Empirical Investigation of Users' Preferred Timing Parameters for American Sign Language Animations

Sedeeq Al-khazraji
Rochester Institute of Technology
Rochester, NY 14623, USA
University of Mosul
Mosul, Iraq
sha6709@rit.edu

Becca Dingman
Matt Huenerfauth
Rochester Institute of Technology
Rochester, NY 14623 USA
bad6955@rit.edu
matt.huenerfauth@rit.edu

Abstract
To make it easier to add American Sign Language (ASL) to websites, which would increase information accessibility for many Deaf users, we investigate software to semi-automatically produce ASL animation from an easy-to-update script of the message, requiring us to automatically select the speed and timing for the animation. While we can model speed and timing of human signers from video recordings, prior work has suggested that users prefer *animations* to be slower than *videos* of humans signers. However, no prior study had systematically examined the multiple parameters of ASL timing, which include: sign duration, transition time, pausing frequency, pausing duration, and differential signing rate. In an experimental study, 16 native ASL signers provided subjective preference judgements during a side-by-side comparison of ASL animations in which each of these five parameters was varied. We empirically identified and report users' preferences for each of these individual timing parameters of ASL animation.

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3382989

Author Keywords
American Sign Language; Accessibility for Deaf or Hard-of-Hearing; Natural Language Processing; Speed; Timing; Human Computer Interaction.

CSS Concepts
• **Human-centered computing → Accessibility → Empirical studies in accessibility.**

Introduction
ASL is a primary means of communication for about 500,000 people, primarily in the United States [9]. ASL is a natural language that incorporates movements of the hands, face, and body to convey meaning [2]. While many Deaf and Hard of Hearing (DHH) individuals have strong English literacy, there is great variation in the DHH community in this regard. Standardized testing has revealed, on average, lower levels of

English reading literacy among deaf adults in the U.S., as compared to their hearing peers [12]. This creates a barrier when users encounter complex English text on websites. While companies could add video of human ASL signers to their websites, this is rarely seen on the web. The challenge is that in order to update content online, new videos must be recorded. This motivates our work on software to generate understandable ASL animations (of a virtual human signer) automatically from an easy-to-update script. The challenge is that this software must configure the details of the animation so that the avatar movements are accurate and easily understood by ASL signers. Prior work has found that adding linguistic pauses and adjusting signing rate improves understandability of ASL animations, with DHH individuals sensitive to tiny errors in these parameters [6, 7]. In prior work, we have used a corpus of videos and motion-capture recordings of human ASL signers to build models to predict ASL timing parameters [1]. Rather than a single "speed" value, the timing of an ASL animation is determined by multiple timing parameters, including:

- Sign duration: The average duration of words is based on the original speed at which words were encoded in the dictionary, as well as an adjustable rate factor that can be set for an animation.
- Transition: This parameter focuses on the time in-between individual ASL words, i.e. the number of seconds that signer's hands spend moving from the end of one sign to the beginning of the next.
- Pausing frequency: This parameter represents how often the signer pauses (e.g. prosodic pauses in between syntactic phrases or sentences) during ASL signing, as expressed as a percentage of the inter-sign locations at which a pause occurs.
- Pausing length: The average number of seconds that signer's hands stop moving when a pause occurs.
- Differential signing rate: The speed of signing varies during a message, e.g. slowing down at the end of phrases. This parameter reflects the degree of consistency versus variability in signing speed.

While we had built models of these parameters based on human recordings [1], earlier research suggested that users prefer ASL *animations* to be slower than human *videos* of ASL [6, 7]. Thus, we seek empirical judgments from native ASL signers for each of these parameters for ASL animation. The **contribution** of this paper is therefore empirical: We report specific values preferred by users, which future sign-language animation researchers may use to guide their work.

Prior Work

Prior research has investigated the timing and pausing parameters of both ASL and spoken English. For instance, Grosjean et al. conducted several studies on speed and timing of ASL based on observing video recordings of human ASL signers [3, 4, 5]. The five-parameter model of ASL speed and timing presented in the instruction reflects the model of ASL that formed the basis for these researchers' work. One of their studies established that some timing parameters, e.g. the location and duration of pauses, are related to sentences' syntactic structure [4]. Overall, they found that human signers pause at approximately 25% of the inter-word locations during an ASL passage. They also found that longer pauses take place at sentence boundaries (approximately 229 milliseconds), with shorter pauses at other phrase boundaries.

