a set of ASL verb signs, for various given arrangements of the subject and object in the signing space. We use an Intersense IS-900 motion capture system with an overhead ultrasonic speaker array and hand, head, and torso mounted sensors with directional microphones and gyroscope to record location (x, y, z) and orientation (*roll*, *pitch*, *yaw*) data for the hands, torso, and head of the signer during the study. We placed colored targets around the perimeter of the laboratory at precise angles, relative to where the signer was seated, corresponding to the points on the arc in Fig. 2. Fig. 4. shows how we set up the laboratory during the data collection with 10cm colored paper squares were attached to the walls; the two squares visible in Fig. 4 correspond to arc positions 0.9 and 0.6 in Fig. 2. These squares served as "targets" for the signer to use as "subject" and "object" when performing various inflected verb forms.

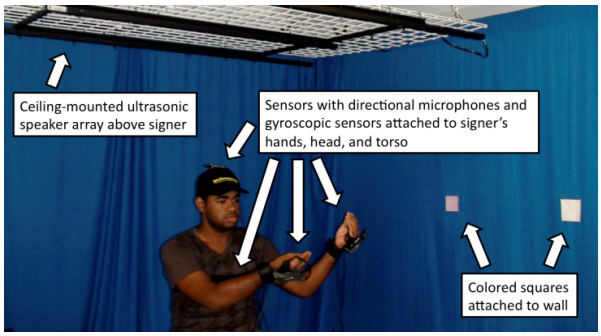

Fig. 4. This three-quarter view illustrates the layout of the laboratory during the motion capture data collection; the signer is facing a camera (off-screen to the right). Sitting behind the camera is another signer conversing with him.

Another native ASL signer sitting behind the video camera prompted the performer to produce each inflected verb form by pointing to the colored squares for the subject and the object for each of the 42 samples we wanted to record for each verb. At the beginning of the session, the signer was asked to make several large arm movements and hand claps (Fig. 5) to facilitate the later synchronization of the motion capture stream with the video data and scaling the data from the recorded human to match the body size of the VCom3D avatar.

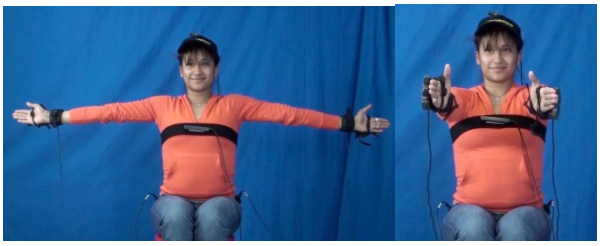

Fig. 5. Arm movements the signer was asked to perform to facilitate calibration of the collected motion capture data.

Fig. 6. The signer signed the number that corresponded to each verb example being performed (left) and a close-up view of the hand-mounted sensor used in the study (right).

Occasionally during the recording session (and whenever the signer made a mistake and needed to repeat a sign), the signer was asked to sign the sequence number of the verb example being recorded (Fig. 6); this facilitated later analysis of the video.

We needed to identify timecodes in the motion capture data stream that correspond to the beginning and ending keyframes of each verb recorded. We asked a native ASL signer to view the video after the recording session to identify the time index (video frame number) that corresponded to the start and end movement of each verb sign that we recorded. (If we had modeled signs with more complex motion paths, we might have needed more than two keyframes.) These time codes were used to extract hand location (x, y, z) data from the motion capture stream for each hand for each keyframe for each verb example that was recorded.

6 Modeling the Verb Path as a Vector

Although experimental evaluations of verb models produced in prior work based on motion-capture data from a single human signer were positive (Lu and Huenerfauth, 2011), this may not have been a realistic test. When constructing a large-scale sign language animation system, it may not be possible to gather all of the needed training examples for all of the verbs for a large lexicon from a single signer. For instance, if you wish to learn performances of a verb from examples of the inflected form of that verb that happen to appear in a corpus, then you would likely need to mix data recorded from multiple signers to produce your training data set for learning the inflected verb animation model.

The challenge of using data from multiple signers is that an ASL verb performance consists of: (1) non-meaningful/idiosyncratic variation in how different people perform a verb (or how one person performs a verb on different occasions) and (2) meaningful/essential aspects of how a verb should be performed (that should be rather invariant across different signers or different occasions). We prefer a model that captures the essential nature of the verb but not the signer-specific elements; models attuned too much to the specifics of a single human's performance may overfit to that one individual's version of the verb (or that one occasion when the signer performed). Further, while motion-capture data recorded from humans with different body proportions can be somewhat rescaled to fit the animated character's body size to be used by the sign language animation system, no "retargeting" algorithm is perfect. If signer-specific idiosyncrasies are captured in the verb animation model, then the variation in data sources used when building a large-scale sign language animation project may be apparent in its output.

