Cyclic Data-Driven Research on American Sign Language Animation

Matt Huenerfauth

The City University of New York (CUNY)
Computer Science Department, CUNY Queens College
Computer Science and Linguistics Ph.D. Programs, CUNY Graduate Center

Invited Keynote Presentation,
International Workshop on Sign Language Translation and Avatar Technology (SLTAT), Berlin, Germany, January 2011

Why Sign Language Animation?

This is how my lab justifies our work whenever we present it:

- American Sign Language (ASL) animation generation software can make information and services accessible for deaf individuals in the U.S. with low English literacy.
  - ASL is not just signs performed in English word order.
  - ASL: primary means of communication for 500,000 people. (Mitchell et al., 2006)
  - Only half of deaf high school graduates (age 18+) in the United States can read English text at a fourth-grade (age 10) level. (Traxler, 2000)
- ASL animation technology ➔ new accessibility applications:
  - English to ASL machine translation software
  - New forms of TV and other captioning
  - Computer user-interfaces incorporating ASL
    - Educational software, web browsers that present information accessibly

Two Ways to Make ASL Animations

- ***Scripted*** by a developer who knows ASL: placing individual signs and facial expressions on a timeline to make sentences.
  - e.g., Elliott, Glauert, Kennaway, Marshall, Safar, 2008; Kennaway, Glauert, Zwitserlood, 2007; commercial product SignSmith Studio (VCom3D, Inc.)
- ***Generated automatically*** (e.g., machine translation software).
  - e.g., Chiu, Wu, Su, Cheng, 2007; Huenerfauth, 2006; Marshall & Sáfár, 2005; Stein, Bungeroth, Ney, 2006; Zhao, Kipper, Schuler, Vogler, Badler, & Palmer, 2000; Veale, Conway, & Collins, 1998; van Zijl & Barker, 2003; Tokuda & Okumara, 1998
- For both, an animation must be *synthesized* with many detailed movement and timing parameters set correctly.

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  – Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  – **Speed and Timing of ASL Animations**
  – **Use of Spatial Reference Points (SRPs)**
  – **Verb Inflection for Spatial Reference Points**
  – **Coarticulation Effects in ASL Signing**
  – **Non-Manual Signals During ASL**

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  – **Speed and Timing of ASL Animations**
    - The speed of individual sign performances
    - The transitional time between signs
    - The insertion of pauses during signing
    - All three 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 timing are imperfect are significantly less understandable to ASL signers (Huenerfauth, 2009).
  – **Use of Spatial Reference Points (SRPs)**
  – **Verb Inflection for Spatial Reference Points**
  – **Coarticulation Effects in ASL Signing**
  – **Non-Manual Signals During ASL**

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  – **Speed and Timing of ASL Animations**
  – **Use of Spatial Reference Points (SRPs)**
    - Entities under discussion associated with 3D points in space around a signer
    - Some Interesting Research Questions:
      – Where in 3D space do signers tend to place the first entity? The second? Third? What factors affect these choices?
      – Can we predict which entities under discussion must be placed in space? What linguistic factors of the entity or of the surrounding discourse affect this decision?
  – **Verb Inflection for Spatial Reference Points**
  – **Coarticulation Effects in ASL Signing**
  – **Non-Manual Signals During ASL**

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  - **Speed and Timing of ASL Animations**
  - **Use of Spatial Reference Points (SRPs)**
  - **Verb Inflection for Spatial Reference Points**
    - Many verbs change their motion paths to indicate the 3D location of their subject, object, or both (Padden, 1988). Complex combination of the verb's lexical motion path and the locations of the arguments of the verb.
    - When an ASL 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.
  - **Coarticulation Effects in ASL Signing**
  - **Non-Manual Signals During ASL**

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  - **Speed and Timing of ASL Animations**
  - **Use of Spatial Reference Points (SRPs)**
  - **Verb Inflection for Spatial Reference Points**
  - **Coarticulation Effects in ASL Signing**
    - Complex factors that affect how performances of signs are affected by the preceding and subsequent signs – or by aspects of performance in parallel.
    - ASL generators that use overly simple interpolation rules to produce these coarticulation effects yield unnatural and non-fluent ASL animation output.
  - **Non-Manual Signals During ASL**

