Evaluation of a Psycholinguistically Motivated Timing Model for Animations of American Sign Language

Matt Huenerfauth
The City University of New York (CUNY)
Department of Computer Science, CUNY Queens College
Doctoral Program in Computer Science, CUNY Graduate Center
65-30 Kissena Boulevard, Flushing, NY 11367 USA
+1 718 997 3264
matt@cs.qc.cuny.edu

ABSTRACT
Using results in the psycholinguistics literature on the speed and timing of American Sign Language (ASL), we built algorithms to calculate the time-duration of signs and the location/length of pauses during an ASL animation. We conducted a study in which native ASL signers evaluated the ASL animations processed by our algorithms, and we found that: (1) adding linguistically motivated pauses and variations in sign-durations improved signers' performance on a comprehension task and (2) these animations were rated as more understandable by ASL signers.

Categories and Subject Descriptors
I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *language generation, machine translation*; K.4.2 [**Computers and Society**]: Social Issues – *assistive technologies for persons with disabilities*.

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
American Sign Language, Animation, Natural Language Generation, Evaluation, Accessibility Technology for the Deaf.

1. MOTIVATIONS AND RELATED WORK
American Sign Language (ASL) is a full natural language; signers use their hands, facial expression, eye gaze, head tilt, and body posture to convey linguistic meaning. ASL is not just a manual presentation of an English sentence; it has its own word-order, syntactic constructions, and vocabulary of signs (which may not have one-to-one equivalence with English words). ASL is used as a primary means of communication for about one half million people in the United States [15]. Because of the differences between English and ASL, it is possible to have ASL fluency yet significant difficulty reading English. In fact, a majority of deaf 18-year-olds in the United States have an English reading level below average 10-year-old hearing students [10]. Unfortunately, websites and other written-English information sources can also pose a challenge for deaf adults with low literacy skills. Software that can display computer-generated animations of American Sign Language can make more information, communication, and services accessible to these users. These ASL animations may be scripted by a content developer [13, 19] or generated by English-to-ASL automatic machine translation software [2, 12, 14, 17].

ASL signing conveys information at the same rate as spoken English [1], but the average speed at which most adults can *read* English text is much faster than the speed of *spoken* English audio. For deaf adults with low literacy to access information at a speed comparable to English-reading adults, ASL animations would need to be displayed quickly (while maintaining their understandability). The literature contains a range of values for "normal" signing speed: from 1.5 to 2.37 signs per second [1, 6]. When ASL videos are played faster than 2.5 times normal, viewers' comprehension of the video drops significantly [4].

Studies of English speech have shown a link between the presence of linguistic pauses and comprehension of the audio. One study increased the speed of an English speech recording, and researchers later inserted pauses at linguistically appropriate locations (between sentences, between clauses, between phrases) [21]. The pauses improved listeners' comprehension. This benefit only arose if pauses were at linguistically appropriate locations in the speech signal (not at random or uniform locations), and the benefit leveled off once the pauses had increased the duration of the whole recording back to its original length before the artificial speeding process [21]. Two explanations for this link between comprehension and the presence of pauses at linguistically-appropriate locations in speech were proposed: (1) pauses may help listeners more easily determine sentence/clause boundaries in the performance or (2) pauses at appropriate phrase boundaries give the listener some additional time to mentally "chunk" units of information and process/remember them more effectively.

In one ASL study, black-screen segments of video were inserted into a double-time video recording of ASL between "semantically unitary statements" [9]. (We understand this to mean that black frames were added between sentences/clauses.) Blank segments were of uniform duration, and enough were added to return the ASL video to its original time. No significant improvement in viewers' comprehension resulted. It has not been examined

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ASSETS'08, October 13–15, 2008, Halifax, Nova Scotia, Canada.
Copyright 2008 ACM 978-1-59593-976-0/08/10...$5.00.

whether inserting pauses (signer remains on screen but the hands stop moving) at linguistically appropriate locations would impact viewers' comprehension of videos of human signers. This paper will examine whether inserting pauses into computer-generated animations of ASL improves comprehension or makes the resulting ASL animations seem more natural-looking to viewers.

Several projects have focused on the synthesis of animations of virtual humans performing sign language [3, 5, 12, 16] (and surveys in [12, 13]). For example, several years of European projects have contributed to the eSIGN project, which creates technologies for content developers to build sign databases in a symbolic notation, assemble scripts of signing performance for use on accessible web pages, and allow viewers to see animations on their web browser [13]. SignSmith Studio [19], discussed in section 3.1, is a commercial tool for scripting ASL animations. Other research has examined automatic generation or machine translation (MT) of sign language [2, 12, 14, 17], but published work does not focus on the timing of sign language animations. Content developers may script individual signs while observing the timing of human signers in video [13] or use motion-capture technology to record individual signs from humans directly [3], but this addresses the timing of isolated signs (not sentences). Also, many sign language animation systems give the *viewer* the ability to adjust a dial that modifies the speed of the performance [3, 13]; however, section 2 will discuss how the speed of an ASL performance is more complex than a single speed value. Previous sign language animation research has not examined how the time-duration of a sign is affected by its surrounding linguistic context (what other signs occur in a sentence or in a performance) nor how pauses should be placed in an animation to mimic how human signers tend to pause at natural linguistic boundaries. SignSmith allows the content developer to *manually* specify pauses to occur, and content-scripting tools from the eSIGN project give similar control over speed, timing, and pauses [13]. Animations from generation or MT projects do tend to include pauses between sentences [3, 5, 12], but a principled linguistic way to select where to insert pauses into a sign language animation has not been described previously in the literature.