Our prior modeling technique explicitly learned the starting and ending location of the hands for each instance of a verb based on a human signer's movements. However, when different signers perform a verb (e.g., GIVE with subject at arc position -0.6 and object at 0.3), they may not select exactly the same point in 3D space for their hands to start and stop. What is common across all of the variations in the performance is the *overall direction* that the hands move through the signing space. We can find empirical evidence for this intuition if we compare motion-capture data of the three different signers we recorded (section 5) performing the same ASL inflecting verbs. When we calculate Euclidean distance between different signer's starting location and their ending locations of the hands for identical verb examples, we see inter-signer variability (Fig. 7). If we instead calculate the Euclidean distance between the vector (direction and magnitude) of the hand movement from the start to the ending location between signers, we see much smaller inter-signer variability (Fig. 7). Section 7 explains the scale and formula used for the distance metrics in Fig. 7 and elsewhere in this paper.

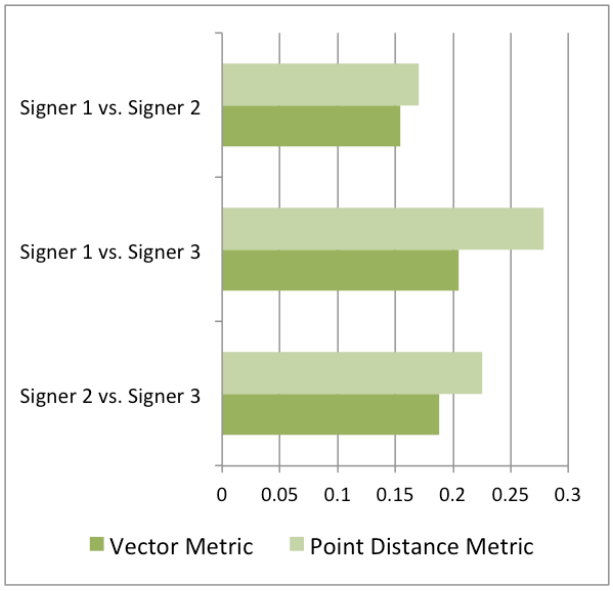

Fig. 7. Inter-signer variability in ASL verb signs, reported using a "point" or "vector" distance metric.

Using these results as intuition, we present a new model of ASL inflecting verbs in this paper, based on this "vector" approach to modeling the movement of the signer's hands through space. We assume that what is essential to a human's performance of an inflected ASL verb is the direction that the hands travel through space, not the specific starting and ending locations in space. Thus, we model each verb example as a tuple of values: the difference between the x-, the y-, and the z-axis values for the starting and ending location of the hand. (The model has three parameters for a one-handed sign and six parameters for a two-handed sign.) Using this model, we followed a similar polynomial fitting technique summarized in section 4 – except that we are now modeling a smaller number of parameters – our new "vector" model uses only three values per hand ($delta_x$, $delta_y$, $delta_z$), instead of six per hand in our prior "point" model, which represented start and end location of the hand as (x_{start}, y_{start}, z_{start}, x_{end}, y_{end}, z_{end}).

This new model can then be used to synthesize animations of ASL verb signs for given subject and object arc positions around the signer – the difference from our prior work is that these new models only represent the movement vector for the hands, not their specific starting and ending locations.

The purpose of building a model of a verb is that we wish to use it as a parameterized lexical entry in a sign language animation synthesis system; thus, we must explain how the model can be used to synthesize a novel verb example, given its input parameters (the arc position of the subject and the object of the verb). While our new vector model predicts the motion vector for the hands, this is not enough; we need starting and ending locations for the hands (an infinite number of which are possible for a given vector). Thus, we need a way to select a starting location for the hands for a specific verb instance (and then based on the vector, we would know the ending location).