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  - **Speed and Timing of ASL Animations**
  - **Use of Spatial Reference Points (SRPs)**
  - **Verb Inflection for Spatial Reference Points**
  - **Coarticulation Effects in ASL Signing**
  - **Non-Manual Signals During ASL**
    - Subtle aspects of facial expression, head-tilt, eye-gaze, and body movement.
    - Indicate important information in ASL about subject/object of verbs, negation, questions, topicalization, etc.
    - Some Interesting Research Questions
      - How do these combine together? Coarticulation for these non-manual signals? Intensity of facial expressions over time?

What makes ASL hard to synthesize?

- Many factors! But let's focus on a few:
  - **Speed and Timing of ASL Animations**
  - **Use of Spatial Reference Points (SRPs)**
  - **Verb Inflection for Spatial Reference Points**
  - **Coarticulation Effects in ASL Signing**
  - **Non-Manual Signals During ASL**
- All of these involve non-discrete values of many complex parameters to be set correctly to produce a good animation.
  - These issues are more complex than selecting the appropriate sequence of glosses/signs to concatenate from a finite lexicon.
  - Even native ASL signers who are hand-producing ASL animations may not have overt intuitions on how to set all of these correctly.
  - A data-driven approach based on collected samples of ASL from humans can benefit this research.

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Various Sources of Data from Signers

- An ASL corpus with motion-capture data and full linguistic annotation would be great (and we're building one – details later), but we can also get useful data from signers by:
  - Published results in the linguistics literature.
  - Grammaticality/understandability judgments.
  - Asking signers to script sentences using ASL scripting tools or asking signers to produce inflected forms of ASL signs using animation software.
  - Recording and annotating video-based ASL corpora.
  - Recording signers using motion-capture equipment while they perform sentences with very specific phenomena of interest + annotating only those.

Some Examples of Data-Driven SL Research

- Motion-capture for lexicon creation (Cox et al., 2002)
- Statistical machine translation based on gloss transcripts of sign language (Morrissey & Way, 2005; Stein et al., 2006)
- Deriving models of coarticulation based on rotoscoping-derived data from LSF video (Seguoat and Braffort, 2009)
- Video-based sign language corpora projects (Neidle et al., 2000; Bungeroth et al., 2006; Crassborn et al., 2004, 2006; Efthimiou & Fotinea, 2007)

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Our Lab's Research Paradigm

- Design-cycle research paradigm that consists of:
  - Directed data-collection of a specific linguistic phenomenon of interest (using some of those data sources listed previously)
  - Creation of models for this phenomenon
  - Synthesis of prototype animations based on the new model (and older models for comparison purposes)
  - Conduct of an experimental study with native ASL signers evaluating animations via comprehension questions and subjective forms of evaluation
  - Iterative refining of the model and re-evaluation
- How does this compare to your labs' approaches?

How do we evaluate animations?

- We conduct studies with native ASL signers who view our animations and answer:
  - Comprehension questions about the information presented in the sign language animation
    - Accessibility of information content that was too difficult to read in written form was our original motivation for ASL animations: measure this!
  - Subjective evaluations of the perceived quality and correctness of the animations
    - This may affect users' willingness to watch these animations over time.

Methodological Details

- We've published details of how we conduct user-based studies with comprehension-question based evaluations.
  - Recruitment and screening for native signers
  - Controlling the experimental environment
  - Making materials accessible and ASL-focused
  - Designing stories/passages as stimuli and embedding the synthesized items we are comparing in the animations.
  - Using ASL to ask questions and using clip-art images or photos for our answer choices.
  - There are often differences between user-reported ease-of-understanding for a passage and their actual scores on comprehension-questions! Participants often "confidently wrong."
  - Sometimes surprising results on what does/doesn't lead to comprehension question differences.

What issues have we addressed using this research paradigm?

- Measuring the importance of spatial reference points on the understandability of ASL animations by native signers. (UAHCI-2009 conference, UAIS journal 2010)
- Creating models of speed, timing, and pauses for ASL animations based on linguistic models of human performance and measuring the effect on comprehension. (TACCESS journal 2009, ASSETS-2008 conference)
- Building models of how ASL verbs change their motion-paths based on the arrangement of their subject/object in 3D space. (ASSETS-2010 conference)

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

John blames Mary.