Stimuli in the Study

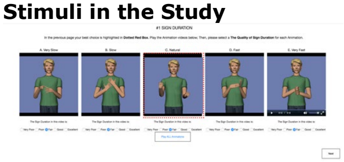

Figure 1: The interface for this study, which displayed five ASL animations of the same passage side-by-side, with each based on a different level of a particular timing parameter. Users indicated a scalar preference score (1 to 5) for each animation. A screen like this occurred 5 times during the study, once for each timing parameter under investigation.

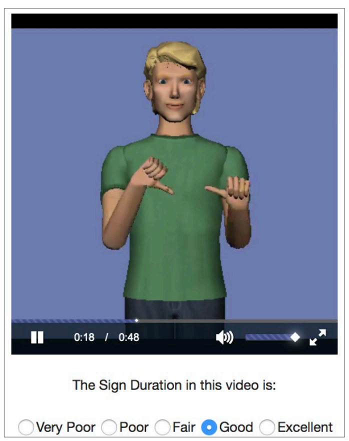

Figure 2: A detail image of a zoomed-in region of Figure 1, showing one of the five ASL animation stimuli on the screen.

Prior researchers have studied how to control an animated character to produce ASL with natural pauses and timing behavior. In early work, we built and evaluated some rule-based methods of predicting various ASL timing parameters [6, 7]. While not a specific question we had set out to investigate, our evaluation study suggested that users preferred ASL animations to be slower than videos of human signers. However, our study had not investigated users' preferences for each of the five timing parameters.

Since data-driven methods are now standard within the field of computational linguistics [10, 11], in later work, we collecting a video and motion-capture corpus of multi-sentence ASL passages [8]. Later, we used this corpus to investigate data-driven models ASL speed and timing trained on these human recordings [1]. In a small evaluation study, users preferred animations with timing values set using these models (as compared to our prior rule-based methods in [6, 7]), but there was a limitation: In [1], **we had assumed that users would prefer ASL animations with speed and timing parameters that were as close as possible to patterns learned from our human recordings of ASL.** In this paper, we investigate whether this assumption was correct, by asking users to evaluate ASL animations with a variety of parameter values. **No prior research study has collected subjective preference judgements about ASL animations of various timing parameter values.**

We had to select a range of values for each of the timing parameters in our study. To select the "midpoint" of our scales for each parameter, we considered the typical speed and timing parameters used in other ASL animation systems. For instance, Sign Smith Studio (SSS) [13] was commercial ASL animation authoring software that enabled a human to create animation. SSS included some "default" values for various speed and timing parameters, and it provided the user with the ability to manually customize many of these timing parameters. For example, SSS used 0.25 seconds as the default transition time between words and 0.5 seconds as the pause duration at a sentence boundary.

Method

For our IRB-approved study, a native-ASL-signer researcher on our team met in-person with native ASL signers on the Rochester Institute of Technology (RIT) campus to obtain their subjective preferences on ASL animations which varied according to each of the five timing parameters.

Participants. At the beginning of the appointment, which was conducted in ASL, the participants answered demographic questions. Our 16 participants included 8 women and 8 men, with median age 22.5 years old (range 18-25). Thirteen identified as Deaf, 1 as hard of hearing, and two were hearing children of Deaf adults who grew up using ASL since infancy. All participants learned ASL in childhood (age 2-8 years). Five had Deaf parents, 9 had Deaf family members or relatives. Fourteen of the participants used ASL at school as a young child. Three currently used only ASL at home, and 13 used both ASL and English at home. Fifteen used ASL at their college or university.

Procedures. On a 15-inch laptop, participants viewed a stimuli page. Figure 1 shows the general organization of the animations used in the page, while Figure 2 illustrates a zoomed image for one of the animations

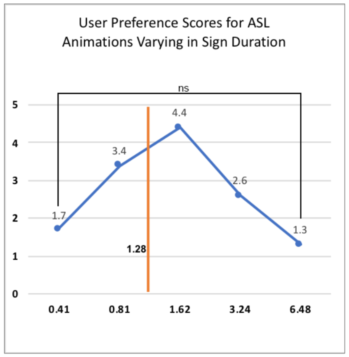

Figure 3: User preference scores for five animations which varied in their average Sign Duration (in seconds). All pairwise differences are significant except between the pair marked with "n.s." Vertical orange line indicates typical value in human signing.