2. LINGUISTIC BASIS FOR THE DESIGN

The timing of an ASL performance is more complex than a single "speed" variable (number of signs per second). In fact, many parameters are needed to specify the speed of an ASL performance: speed of fingerspelling relative to speed of other signs, time spent in transition movements between signs, time spent pausing during signing, etc. Further, ASL signs are not all performed in the same amount of time: each sign has a standard time duration at which it is performed. Thus, the final timing of ASL is a complex interaction between several speed parameters and the lexical durations of the specific signs performed. Although several psycholinguistic experiments have examined human ASL speed and pausing [1, 4], the most detailed analysis was conducted by Grosjean and Lane [6, 7, 8] in the late 1970s.

They studied the interaction of three component variables of signing rate: (1) articulation rate at which the hands move through space, (2) number and location of pauses in a sentence, and (3) length of each pause. They define a "pause" as a period of time between two signs when the hands are not moving – there is not a pause between every pair of signs [8]. In their view, signs consist of an "in-transition" to get the hands in the proper position, the main movement of the sign, and an optional "hold" at the end of a sign where a pause could occur [6].

They observed that ASL signers (in videos) perform signs before sentence boundaries more slowly (12% longer) than their normal duration [6]. Also, when a sign occurs more than once in a performance, then the durations of the later occurrences differ from the typical duration for that sign. If later occurrences of the sign appear in a syntactic position where the sign has appeared before (e.g. as the subject or as the direct object of the sentence), then the later occurrences are 12% shorter. If the later occurrence of a sign appears in a new syntactic position where it had not previously appeared, then the later occurrence of the sign is 12% longer [6]. For example, if a sign appears early in a performance in the subject position of a sentence, and the same sign is used as a direct object in a later sentence in the performance, then the second occurrence is longer.

3. TWO ALGORITHMS FOR ASL TIMING

Based on these ASL linguistics studies, we have built algorithms for calculating the duration of signs and the location/length of pauses during an ASL animation. Our two algorithms thus attempt to set timing values for an ASL animation so that it mimics the behavior of human signers. Our goal is to improve the understandability and perceived naturalness of ASL animations.

3.1 Using SignSmith Studio for Prototyping

To build a prototype of our algorithms and evaluate whether they produce ASL animations that signers find more understandable and natural, we had to create animations for signers to examine. In earlier work, we built a system for generating animations of a character performing ASL sentences (containing constructions called "classifier predicates," which describe 3D spatial concepts) [11]. Our generator was designed with a primary focus on these constructions, and it has a small *ad hoc* vocabulary of signs in its repertoire, which had been expanded as needed to construct full sentences for past evaluation studies. To conduct tests of algorithms that operate on sign-duration and pause-insertion on wide variety of ASL sentences (not just classifier predicates), we needed an animation platform with a larger vocabulary of signs. We have selected to use SignSmith Studio, a commercial system from VCom3D [19]. This product allows users to script an ASL performance (using a dictionary of signs, a fingerspelling generator, limited eye-gaze and head tilt, limited shoulder-tilt, a list of around 50 facial expressions, optional speech audio and mouth movements, and other ASL features). The user is presented with an interface that looks like a set of parallel tracks (like a musical score), and the user arranges signs and facial expressions on these parallel timelines. When a "play" button is pressed, then an animation is generated in which a virtual human character (there are several to choose from) performs the script.

The advantage of using a commercial system that we did not develop is that the movements and speed/durations of the signs in its dictionary were developed external to our research project. The signs were not built solely for the purpose of this study, and so, each sign's duration in the dictionary was set independently of our experiments. Another advantage of SignSmith is that its representation of the timing of signs is compatible with that of our generation software [12]; so, progress made running experiments on their animations can translate into later improvements for our

ASL generation system under development. Specifically, SignSmith stores three timing values for each sign: (1) transition time during which the hands get to the starting position, (2) main movement of each sign, and (3) hold time at the end of the sign when the hands remain motionless. In SignSmith, the user can manually override the default values for these three sub-times for each sign in the animation. Signs have a basic "duration" value for their main movement, and this is multiplied by a "multiplier" factor (that the user may optionally specify) to vary the time of the sign's main movement in the resulting animation.

To produce ASL animations, users of SignSmith are expected to be knowledgeable of the language; however, even fluent signers may not have intuitions about how the speed and pauses of a performance should be numerically specified to create a natural result with good timing. The documentation for SignSmith mentions that users may want to insert longer transition times between two signs at a sentence boundary. (So, this appears to be the recommended approach for users to manually add pauses between sentences.) For this project, we created several multi-sentence ASL performances using SignSmith – leaving the default timing values for each sign in the script (more details in section 4.2). The script for these animations was used as input to the sign-duration and pause-insertion algorithms discussed below.

3.2 Sign-Duration Algorithm

We have implemented an algorithm for calculating linguistically motivated durations for the signs in an ASL animation – based on the standard duration for each sign and its surrounding linguistic context. The input to our algorithm is an initial script of the ASL performance from SignSmith (sequence of signs with standard durations) and some linguistic information: (1) part-of-speech for each sign (with nouns sub-categorized according to their syntactic role as described below) and (2) sentence/clause boundaries identified. We implemented our algorithm using the Python programming language (good for processing strings of text), and it operates on the script produced by SignSmith. The output is a script in which the durations of some signs are modified.