JOHN POINT$_x$ $_x$BLAME$_y$ MARY POINT$_y$

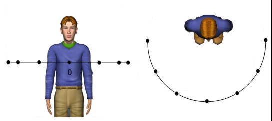
The signer in the videotape finger-spells John and set up "John" by pointing to a location in space on the left of his body, then he makes the sign "blame" with the motion path moving from the point "John" towards the point in space where "Mary" is going to set up. Then he makes the sign of "Mary" and set up a location in space for "Mary".

Two versions of this sentence.

Placement of John and Mary is swapped in the 2nd version.

Where do you put everything?

- ASL linguists debate where in 3D space signers set up these reference points:
  - Semi-circular arc floating at chest height in front of their torso?
  - Arbitrary 3D location in the surrounding signing space (at different heights and distances from the signer)?
  - Regardless, there is an infinite number of locations possible on the arc or in 3D space.



Current Systems: Verb Inflection

- ASL animations without correctly infected verbs have limited understandability for deaf users (Huenerfauth, 2009).
- Generation/Translation Systems:
  - Usually do not to include inflected forms of verbs in their output.
  - Marshall & Safar's (2005) British Sign Language generation system included multiple copies of several verbs in its dictionary for a finite number of combinations of subject/object arrangement in space.
  - Toro (2004) also investigated inflection of some ASL verb signs.
- Scripting Systems:
  - E.g. Vcom3D includes only 1 uninflected version of each verb.
  - If a user of the scripting system *really* wants to add a verb with spatial inflection (or any other sign that isn't in the system's dictionary), then the user needs to specify all the movements for that sign manually – and then import it into the sentence.

Our Goal

- Given a 3D location of where in the signing space the subject and object of a verb is placed,
  - We want software that can produce an *instance* of that verb that has been properly inflected
  - i.e., its motion-path has been modified to reflect the subject and object locations in 3D space
- We can't collect infinitely many signs (for all subject/object locations); so, our model must be able to synthesize previously-unseen instances.
- Initially, we focused on 5 example ASL verbs.

Five Verbs We Examined

Verb	*Inflection*	*#Hands*	*Description*
ASK1h	Subj+Obj	1	'ask a question': a bending index finger moves from Subj ('asker') to Obj ('askee')
GIVE2h	Subj+Obj	2	'give to someone': hands move as a pair from the Subj ('giver') to Obj ('recipient')
MEET	Subj+Obj	2	'two people meet': hands move from Subj and Obj toward each other symmetrically
SCOLD	Obj only	1	'scold/reprimand': extended index finger wags at the Obj ('person being scolded')
TELL	Obj only	1	'tell someone': index finger moves from signer's mouth to Obj ('person being told')

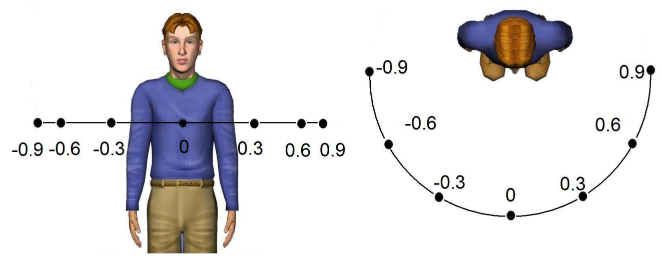
Other Assumptions

- We'll use inverse kinematics algorithms.
  - So we just need to figure out where the hand goes.
- We'll use motion-interpolation through keyframes to produce our animations.
  - So we just need to figure out where the hands should be at specific moments on the timeline.
- The locations in 3D space around a signer where subject and object are placed fall on an "arc."
- A verb's handshape is not affected by different subject and object positions on the "arc."
  - This is true most of the time.