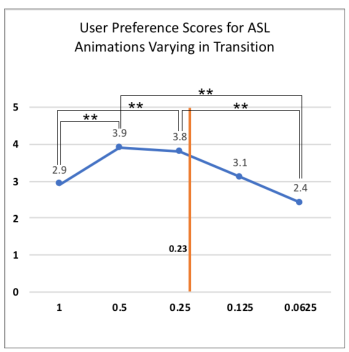

Figure 4: User preference scores for five animations which varied in their average Transition Time Between Signs (in seconds). Pairwise significant differences are marked with "**" (p<0.01).

used in Figure 1. Our stimuli consisted of time-parameter-adjusted versions of a set of ASL passages which we had used in prior work [1], with each passage approximately 75 words in length. Sign Smith Studio (SSS) [13] was used to generated ASL animations [6, 7], and a researcher adjusted the timing parameters in the animations to produce stimuli with particular timing properties. Each screen displayed five variations side-by-side of an ASL animation of the same message, with each version using a different value for a particular timing parameter. For instance, one screen displayed five ASL animations with different values for the "sign duration" parameter. A screen like this appeared a total of five times, with each screen showing animations that varied based on a different timing parameter, e.g. transition time, pause duration, etc.

Because the appearance variations of each animation can be somewhat subtle, some initial pilot testing with participants prior to launching the study revealed that users found it less confusing if the animations shown side-by-side were not displayed in randomized order: instead, pilot participants found it easier to compare the animations when they were presented in an arrangement with slower animations on the left and faster on the right. Pilot testing also revealed that participants preferred to know "what was different" among the five variations displayed on a screen. Thus, prior to each screen of the study, users were informed of how the animations would vary, e.g. "On the next screen, you will see 5 animations in which the amount of time in-between words is different. Please evaluate each animation to indicate how you rate its quality."

For each of the timing parameters (sign duration, transition, pausing frequency, pausing length, and differential signing rate), after the participant provided their quantitative scores for the five side-by-side animations, the process was repeated for the next timing parameter. Another stimuli page was shown with five variations of an animation to collect judgements for the next parameter. Thus, each participant saw 25 videos during the study, providing numerical scores for each on a 1-to-5 scale (1 very poor to 5 excellent).

Figures (3-7) show the average subjective score rating of participants for animations for each parameter. The Y axis in these figures represents the (1-5) user rating, while the X axis represents the five values of the timing parameter that had been displayed side-by-side.

In each graph (Figures 3-7), the vertical orange line serves as informal visualization of typical values for *human ASL signers* as reported in the linguistics literature. As the X axis is not plotted in a proportional manner (for space considerations), the vertical orange line has been plotted by linearly interpolating between the two adjacent levels (on the left and on the right).

To determine whether there were significant differences in users' subjective scores, we analyzed responses using a Kruskal-Wallis test for each of the parameters; we found a significant main effect for each parameter. We then performed post-hoc pairwise comparisons using a Wilcoxon test (with Bonferroni correction for multiple comparisons), to compare responses for each pair of levels for each parameter. Results of these tests are indicated as follows: In Figure 4, the "**" indicates statistical significance two levels (p<0.01); no other pairs in that figure had significant pairwise differences. In figures 3, 5, 6, and 7, since nearly all pairs were

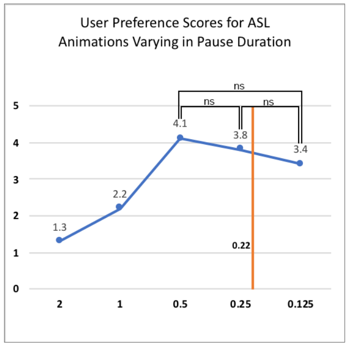

Figure 5: User preference scores for five animations which varied in their average Pause Duration (in seconds). All pairwise differences are significant except between pairs marked with "n.s." Vertical orange line indicates typical value in human signing.