Our two-phase algorithm builds upon linguistic research in section 2. In phase 1, signs occurring more than once in the script are detected. Only content signs like nouns, verbs, adjectives, and adverbs are modified in this sign duration algorithm – function signs like prepositions, pronouns, or other grammatical markers are not. If the repeated sign is a verb, adjective, or adverb, then later occurrences of the sign are shortened by 12% in duration. If the repeated sign is a noun, then changes to later occurrences depend on the syntactic role of each occurrence (it may lengthen or shorten by 12%). Nouns are categorized as being in: topic-position, when/conditional-clause position, subject position, direct/indirect object position, object-of-preposition, etc. (Topic and when/conditional clauses occur at the beginning of ASL sentences and are accompanied by special facial expressions.) In phase 2 of the algorithm, signs that appear just before sentence or clause boundaries are lengthened (by 12% or 8%, respectively).

3.3 Pause-Insertion Algorithm

We have also implemented an algorithm for determining where to insert pauses (and how long they should be) during a multi-sentence ASL animation. The input to the algorithm is a script of an ASL performance that includes the sequence of signs with the time duration of each (output of the previous algorithm can be input to this one). Our algorithm also requires linguistic data: (1) location of sentence boundaries and (2) a syntactic parse tree for each sentence. (This data is again supplied manually to the algorithm for this study; tools for automatically parsing ASL sentences are a focus of future work.) The algorithm's output is a script of the ASL performance in which the "hold" times at the end of signs have been modified to produce pauses during the performance at linguistically appropriate locations. Like the sign-duration algorithm above, this algorithm has been implemented as Python code that operates on the ASL scripts produced by the SignSmith animation system, but it could be modified to accommodate other file formats containing similar information.

3.3.1 The Original Grosjean and Lane Model

Our algorithm builds upon ASL psycholinguistic research on speed and timing. When analyzing video recordings of human ASL signers, Grosjean and Lane (G&L) proposed a model to account for where in a performance signers would insert pauses (and how long pauses would be) [7, 8]. Their model assumed that a syntactic parse tree for the sentence was known, and it predicted the percentage share of total pause time that should be inserted between adjacent pairs of signs. Their model only accounted for pauses in a single sentence – not for a multi-sentential passage. We have used their model as a basis for our multi-sentential ASL pause-prediction algorithm. The key idea of their model was that two factors determined how much pause time occurs between two signs: (1) the syntactic significance of that boundary and (2) whether the boundary is near the midpoint between the nearest pauses. Thus, boundaries between sentences, between clauses, and between major phrases in a sentence were preferred locations for pauses, but this preference was balanced with a "bisection tendency" [7] to pause near the middle of long constituents.

G&L describe an iterative method for calculating the percentage share of total pause time that should be inserted between each pair of adjacent signs in a performance [7]. They first assign a syntactic "complexity index" value (CI) to each sign boundary; this value is calculated using the syntactic parse tree for the sentence. In the parse tree, there exists some lowest node that spans the boundary between each pair of adjacent signs; the syntactic importance of that node determines the CI value of that boundary. Specifically, the total number of descendant nodes below that node in the parse tree is the CI value for the boundary. Thus, the boundary between two clauses would have a larger CI value than the boundary between two signs inside a noun phrase.

G&L's method iterates until all boundaries between signs have been assigned a percentage share of pause time. One pause is added during each iteration. An iteration of the algorithm begins by selecting the longest span of signs not yet broken by a pause (based on number of signs, not on sign durations). For each boundary between adjacent signs within that span, the relative proximity (RP) of the boundary to the midpoint of the span is calculated (RP = 100% at the midpoint, RP = 0% at the ends of the span, etc.). For each boundary inside the span, the CI value is multiplied by the current RP value. The boundary with the greatest (CI*RP) product inside the span is chosen as a location for a pause; the percentage share of pause time assigned to that boundary is calculated based on the product. The algorithm then iterates (selecting the longest remaining unbroken span of signs in the whole performance and then calculating fresh RP values for that span under consideration) [7].

```
Algorithm: Pause-Insertion
Input: Multi-sentence script with CI values for each sign boundary.
Output: Script with pause-share (PS) values assigned at boundaries.
  Loop until all boundaries are assigned a pause share (PS) value:
      Select longest span of signs without PS values between them.
      Call AssignOneBoundary( the_longest_span ).
  Identify the boundaries with the top 25% of PS values.
  Assign pauses to the top-25% boundaries, proportional to their PS.
Algorithm: AssignOneBoundary
Input: Span of signs in an ASL performance with no PS values yet.
Output: Assign a PS value to one boundary in this span.
  For each sign boundary inside this span:
      Calculate RP for this boundary (as done by G&L model).
      Calculate the product (CI * RP) for this boundary between signs.
  Find boundary with highest (CI*RP) value.
  Set the pause share for this boundary as follows: PS = (CI*RP).
```

Figure 1. Pseudocode for Pause-Insertion Algorithm.

3.3.2 Our Pause-Insertion Algorithm

Our algorithm implements and extends the G&L method in several ways. First, it takes into account the results of our sign-duration algorithm. When calculating the RP values, the G&L model operates on signs as if they were all uniform unit duration. Our algorithm uses the actual timing values (in-transition time + sign-duration) of the signs in a span to calculate the RP values.