Overall Methodology

1. Collect samples of instances of ASL inflecting verbs (for a variety of subject and object positions) from a human animator
2. Fit low-order polynomial models to the data
   - So that we can predict hand location and orientation for an ASL verb for any given subject and object positions around the signer
   - We could synthesize any instance of a verb that we want for any subject/object layout
3. Evaluate our models in several ways

Arc Positions Around the Signer



When we collect our data to train our animation models, we'll just collect examples of the subject and object at these seven positions on the arc (labeled as "-0.9," "-0.6," etc.)
However, the models we will create can produce instances of ASL verbs for subject/object positions between these seven values.

Collecting Verb Instances from Humans

- A native ASL signer used VCom3D Gesture Builder to produce examples of all possible subject/object combinations for all the verbs.
  - Experienced user of this software.
  - GUI lets you drag, reorient, and set handshape for the hands for keyframes on a timeline to produce an ASL sign.

What data did we gather?

For all five of the verbs we were studying:

For all possible combinations of subject and object positions on arc: {0.9, 0.6, 0.3, 0, -0.3, -0.6, -0.9}

- We noted the location of the subject of this instance of the verb, represented as a real-number specifying a position on the "arc"
- We noted the position of the object on "arc"
- For each keyframe of this instance of the verb:
  - Location of the hand: (x, y, z) coordinates
  - Orientation of the hand represented as a quaternion (a quadruple of numbers, details in the paper)

Building Models of Verb Inflection

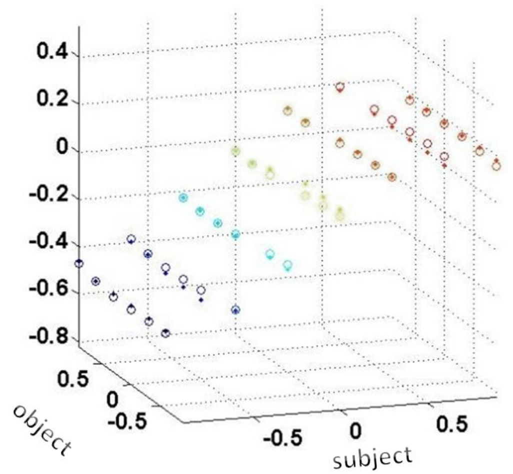
- For each keyframe of a verb instance, there are 7 values to be fit:
  - 3 parameters for location (x, y, z)
  - 4 parameters for orientation quaternion (q_0, q_1, q_2, q_3)
- For verbs whose movement is only affected by their object's position on the arc ("o"):

  parameter = f(o)
- For verbs whose movement is affected by both their subject's ("s") and object's position ("o"):

  parameter = f(s, o)

Modeling Each Parameter

- What's the function "f" on previous slide?
- We fit a 3rd order polynomial (least squares)
  …for each parameter
    …for each keyframe
      …for each ASL verb.
  - Details:
    - For verbs parameterized on object position only, model contained all terms o^b such that $0 \le b \le 3$
    - For verbs parameterized on subject and object position, the model contained all possible cross product terms $s^a o^b$ such that $a+b \le 3$, $a \ge 0$ $b \ge 0$.
    - Wrote MATLAB code to calculate coefficients.



Example of our 3rd order polynomial model fit to the data (least squares).

This is the x coordinate for the right hand for keyframe #1 of GIVE

Dots:
Data from human-produced samples.

Open Circles:
Model prediction.