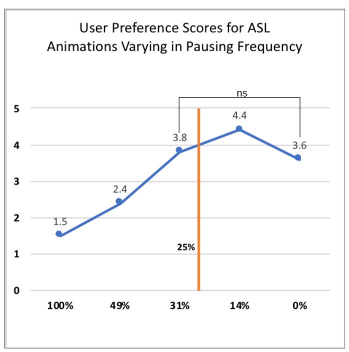

Figure 6: User preference scores for five animations which varied in their Pausing Frequency (represented as the percentage of inter-sign gaps where a pause occurs). All pairwise differences significant except one marked.

significantly different, it was simpler to mark only those pairs which did were not significant ("n.s.").

For each parameter, the values shown in the study were selected as follows: First a "neutral/default" value was selected for each value (which was used as the middle option in each graph). Next, two slower and two faster variations were selected, so that five levels were investigated for each timing parameter. More details about each parameter and its values are below.

Results

We compared five levels of **sign duration** (Figure 3): The middle level was 1.62 seconds, which was the average duration of signs in the ASL animations generated using SSS [13], using default timing of signs from the system's dictionary. We then quartered, halved, doubled, and quadrupled this value, to produce animation stimuli with average sign durations ranging from 0.41 seconds to 6.48 seconds. Larger sign duration values yield slower animations. Participants preferred the default 1.62 seconds/sign average timing, which was slightly slower than the typical value of 1.28 for this parameter in human signing. (This 1.28 value was calculated by identifying a set of 30 signs used in our stimuli, finding examples of these signs in our corpus of videos of human signing [8], and then calculating the average duration in seconds of those words in the corpus.

We compared five levels of **transition time** (Figure 4): The middle value was 0.25 seconds (which was the default transition time between words in SSS), and the other stimuli used values of 0.125, 0.25, 1.0, and 2.0. Participants preferred the 0.25- or 0.5-second stimuli (no significant difference between these two). Human ASL signers typically have a transition time of 0.23 seconds. This 0.23-second value was calculated by computing the average transition time in a set of ASL video recordings for the three human signers in our previously collected corpus of ASL recordings [8].

We compared five levels of **pause duration** (Figure 5): The middle value was 0.5 seconds (the default time of a pause between sentences in SSS), and the other levels included: 0.125, 0.25, 1.0, and 2.0. Users preferred animations with pause duration of 0.5, 0.25, or 0.125 (no significant difference between these three levels), and human ASL signers typically have a pause duration of 229 milliseconds, as reported in [4].

The **pausing frequency** parameter (Figure 6) is how often the signer pauses (e.g. stops moving between phrases or sentences), represented numerically as a percentage of the between-sign locations where a pause occurs. However, inserting pauses randomly among these locations would have yielded an unnatural result; so, we used the following rubric to define our five levels for this parameter in the stimuli: **0%** (no pauses inserted), **14%** (pauses inserted after every sentence), **31%** (same as previous, plus pauses inserted after every clause and verb phrase), **49%** (same as previous, plus pauses after every noun phrase), and **100%** (pauses inserted after every word). Users preferred animations with a pause after every sentence, which, in the stories shown in our study, was at 14% of the inter-sign locations. Humans typically pause at 25% of inter-sign locations [4].

The **differential rate** parameter (Figure 7) represents the tempo dynamics of an ASL signing passage, in which signers vary their speed throughout a passage,

Equation Explaining the Differential Rate

$$v_{final} = v_{original} \cdot Factor^{Exponent}$$

Equation 1: The final velocity (v_{final}) of a sign is based on its original velocity ($v_{original}$) multiplied by a speed adjustment factor (Factor), which may be raised to some power (Exponent). In Figure 7, the values shown along the x-axis represent this Exponent.

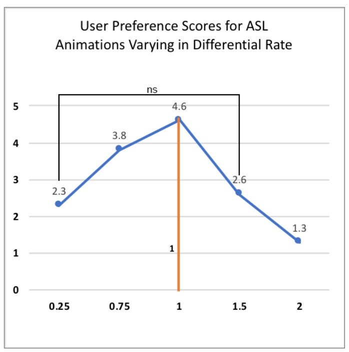

Figure 7: User preference scores for five animations which varied in the exponent used when applying their Differential Rate factor (see Equation 1). All pairwise differences significant except pair marked with "n.s." The vertical orange line indicates typical value for differential rate in human signing, i.e. with "1" indicating the speed adjustment factor applied as learned by the model trained on human data.