Second, we had to extend the G&L model to account for multi-sentential performances. Syntactic parse trees span over single sentences; so, the CI for a boundary between sentences is undefined in the G&L model. Our algorithm sets the CI value of the sentence boundary between any two sentences (*S1* and *S2*) as equal to *min*(18, *length*(*S1*)+*length*(*S2*)-2), where "*min*" is the minimum function and "*length*" is the number of signs in each sentence. The logic behind this is that if the two sentences had been joined by a conjunction, then the root of the parse tree that joins them would dominate all of the nodes in the tree for *S1* and all nodes in the tree for *S2*. Assuming binary-branching syntactic parse trees, the number of internal nodes in a tree for a sentence *S* would be *length*(*S*)-1. To ensure sentence boundaries adjacent to short sentences still receive sufficient syntactic heft, our algorithm will assign a CI value of at least 18 to any sentence boundary.

Our algorithm uses other results from the psycholinguistic literature. ASL signers insert a pause at 25% of the boundaries between signs [6]; so, our algorithm adds pauses at boundaries that have been assigned the top 25% of pause-percentage weight by the iterative method. ASL signers spend about 12% of their time pausing during rehearsed sentences [6] and 10%-35% during spontaneous signing [1]. So, we insert pause-time into the animation such that 17% of the final animation time is spent in pauses (a middle-ground between published percentages). Pause time is added to the "hold" time of the sign before the boundary. Figure 1 shows pseudocode for our pause-insertion algorithm. Boundaries with the top 25% Pause-Share (PS) values in Figure 1 receive a share of the pause time proportional to their PS value.

4. DESIGN OF OUR EVALUATION STUDY

We conducted a study in which native ASL signers evaluated the animations processed by our timing algorithms to test two hypotheses: (H1) adding linguistically motivated pauses and variations in sign-durations will help ASL signers understand and remember information from these animations and (H2) these new ASL animations will appear more natural-looking to signers.

4.1 Participant Recruitment and Interaction

In an earlier evaluation study [11], we outlined a set of best-practices for the conduct of studies involving ASL signers that are designed to evaluate ASL animations. We described how it was important that the signers recruited for the study be native signers, how to best ask questions to screen for such signers, and how the experimental environment around signers should be ASL-focused (with as little English or English-like signing as possible). All of these factors help to ensure that the responses given by participants about the correctness of the animations are as ASL-accurate as possible. Non-native signers who learned ASL later in life may be more lenient when judging ASL animations, and signers subjected to an English environment may switch their own signing to a more English-like form. This can also result in their being overly tolerant of animations that are too English-like [11]. For the current study, all instructions and interactions were conducted in ASL, and 8 of the 12 participants arrived accompanied by another ASL signer (thereby producing an ASL conversational environment immediately prior to the study).

Advertisements posted on Deaf community websites in New York City asked whether potential participants had grown up using ASL at home or whether they attended an ASL-based school as a young child. Of the 14 people who came to participate in the experiment, 2 answered pre-screening questions in such a way that they did not meet the screening criteria. Their data was excluded from the study. Of the 12 participants whose data was included, nine grew up with parents who used ASL at home. Of the remaining three, two began attending a school using primarily ASL before the age of 7, and the final participant began using ASL before the age of 7 through another circumstance. Of our 12 participants, 5 were married to someone deaf/Deaf, 9 used ASL as the primary language in their home, 11 used ASL at work, and 11 had attended a college where instruction was primarily in ASL. There were 7 men and 5 women of ages 25-58 (median age 37).

4.2 Animations Shown in the Study

Twelve ASL passages of length 48-80 signs (median 69 signs) were created in SignSmith on a variety of topics: four short news stories, two adaptations of encyclopedia articles, four fictional narratives, and two personal introductions. Passages contained sentences of a variety of lengths and complexity; some included topicalized noun phrases, condition/when clauses before a main clause, rhetorical questions, contrastive role-shift (signers may tilt their shoulders to the left or right as they contrast two concepts), or association of entities with locations in space (for use by later pronouns during which the signer points to those locations).

An ASL interpreter verified the accuracy of the twelve animations – to a degree. While SignSmith gives the user many options in controlling the animated character, there are many phenomena in fluent ASL signing that are beyond the capabilities of the system: inflection of ASL verbs for locations in 3D space, separate use of head-tilt and eye-gaze during verb signs to indicate subject/object agreement, association of entities under discussion with more than two locations around the signer, etc. In addition to these limitations, SignSmith's dictionary does not contain all possible ASL signs. Therefore, the animations produced for this study are not perfectly fluent ASL. To evaluate our timing algorithms, we feel that "rather fluent" ASL animations are acceptable. State-of-the-art ASL generation technology will not be able to produce

Figure 2. Screenshot from an animation.

Figure 3. Illustration of the sequence of signs in an original animation (top row) and after being processed by our timing algorithms (bottom row). Additional space between signs represents pauses added to the animation script by our algorithms.

Figure 4. Sample set of answer choices for one comprehension task question.

"fully fluent" ASL animations for many years in the future. So, the degree to which the timing algorithms we have developed can improve the understandability and naturalness of semi-fluent ASL animations is still an important research question (and is perhaps a more realistic evaluation our timing algorithms). See Figures 2 and 3 for a screenshot and timeline of animations from our study.