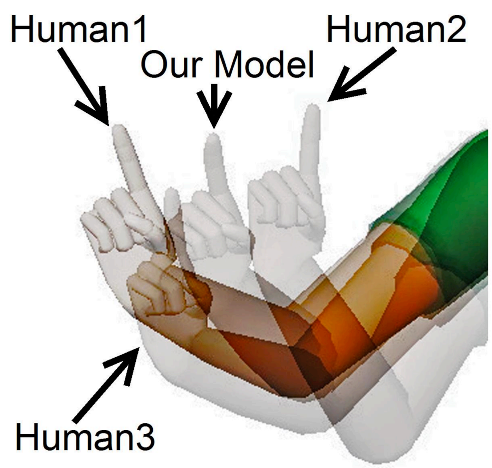
Sample Verbs: GIVE

Produced by Human Animator

Produced by Our Model

Comparing Our Models to Human Data

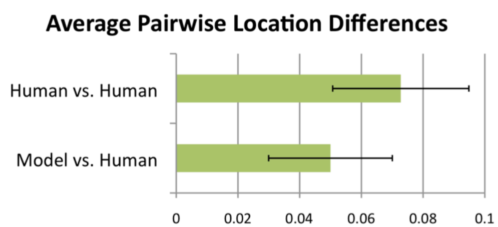
- We conducted a detailed analysis of one instance of all five of the verbs
  - Subject at -0.6 and object at 0.3.
- We asked our ASL-native human animator to return to the lab on three different days.
  - Each day, he produced an instance of each of the five verbs with subject at -0.6 and object at 0.3.
  - His animation from different days varied slightly.
  - We compared the instance of verbs produced by our model to the three collected human samples.

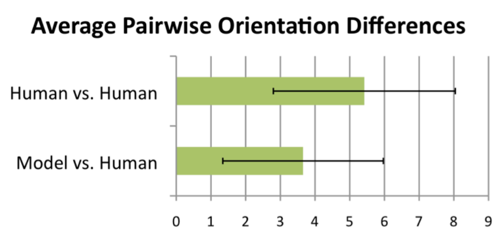


Close-up view of differences between keyframe #1 of ASK1h produced by our model or by the human animator.

Calculated Differences

- We calculated the average of the differences between our model and each of the three human-produced versions.
  - Euclidean distance for locations.
  - Distance formula for quaternions (see paper).
- We also calculated the average of the pairwise differences between the three human-produced versions to estimate the variance in human-produced signs.



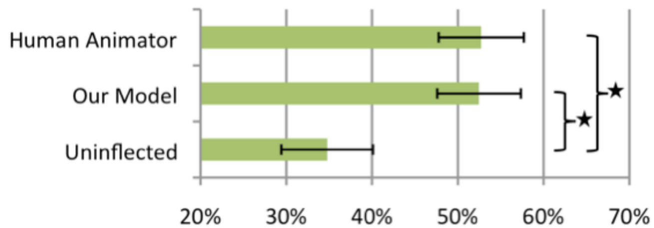


No significant differences between the bars ($p>0.05$, t-test)

User-Based Comprehension Study

- 18 native ASL signers, ASL-focused environment.
- Participants viewed 9 animations of a virtual human character telling a short story (in ASL).
  - Included instances of the five verbs we are studying.
  - Typically, 5-6 verbs per story. Average length: 55 signs.
  - Different characters in the story set up at different “arc” positions; served as subjects/objects for the verbs.
  - Each story produced in three forms, depending on whether its verbs were: (1) from our model, (2) from a human animator, or (3) uninflected form of the verb.
  - **Comprehension questions:** information conveyed by the verbs, identity of the subject and object of the verb.

Comprehension Scores

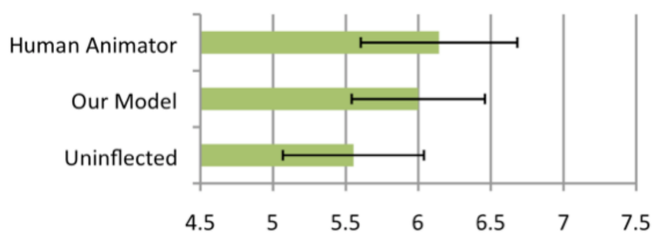


Error bars: standard error of the mean for each value
Significant pairwise differences: marked with a star
ANOVA, alpha = 0.05

Side-by-Side Comparisons

- Three identical sentences side-by-side, identical subject and object positions, same verb, but produced in different ways:
  - Produced by our model
  - Created by a human animator
  - Uninflected version of the verb from dictionary
- Instructed to focus on the verb and consider its grammaticality, understandability, and naturalness.
- Assigned a 1-to-10 Likert-scale score to each of the three versions of the animation.

Side-by-Side Comparison



Error bars: standard error of the mean for each value
No significant pairwise differences
Pairwise Mann-Whitney U-tests with Bonferroni-correction, alpha = 0.05

Summary of Our Work on Verb Inflection

- A novel approach to synthesizing animations of ASL signs affected by signing space.
  - Modeling hand location/orientation for ASL inflecting verbs to enable the synthesis of *infinitely many* versions of a verb – based on the values of input parameters that specify the position of the subject and object of a verb.
  - Creators of sign-language generation software can enable an infinite variety of inflecting verb instances to be included in the repertoire of their software.
  - Creators of sign-language scripting software can build models of signs for their dictionaries – allow users to easily insert a specific instance of an inflecting verb.