We observe that, for a given verb, there are some locations in the signing space that are likely for the signer's hands to occupy and some regions that are less likely. Some motion paths through the signing space travel through high-likelihood "popular" regions of the signing space, and some, through less likely regions. Thus, we can build a Gaussian mixture model of the likelihood that a hand might occupy a specific location in the signing space during a particular ASL verb. For a given motion vector, one possible starting point in the signing space will lead to a path that travels through a maximally likely region of the signing space. Thus, we can search possible starting points for the hands for a given vector and identify an optimal path for the hands given a Gaussian mixture model of hand location likelihood.

Fig. 8 shows a (two-dimensional) illustration of our approach for selecting a starting location for the hand when synthesizing a verb. The concentric groups of ovals in the image represent the component Gaussians in the mixture model, which was fit on the data from the locations that one hand occupied during a signer's performances of a verb. Given the vector (direction and magnitude) for the hand's motion path for a verb (predicted by our model), we can systematically search the signing space for all possible starting locations for the hand – to identify the starting location that yields a path through the signing space with maximum probability (as predicted by the Gaussian model). The arrows shown in Fig. 8 represent a few possible paths for the hand given several possible starting locations, and one of these arrows travels a path through the model with maximum probability.

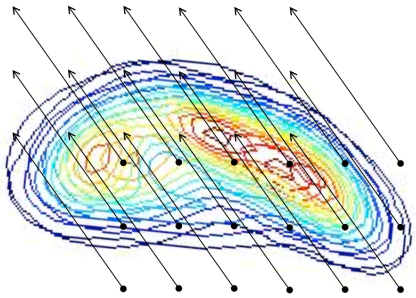

Fig. 8. This 2D diagram illustrates how the starting location for the hand can be selected that yields a path through the mixture model with maximum probability.

Specifically, for each signer, for each hand, for each verb, we used the recorded motion-capture data stream between the start-times and end-times of all of the verb examples as training data, and then we fit a 3D Gaussian mixture model for each, to represent the probability that the hand would occupy each location in the signing space during that verb. We used a model with 6 component Gaussians for modeling the signing space for each of the verbs SCOLD, GIVE, ASK, and MEET. Due to the fast movement (and thus short clips of recorded motion-capture data) for the verb TELL, we only had sufficient data to fit a 5-component

Gaussian model for the locations of the hand during this verb (TELL is a one-handed verb). When we need to synthesize a verb, then we use our vector model to predict a movement vector for the hands, and then we perform a grid search through the signing space (in the x, y, and z dimensions) to identify an optimal starting location for the hand. If run-time efficiency is a concern, optimization or estimation methods could be applied to this search.

In summary, the vector direction and magnitude of the hands are based on a model that is parameterized on: the verb, the location of the subject on an arc around the signer, and the location of the object on this arc. When a specific instance of a verb must be synthesized, a starting point for the hand is selected that maximizes the probability of the entire trajectory of the hands through space, based on a Gaussian mixture model specific to that verb (but not parameterized on any specific subject/object locations in space). All instances of the verb in the training data were used to train the mixture model, due to data sparseness considerations.

7 Distance Metric Evaluation

Because the premise of this paper is that models of ASL verbs based on a motion vector representation would do a better job of capturing the essential aspects of a verb's motion path across signers, we conducted an inter-signer cross-validation of our new model. We built separate models on the data from each of our three signers, and then we compared the resulting model's predictions for all 42 verb instances collected from the other two signers. For comparison purposes, we also trained three models (one per signer) using the "point"-based model from our prior work (Lu and Huenerfauth, 2011). Fig. 9 presents the results; the values of each bar are the average "error" for each synthesized verb example for all five ASL verbs in Table 1. The error score for a verb example is the average of four values: (1) Euclidean distance between the start location of the right hand as predicted by the model and the start location of the right hand of the human signer data being used for evaluation, (2) same for the end location for the right hand, (3) same for the start location for the left hand, and (4) same for end location for the left hand.

Fig. 9 shows that the new "vector" model has lower error scores than our older "point" model presented in prior work. To interpret the Euclidean distance value, it is useful to know that the scale of the coordinate space used for the verb model is set such that shoulder width of a signer would be 1.0. As a baseline for comparison, the average inter-signer variation (based on the values shown in Fig. 7) is also plotted in Fig. 9.