4.3 Responses and Experimental Design

Participants viewed animations of six types: 2 groups (no-pauses vs. pauses) × 3 speeds (normal, fast, very fast). *No-pauses* animations were not processed by our timing algorithms; *pauses* animations have been processed by our sign-duration and pause-insertion algorithms. We also examined how quickly animations can be played while remaining understandable. *Normal*-speed animations are at a rate of 1.5 signs per second, *fast* are 2.25 signs per second, and *very fast* are 3 signs per second. (Values for average signing speed in the linguistics literature vary from 1.5 to 2.37 signs per second [1, 6].) Another reason to study both pause-insertion and speed in one study is that we can determine whether any benefit observed from pause-insertion is from (1) pauses being placed at linguistically appropriate locations or (2) simply from the additional time added to the animation.

Each of the 12 ASL passages was generated in all six combinations of group × speed, producing 72 animation files. A fully-factorial within-subjects design was used to assign 12 files to each participant in the study such that: (1) no participant saw the same passage twice, (2) the order of presentation was randomized, and (3) each participant saw two animations of each of the six combinations of group × speed. Animations were viewed on a 17" LCD screen at a distance of less than one meter. The animations occupied a 10cm×10cm region of the screen.

Several formal evaluation studies of ASL animations have been conducted that involve native signers [11, 12, 13, 16, 20]. In our earlier study [11], participants saw ASL animations and were asked to circle numbers on 10-point Likert-scales to indicate how Grammatical, Understandable, and Natural-moving the animations were. Participants in that study appeared comfortable with the instructions and giving their opinions about these aspects of the ASL animations. We have used the same three subjective criteria in this study. For this study, we have added an additional Likert-scale to enable participants to indicate whether the ASL animation is too-slow, perfect, too-fast, or somewhere in-between. As in our previous study, instructions were given to participants in ASL to explain the meaning of each of these Likert-scales.

Evaluations of the understandability of sign language animations have given viewers various tasks to demonstrate understanding of the animation: decide if two animations say the same thing [20], match a signing animation to a movie of what was described [11, 12], summarize the animation [16], or answer comprehension questions about the animation's content [16]. Studies of ASL videos have also pioneered techniques useful for evaluation of animation [1, 4, 6, 7, 8, 9, 18]. Heiman and Tweney [9] showed signers multi-sentence ASL videos followed by ASL videos of comprehension questions. Their participants wrote down English answers to the questions. Tartter and Fischer [18] asked signers to select a cartoon drawing that corresponded to the content of video of an ASL sentence they had seen.

In our study, after viewing the ASL passage, participants will answer the four Likert-scale questions discussed above. Then, participants are shown a set of four comprehension questions about the information in that ASL passage. The same animated signing character used during the passage performs these four questions; after each question, the animated character gives a list of multiple-choice answers. These answers correspond to a list of cartoon clip-art pictures on a paper survey form, and the participant circles the correct answer(s) to each question. We have adapted the design of Heiman and Tweney so that our participants do not need to write any English text during the experiment – they circle a picture as their answer. We had planned to omit labels below the cartoons to avoid English influence in the environment; however, during pilot tests, participants requested we add short English captions (Figure 4). We included an example story with comprehension questions prior to the data collection process. This allowed participants to become comfortable seeing the animated signing character, and it ensured that the instructions for the study were clear.

Most questions ask basic who/what/when/where facts from the passage, and about 10% are less direct. For an example of a less-direct question, in one passage, a person is said to be vegetarian, and a later question asks what foods this person eats – choices include: hamburgers, hot dogs, salad, etc. To minimize the effect of variation in skills across participants, questions focused on shallow comprehension. Also, our "comprehension" questions actually measure a mixture of the participant's recall and

comprehension of the passage. Participants were not allowed to watch the passage more than once; so, they could not replay it to look for an answer. (Participants could replay the questions.) Since ASL animations should both be comprehensible and convey information memorably, we decided it was acceptable for our questions measure both recall and comprehension. In future work, we may study these two phenomena more independently.

After viewing 12 ASL passages and answering the corresponding Likert-scale and comprehension questions, subjects gave open-ended feedback about the animations. They were shown two animations (with pauses included) while they gave their feedback to give them something specific to comment about if they preferred. Participants were given the option to sign their comments in ASL or to write them down in English themselves.

Statistical tests to be performed on the data were planned prior to data collection. To look for significant differences between scores for Grammaticality, Understandability, Naturalness, Speed, and Comprehension-task performance, a Kruskal-Wallis test was performed (this is an extension of a Mann-Whitney U-test with corrected p-values for multiple groups). Non-parametric significance tests were used because the Likert-Scale response data was not known to be normally distributed. In addition, tests for correlation were planned between the major dependent variables (G, U, N, S, and C), and between various demographic variables (subject's age, subject's gender, the presentation order in which individual animations were seen by that subject, etc.).

5. RESULTS OF EVALUATION STUDY

To score the comprehension task for each passage, we subtracted the number of correctly circled pictures minus 25% of the incorrectly circled pictures. This difference was then divided by the highest possible score for that question (to enable comparisons across questions with different numbers of correct answers). Figure 5 shows the average comprehension task score for the no-pauses vs. pauses groups; the white bar represents the average across all responses for that group, and the shaded bars represent the values for each speed sub-group (normal, fast, very fast). Tests for statistical significance were performed between the pause and no-pause groups for each speed level and for the combination of responses across all speeds. Statistically significant differences between pairs of compared values are marked with an asterisk.