What's Next?

- More ASL Verbs, with…
  - Very complex motion paths or interaction of the two hands
  - That become uncomfortable when you try to perform them for certain subject and object positions
  - We want to find verbs that may push our very simple polynomial model to its limits to inspire better models.
  - We will repeat our design cycle!
- Relax some of our assumptions in this paper
  - Subject and object not limited to an arc
  - Allow for flexible sign durations/timings that are also parameterized on the subject and object positions
- Train models based on motion-capture data from people
  - Special challenges to using such data (identifying keyframes, cleaning up "noise," retargeting to a virtual human character, etc.)

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Using Motion-Capture Data

- While we have been able to model some simple verbs in this manner using the data gathered from signers using the animation software, perhaps the way that signers *think* they move is different than *how they really move*.
- Instead, we can record signers performing specific inflected ASL verb forms using motion capture equipment in our laboratory.
- We can set up "targets" around the room for different subject and object locations.
- We can record the verb forms performed.

Our Equipment: The Gloves

- Immersion 22-sensor CyberGloves.
  - Sensor strips record how much they bend.
  - Calculates hand pose.
  - Needs calibration before use.

Our Equipment: The Body Suit

- Animazoo IGS-190 Magnetic/ Inertial Body Suit
  - Spandex suit covered in soft Velcro
  - Small sensors attach to the suit
    - Record magnetic north and gravity
  - Track displacement from their starting position (standing T-pose)
  - You provide the suit with the bone measurements for the person wearing it. It calculates skeleton.

Our Equipment: Eye-Tracker

- Applied Science Labs "H6" head-mounted eye-tracker.
  - Small camera tracks eye movement.
  - Pupil-position vs. reflection of a low-power laser off the person's cornea.

Our Equipment: Head Tracker

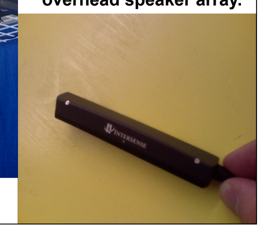
- We want to know where in a room the person is looking (compensating for head movement). So, we need to also use a head tracker that can tell us the 3D position (X, Y, Z) and orientation (roll, pitch, yaw) for the top of the head.
- This information is combined with the eye tracker data to calculate a 3D vector for the eye-gaze direction in a room.

Our Equipment: Head Tracker

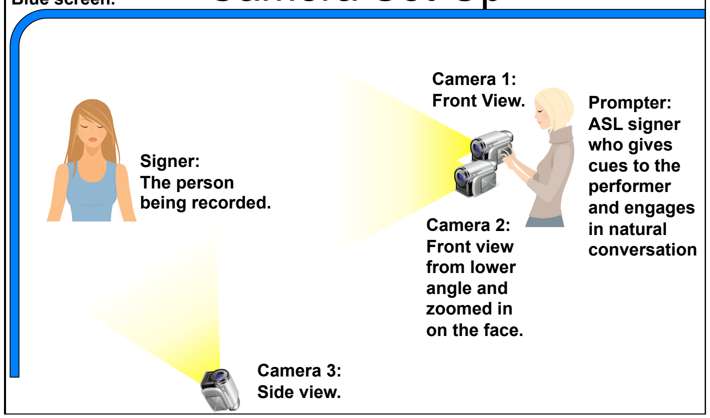
Intersense IS-900

Overhead ultra-sonic speaker array.

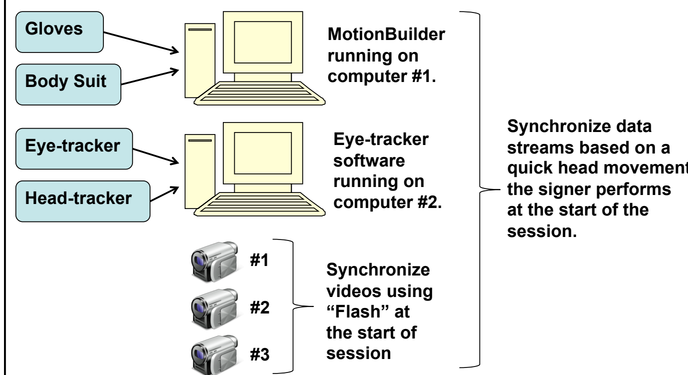
Inertial sensor with directional microphones for triangulating exact 3D location under the overhead speaker array.