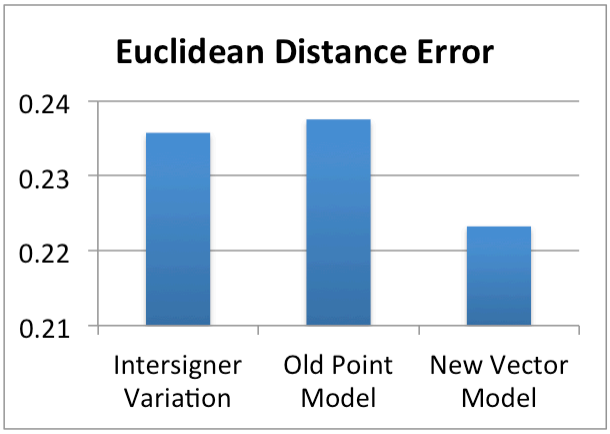

Fig. 9. Evaluation of the "Point" and "Vector" models for all five ASL verbs listed in Table 1.

Next, we wanted to compare the two models under two assumptions: (1) it may not be possible to gather a large number of examples of a verb from a single signer and (2) it may be necessary to mix data from multiple signers when assembling a training data set for a verb model. For instance, these conditions would hold if a researcher were using examples of a verb performance extracted from a multi-signer corpus to assemble a training set. Due to the limited size of most sign language corpora (and the many possible combinations of subject and object position in the signing space), a training set gathered in this manner would likely contain a relatively small number of training examples – possibly gathered from multiple signers.

To test the models under these conditions, we assembled three training data sets – using the data from our three recorded signers. Each data set included 22 examples of the performance of an ASL inflected verb for a subset of the various possible combinations of subject and object locations in the signing space – with half of the examples from one signer and half from another. After training a model on each data set, then the model was evaluated against the 42 examples of each verb performance recorded from the third signer (who was not part of the training data used for that model). This process was repeated for a total of three times (for all combinations of the data from the three sign-

ers). For comparison purposes, we also trained three models (one based on each of the three two-signer data sets) using the "point"-based model from our prior work (Lu and Huenerfauth, 2011).

Fig. 10 shows the results for two of the verbs in Table 1 (ASK and GIVE); the "vector" model has lower error scores than our older "point" model.

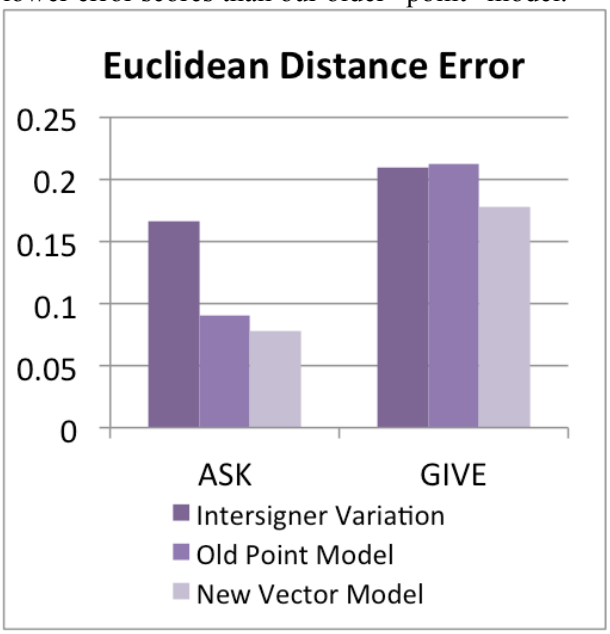

Fig. 10. Evaluation of the "Point" and "Vector" models trained on a small "mixed" data set from two signers.

Examples of animations of the ASL verbs synthesized using each of these models are on our lab website: http://latlab.cs.qc.cuny.edu/slpat2012/

8 Conclusion And Future Work

This paper presented and evaluated a new method of constructing a lexicon of ASL verb signs whose motion path depends on the location in the signing space associated with the verb's subject and object. We used motion capture data from multiple signers to evaluate whether our new models do a better job of capturing the signer-invariant and occasion-invariant aspect of an ASL inflected verb's movement, compared to our prior modeling approach. The parameterized models of ASL verb movements produced in this paper could be used to synthesize a desired verb instance for a potentially infinite number of arrangements of the subject and object of the verb in the signing space – based on the collection of a finite number of examples of a verb performance from a human signer.

Using this technique, generation software could include flexible lexicons that can be used to synthesize an infinite variety of inflecting verb instances, and scripting software could more easily enable users to include inflecting verbs in a sentence (without requiring the user to create a custom animations of a body movement for a particular inflected verb sign). While this paper demonstrates our method on five ASL verbs, this technique should be applicable to more ASL verbs, more ASL signs parameterized on spatial locations, and signs in other sign languages used internationally.