Figure 6 shows average Likert-scale responses for each of the six animation types for all three speeds. The graphs include white bars that indicate the average value across 72 responses, while the shaded bars for each speed level indicate the average of 24 responses for animations of that type (group × speed).

It is somewhat unusual for S to be shown on the same graph as G, U, and N in Figure 6. While G, U, and N are reported on 1-to-10 scales, S was reported on a 21-point scale: from 1 ("too fast") to 10 ("perfect") to 21 ("too slow"). Out of the 144 S scores collected, only six scores were above 10 (range 11-13, median 11.5), all for *normal* animations (3 *pauses*, 3 *no-pauses*). Another important note about Figure 6 is that significant differences are indicated only between *pause* vs. *no-pause* pairs of animations. Impact of speed differences alone was not a focus of this project.

The use of our timing algorithms led to a significant increase in Comprehension task performance. This was true in the overall case (all speeds combined), and it was also significant when only considering the "normal" speed data. We also see significant differences for the scores participants gave to the animations for Understandability. Since the "normal" speed animations received Speed scores closest to 10 (Perfect), it is important that our Comprehension and Understandability trends hold true in the

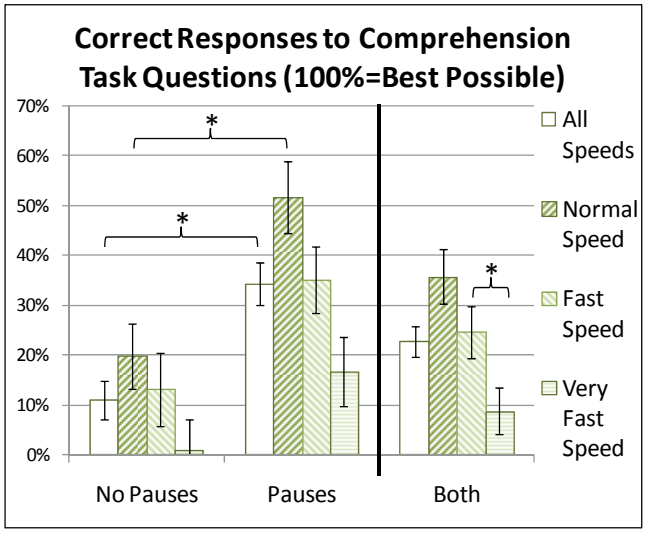

Figure 5. Graph of comprehension task scores.

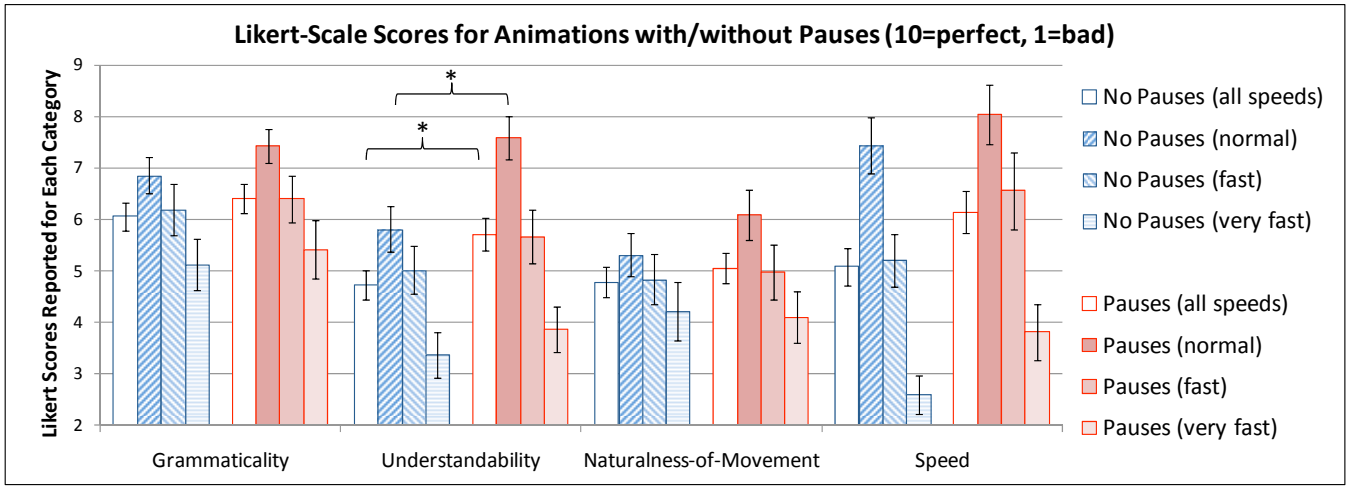

Figure 6. Graph of Likert-Scale scores for subjective responses from the evaluation study.

"normal" speed case. One of our hypotheses was not supported by the data; we did not measure significant difference in the Naturalness scores of animations in the pauses vs. no-pauses groups. No significant interaction effects were observed between the group (no pauses, pauses) and animation-speed (normal, fast, very fast) for any of the categories of response (G, U, N, S, or C).