Camera Set-Up



Synchronizing the Data Streams



Future Work

- We've already collected some verb forms using this motion-capture approach
  - With little targets set up around the room for different combinations of subject/object location
- We're currently cleaning up this data and extracting frames of data to be used for training our verb inflection models.
- In progress.

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Our ASL Motion-Capture Corpus Collection Project

With all this equipment, we wanted to build a resource for our research in the long-term…

An ASL Motion-Capture Corpus

- Mocap data in a full ASL corpus could drive more data-driven ASL animation synthesis research.
  - This rich dataset could be used for a variety of future purposes by animation and linguistic researchers.
  - Linguistic and Assistive Technologies Lab (LATLab) has begun a multi-year project to build the first motion-capture corpus of multi-sentential ASL utterances.
  - We record native ASL signers performing spontaneous and directed ASL passages while wearing motion-capture body suits, gloves, eye-trackers, etc.
  - This data is being linguistically annotated by native ASL signers to produce a permanent research resource.

Our Corpus

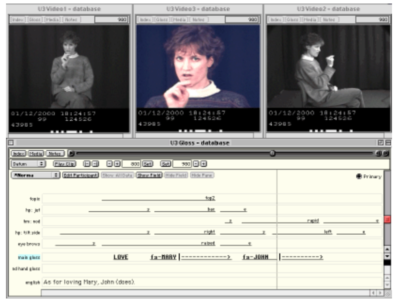
- Our main interest: spatial reference point (SRP) usage and inflected ASL verbs.
- Spontaneous single-signer multi-sentence ASL performances on non-narrative topics.
  - We want multi-sentence performance in which the signer is likely to set up different numbers of spatial reference points in the signing space.
  - We want to avoid some genres of signing that contain a large percentage of classifier predicates – different use of space than SRPs

Eliciting the Corpus

- To elicit the kinds of sentences that we want, we're pilot testing various forms of prompting.
  - Lists of topics or concepts that the signer is asked to discuss, asking for a short bio, asking the signer to relate the plot of a short story, asking the signer to describe the events of a short video.
- In all of our recording scenarios, a native ASL signer is behind the camera – giving the performing someone to focus on and converse with in a more natural manner.

SignStream

- Software developed by researchers at Boston University for linguistically annotating ASL performances.
  - Multiple views of signer. Timeline to add linguistic data.



(C. Neidle, 2002)

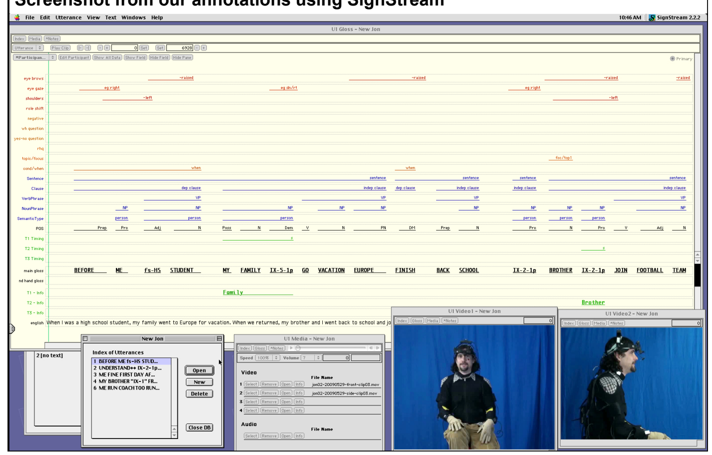
What Information Are We Adding?

- English translation for the performance.
- Glosses for each sign in the performance with part-of-speech of each.
- Grammatically meaningful non-manual signals (e.g. facial expressions used to indicate a question or a topicalized constituent at the start of a sentence).
- Marking syntactic constituents: NP, VP, S.
- Marking when a spatial reference point (SRP) is established in space, when it is referred to again, what its identity is, where it is in 3D space.