In this paper, we studied a set of ASL verbs with relatively simple motion-paths (consisting of straight line movements, which therefore only required two keyframes per verb); in future work, we may analyze verbs with more complex movements of the hands. Further, our vector models represent the magnitude (length) of the hands' motion path through space; in future work, we may explore techniques for rescaling these vector lengths. In future work, we will also use hand orientation data from our motion capture sessions to synthesize hand orientation for sign animations. We also plan to experiment with modeling how the timing of keyframes varies with subject/object positions.

Finally, we also plan on conducting a user-based evaluation study using animations synthesized by the models presented in this paper – to determine if native ASL signers who view animations containing such verbs find them to be more grammatical, understandable, and natural.

Acknowledgments

This material is based upon work supported in part by the US. National Science Foundation under award number 0746556 and award number 1065009, by The City University of New York PSC-CUNY Research Award Program, by Siemens A&D UGS PLM Software through a Go PLM Academic Grant, and by Visage Technologies AB through a free academic license for character animation software. Jonathan Lamberton assisted with the recruitment of participants and the conduct of experimental sessions. Kenya Bryant, Wesley Clarke, Kelsey Gallagher, Amanda Krieger, Giovanni Moriarty, Aaron Pagan, Jaime Penzellna, Raymond Ramirez, and Meredith Turtletaub have also assisted with data collection and contributed their ASL expertise to the project.

References

Cormier, K. 2002. Grammaticalization of Indexic Signs: How American Sign Language Expresses Numerosity. Ph.D. Dissertation, University of Texas at Austin.

Cox, S., M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, S. Abbott. 2002. Tessa, a system to aid communication with deaf people. In Proceedings of Assets '02, 205-212.

Duarte, K., and Gibet, S. Presentation of the SignCom Project. In Proceedings of the First International Workshop on Sign Language Translation and Avatar Technology, Berlin, Germany, 10-11 Jan 2011.

Elliott, R., Glauert, J., Kennaway, J., Marshall, I., Safar, E. 2008. Linguistic modeling and language-processing technologies for avatar-based sign language presentation. Univ Access Inf Soc 6(4), 375-391. Berlin: Springer.

Fotinea, S.E., Efthimiou, E., Caridakis, G., Karpouzis K. 2008. A knowledge-based sign synthesis architecture. Univ Access Inf Soc 6(4):405-418. Berlin: Springer.

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

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.

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.

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

Huenerfauth, M., Lu, P. 2010. Modeling and synthesizing spatially inflected verbs for American sign language animations. In Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '10). ACM, New York, NY, USA, 99-106.

Huenerfauth, M, P. Lu. (2012. in press). Effect of spatial reference and verb inflection on the usability of American sign language animation. In Univ Access Inf Soc. Berlin: Springer.

Klima, E., U. Bellugi. 1979. The Signs of Language. Harvard University Press, Cambridge, MA.

Liddell, S. 2003. Grammar, Gesture, and Meaning in American Sign Language. UK: Cambridge U. Press.

Lillo-Martin, D. 1991. Universal Grammar and American Sign Language: Setting the Null Argument Parameters. Kluwer Academic Publishers, Dordrecht.

Lu, P., Huenerfauth, M. 2011. Synthesizing American Sign Language Spatially Inflected Verbs from Motion-Capture Data. Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), in conjunction with ASSETS 2011, Dundee, Scotland.

Marshall, I., E. Safar. 2005. Grammar development for sign language avatar-based synthesis. In Proc. UAHCI'05.

McBurney, S.L. 2002. Pronominal reference in signed and spoken language. In R.P. Meier, K. Cormier, D. Quinto-Pozos (eds.) Modality and Structure in Signed and Spoken Languages. UK: Cambridge U. Press, 329-369.

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

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.

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

Padden, C. 1988. Interaction of morphology & syntax in American Sign Language. New York: Garland Press.

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

Toro, J. 2004. Automated 3D animation system to inflect agreement verbs. Proc. 6th High Desert Linguistics Conf.

Toro, J. 2005. Automatic verb agreement in computer synthesized depictions of American Sign Language. Ph.D. dissertation, Depaul University, Chicago, IL.

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

VCom3D. 2012. Homepage. http://www.vcom3d.com/

Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., Palmer, M. 2000. A machine translation system from English to American Sign Language. In Proc. AMTA'00, pp. 293-300.