There are two ways to define the speed of an animation with pauses: the signs-per-minute of the original animation or the recalculated signs-per-minute taking into account the additional time from pauses. In this study, we have used the first definition to label animations as normal, fast, or very fast; so, a "normal" animation without pauses is shorter than the corresponding "normal" animation with pauses. So, we should examine whether it was actually important that the pauses are in linguistically appropriate locations or whether the increased C scores arose merely from the additional time added to the animation. Figure 7 shows C scores plotted according to the total number of signs divided by total time (including pauses). If the only benefit of adding pauses had been from the change in total animation time alone, then we'd expect these two lines to overlap. It appears that the placement of pauses at linguistically appropriate locations has a benefit beyond just an overall increase in the animation time. In future work, we will compare: (1) animations without pauses and (2) animations with pauses that have later been sped up so that they have the same total time as the original no-pauses animation.

Table 1 shows the correlations calculated between scores for G, U, N, S, and C. Statistically significant values ($p<0.05$) are marked with an asterisk. Understandability was the value best correlated with a participant's success on the Comprehension task for an animation, but it was not a strong correlation (R=0.473). In an earlier study [11], we also saw low R-value between Understandability and Comprehension task success. In that study, the animations were single sentences containing ASL phenomena called classifier predicates (not present in animations for this study). Low correlation between U and C suggests that asking participants to rate the understandability of animations is no substitute for measuring comprehension task performance. There is a difference between perceived and actual understanding.

Correlations were also examined between the order in which each animation was shown to a participant and the evaluation scores.

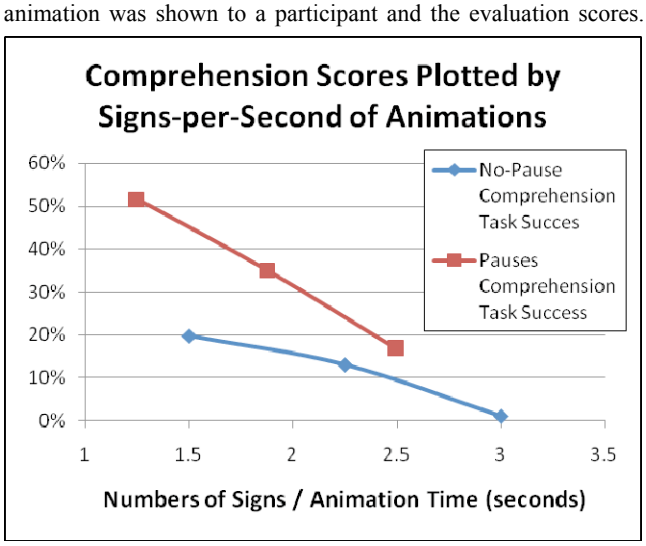

Figure 7. Comprehension scores plotted by signs-per-second.

Table 1. Pearson's R-values between Compared Values

	Gram.	Under.	Natural.	Speed
Under.	0.687*			
Natural.	0.537*	0.673*		
Speed	0.506*	0.761*	0.473*	
Compre.	0.406*	0.473*	0.317*	0.413*

There was a slight correlation between presentation order and reported speed score (R=0.11) and between presentation order and comprehension task success (R=0.15); however, neither of these was statistically significant. Since we presented the individual passages in this study in randomized order for each participant, then the practice effect would have minimal impact on the results.

There were weak but significant negative correlations between a participant's age and their scores for G, U, and N (R=-0.26, -0.22, -0.29 respectively). There was no significant correlation between age and Speed nor between age and Comprehension success. So, older participants rated animations more critically but this did not lead to differences in their rating of the animation's Speed nor success at the Comprehension task. No age-related differences were observed in Comprehension scores for pauses vs. no-pauses.

Most feedback comments from participants were on aspects of the animation inherent to SignSmith or the specific passages (quality of facial expression, movement of individual signs, geographic or regional specificity of some signs, "stiffness" of the character's torso, or limited association of entities with points in 3D space for pronouns). Interestingly, no one commented on the presence or absence of pauses. This feature of the animations did not seem to draw their overt attention. Some participants' comments were relevant to the timing. For example, nine participants mentioned that the animations at the "normal" speed were still too fast, but three felt the speed was OK. Three participants felt that the fingerspelling was relatively too fast compared to other signs. In the "normal" speed animations, fingerspelling occurred at a rate of 4.1 letters per second with a 0.8 second hold at the end.

6. CONCLUSION AND FUTURE WORK

This project has identified a way to improve the comprehension of ASL animations by modifying two parameters that have received little systematic research attention before: insertion of linguistically-appropriate pauses and modifying the duration of signs based on surrounding linguistic factors. This is the first study to show that signers perceive a difference in the quality of ASL animations based on these speed and timing factors. While this study focused on ASL, these techniques should be applicable to animations of other sign languages around the world. Other contributions of this work include: (1) a prototyping/experimental framework for measuring comprehension of ASL sentences (so that additional factors can be explored in later work) and (2) motivation for future computational linguistic work on ASL.

Now that this study has shown that ASL animations can be made more understandable if we have a syntactic parse of the sentence and a part-of-speech for each sign, there is motivation to design automatic tools to identify this linguistic information. Currently, for our algorithms to be used in a real-world setting as part of a tool like SignSmith, the user (who is scripting an ASL animation) must supply information: (1) clause/sentence boundaries, (2) part-of-speech of each sign, (3) syntactic role of each noun, and (4) a

syntactic parse tree for each sentence. It may be possible to develop automatic natural language processing tools to assign part-of-speech tags to signs in a sentence and provide some form of syntactic parse of the sentence that has been scripted (to automate the process). If our sign-duration and pause-insertion algorithms were used as part of an ASL animation generator (that automatically plans an ASL sentence – as in the final step of an English-to-ASL machine translation system), then the linguistic data required by our algorithms would have already been calculated by the generator as part of its planning work.