Why Add this Information?

- Linguistic features of the sentences may allow us to (partially) predict when during a performance a SRP will be established and where it will be placed in 3D space.
  - e.g. first noun phrase reference to an entity that recurs more than N times in the discourse.
  - e.g. signers may tend to place SRPs in favorite regions of 3D space: we may be able to learn a "cloud" of where in 3D space tend to go.
- We plan on using ML techniques to analyze.

Screenshot from our annotations using SignStream



Who's Annotating the Videos?

- One permanent member of the research team (graduate assistant), native signer.
- Each summer for five years (2009-2013), the lab will be hosting 2-3 deaf high school students from New York City.
  - Students selected based on being native signers and interest in science/computers.
- In summer 2009 and 2010, we also hosted a visiting undergraduate student who was a native ASL signer.

What do we have so far?

- Summer 2009:
  - 58 passages, 39 minutes, 5073 glosses
- Summer 2010:
  - 66 passages, 75 minutes, 7898 glosses
- Annotations, Videos (front, side, face), and Autodesk MotionBuilder files.
- We're still working on integrating the eye-tracking data into the files, and we want to do some more mocap clean-up work.

Not Enough Data Yet

- Trying to apply machine learning to model all of the challenging aspects of ASL (timing, space, verb inflection, coarticulation, etc.) will require a very large corpus.
  - Especially if you want to learn lexically specific information.
  - While lexicalized machine-learning models are mainstream in the computational linguistics community, the size of ASL corpora will remain small (relative to those for written languages) for the foreseeable future -- due to the time-consuming nature of sign language corpora creation and annotation.
- While we are gathering an ASL corpus, at the same time, we consider this a ***long-term investment***.
- For our short term research, we are gathering very specific data on particular ASL phenomena of interest, and using the design-cycle research paradigm described earlier.

Overview of This Talk

- Challenging aspects of ASL animation synthesis that are common to scripting or generation/MT.
- Sources for data-driven ASL animation research.
- Our lab's "design cycle" research paradigm.
  - Data from native ASL signers on phenomena of interest, creation of computational models of these phenomena, and user-based evaluation of our animations.
- Example: Our recent work on ASL verb-inflection.
- Using Motion-Capture in Our Research.
- Our Motion-Capture Corpus Collection Project.
- Summary and Future Work

Key Features of Our Research

- Use of linguistic data on targeted ASL phenomena of interest.
  - Various possible sources of such data.
- Use of experimental studies with native ASL signers answering comprehension questions about animations.
  - How we tell if our models are actually improving.
- The involvement of native ASL signers in the research process as informants, annotators, and research team members.

Short Term and Long Term Focus

- In the short term, we're using this directed data-collection, model-building, and evaluation approach to address particular phenomena for ASL synthesis.
- In the long term, we're constructing an ASL corpus based on motion-capture data with full linguistic annotations.
  - We have plans to use this corpus first to explore issues related to spatial reference points.
  - Should also be useful for many other studies.

Linguistic and Assistive Technologies Laboratory
Queens College of The City University of New York

Projects at the Lab:

- Generation of sign language animations.
- Text simplification for people with intellectual disabilities.
- User-interface design and evaluation studies with deaf users.

Ph.D. Students:
Pengfei Lu, Allen Harper, Josh Waxman

Undergraduate Students:
Rea Bhasin

High School Students:
Kelsey Gallagher, Sheldon Clarke, Aaron Pagan, Jaime Penzellna

Research Staff:
Jonathan Lamberton

Alumni, Former Visiting Students:
Lijun Feng(PhD), Jackson Yeh(MA), Amanda Krieger, Meredith Turtletuab

Funding Support for this Research

- NSF CAREER Award "Learning to Generate American Sign Language Animation through Motion-Capture and Participation of Native ASL Signers," 2008-2013, $581,496.
- Siemens/UGS: "Generating Animations of American Sign Language," Go PLM Grant, 2007-2011, $633,150.
- PSC CUNY Research Award "Evaluating Parameters for ASL Animation," 2007-2008.