In future work, we will use the experimental techniques developed in this study for measuring sentence comprehension to evaluate slower animations (since the animations in this study were rated as too fast by participants) and modify the relative speed of fingerspelling vs. other signs. ASL psycholinguistic studies have also suggested that the overall speed of a performance may affect where pauses are inserted (and their length), these studies have also suggested syntactic, discourse, and emotional features [6] that affect speed and timing of ASL. We may modify our sign-duration and pause-insertion algorithms to incorporate these other factors. Finally, it was an earlier study of ours on ASL classifier-predicate sentences that first prompted us to explore the issue of ASL speed and timing [11]; so, we also plan to develop timing algorithms for those ASL constructions.

When researchers like Grosjean and Lane [8] described patterns in ASL sign-durations over thirty years ago, they were analyzing the data as a matter of linguistic interest. It would have been impossible to predict the development of computer and animation technology that has made the generation of ASL animations possible today – opening up a new application for their research. The promising initial results of this study open the question of whether there may be additional published linguistic research on ASL that can be applied to current research on ASL animation. This investigation paradigm (of seeking inspiration from the ASL linguistics literature, creating prototype systems to explore the potential benefits of new algorithms, and conducting controlled user experiments to measure any benefit) may lead to additional advancements in the state-of-the-art of ASL animation.

7. ACKNOWLEDGMENTS

This work was supported (in part) by a grant from The City University of New York PSC-CUNY Research Award Program and by a grant from Siemens A&D UGS PLM Software (Generating Animations of American Sign Language, Go PLM Grant Program, 2007). Allen Harper and Julian Bayona assisted with data analysis and the preparation of experiment materials.

8. REFERENCES

[1] Bellugi, U., Fischer, S. 1972. A comparison of sign language and spoken language. *Cognition*, **1**:2-3.

[2] Chiu, Y.-H., Wu, C.-H., Su, H.-Y., Cheng, C.-J. 2007. Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**:1.

[3] 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.

[4] Fischer, S., Delhourne, L., Reed, C. 1999. Effects of rate of presentation on the reception of American Sign Language. *J Speech Lang Hear Res*, **42**.

[5] Fotinea, S.-E., Efthimiou, E., Caridakis, G., Karpouzis, K. 2008. A knowledge-based sign synthesis architecture. *Univ Access Inf Soc*, 6:4.

[6] Grosjean, F. 1979. A study of timing in a manual and a spoken language: American sign language and English. *J Psycholinguist Res*. **8**:4.

[7] Grosjean, F., Grosjean, L., Lane, H. 1979. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, **11**.

[8] Grosjean, F., Lane, H., Teuber, H. and Battison, R. 1981. The invariance of sentence performance structures across language modality. *J Exp Psychol Hum Percept Perform*, **7**.

[9] Heiman, G., Tweney, R. 1981. Intelligibility and comprehension of time compressed sign language narratives. *J Psycholinguist Res*, **10**:1.

[10] Holt, J.A. 1993. Stanford Achievement Test - 8th Edition: Reading comprehension subgroup results. *American Annals of the Deaf*, **138**.

[11] Huenerfauth, M., Zhou, L, Gu, E., and Allbeck, J. 2007. Evaluating American Sign Language generation through the participation of native ASL signers. *In Proc. of ACM SIGACCESS Conf. on Comp. and Accessibility (Assets '07).*

[12] Huenerfauth, M. 2006. *Generating American Sign Language classifier predicates for English-to-ASL machine translation*. Dissertation, Comp. & Info. Science, U. Penn.

[13] Kennaway, J., Glauert, J., Zwitserlood, I. 2007. Providing signed content on the Internet by synthesized animation. *ACM Trans. on Computer-Human Interaction,* **14**:3.

[14] Marshall, I., Sáfár, E. 2005. Grammar development for sign language avatar-based synthesis, In *3rd International Conference on Universal Access in HCI*, Las Vegas, NV.

[15] Mitchell, R., Young, T., Bachleda, B., & Karchmer, M. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies*, **6**:3.

[16] Sheard, M., van der Schoot, S., Zwiterslood, I., Verlinden, M., Weber, I. 2004. Eval. reports 1 & 2, EU project *Essential sign language information on government networks*.

[17] Stein, D., Bungeroth, J., Ney, H. 2006. Morpho-syntax based statistical methods for sign language translation. In *Proc. of the European Assoc. for Machine Translation*.

[18] Tartter, V., Fischer, S. 1983. Perceptual confusion in ASL under normal and reduced (point-light display) conditions. In J. Kyle & B. Woll (eds.), *Language in sign: An international perspective on sign language*, London: Croom Helm.

[19] Vcom3D. 2008. Company website. http://www.vcom3d.com

[20] Vink, M., Schermer, T. 2005. Report of research on use of an avatar compared with drawings or films of gestures. Technical Report, Dutch Sign Language Center.

[21] Wingfield, A., Tun, P., Koh, C., Rosen, M. 1999. Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speed. *Psychol Aging*, **14**:3.