e.g. slowing down at the end of sentences. In prior work [1], we trained a Gradient Boosting Regressor model on motion-capture recordings of human ASL signing [8]. Our resulting model can predict, for each word in an ASL passage, a "speed adjustment" factor that should be used to adjust the speed of an individual ASL sign, based on some surrounding linguistic properties around that word, e.g. how close it is to the end of a sentence. Thus, the value of the differential rate parameter in this study represents the power ("Exponent" in Equation 1 in sidebar) to which we raise the speed adjustment factor for this word (as predicted by our model). Thus, in our prior work when we developed this model [1], we used a value of 1 for this Exponent; essentially, we had directly used the output of the model as has been trained on the tempo dynamics of human ASL signers in our corpus. By adjusting this Exponent, we can dampen or magnify the effect of this speed adjustment factor, to produce animations that are more consistent or more variable in their signing speed. In our stimuli, we presented users with animations that used an Exponent value of 0.25, 0.75, 1, 1.5, or 2. We found that participants preferred animations that had differential rate similar to human signers, i.e. with an Exponent of 1 for the speed adjustment factor, which was based on our model trained on human recordings of ASL.

Discussion, Limitations, and Future Work

Our study has individually investigated the preferences of native ASL signers for the values of five key timing parameters of ASL animations. Models based on these five parameters had formed the basis of prior research in linguistics, e.g. [3, 4, 5], and sign-language animation, e.g. [1, 6, 7]. While some preliminary findings in prior work had suggested that users may prefer slower overall speed for animations of ASL, as compared to the speed of human ASL videos, no prior study had systematically investigated users' preferences for these five parameters individually. Specifically, in a comparison of five levels for each parameter, our study has empirically identified preferred values for sign duration, pausing frequency, and differential rate. In the case of transition time and pause duration, our study revealed some significant differences between pairs of values, but it did not identify a single preferred value for those two parameters, among our five stimuli.

In Figures 3-7, we have included some vertical lines indicating of typical values for each timing parameter for human ASL signing. In general, the preferred values among our participants were close to human values for each parameter, but a **limitation** of our study is that, for each parameter, we had not included an animation stimulus with a value identical to this human level – which would have enabled a statistical comparison. In **future work**, we will replicate this study with one of our levels set at this human level. Furthermore, now that we know the general range of values which users preferred, we can select levels for our parameters that "zoom in" on the most-preferred range from this study. Finally, another limitation of our study is that we have investigated levels of each parameter individually (using default values for the other parameters while we investigate users' preferences for each); a future study would be needed to investigate interaction effects between these parameters.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. 1462280.

References

[1] Sedeeq Al-khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2018. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. *The 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. https://doi.org/10.1145/3234695.3236356

[2] Ursula Bellugi and Fischer Susan. 1972. A Comparison of Sign Language and Spoken Language. *Cognition* 1, 173-200. https://doi.org/10.1016/0010-0277(72)90018-2

[3] François Grosjean. 1977. The Perception of Rate in Spoken and Sign Languages. *Perception & Psychophysics* 22, 4, 408-413. https://doi.org/10.3758/BF03199708

[4] Francois Grosjean, Lysiane Grosjean, and Harlan Lane. 1979. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology* 11, 58-81. https://doi.org/10.1016/0010-0285(79)90004-5

[5] François Grosjean, and Harlan Lane. 1977. Pauses and syntax in American sign language. *Cognition* 5, 2, 101-117. https://doi.org/10.1016/0010-0277(77)90006-3

[6] Matt Huenerfauth. 2009. A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. *ACM Trans. Access. Comput.* 2, 2, Article 9 (June 2009), 31 pages. http://dx.doi.org/10.1145/1530064.1530067

[7] Matt Huenerfauth. 2008. Evaluation of a psycholinguistically motivated timing model for animations of American Sign Language. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08).* ACM, NY, NY, USA, 129-136. http://dx.doi.org/10.1145/1414471.1414496

[8] Pengfei Lu, Matt Huenerfauth. 2014. Collecting and Evaluating the CUNY ASL Corpus for Research on American Sign Language Animation. *Computer Speech & Language*, 28, 3, 812–831. DOI: https://doi.org/10.1016/j.csl.2013.10.004

[9] Ross Mitchell, Travas Young, Bellamie Bachleda, and Michael Karchmer. 2006. How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies* 6, 3 (Jan 2006). https://doi.org/10.1093/deafed/enj004

[10] Alok Parlikar and Alan W. Black. 2012. Data-driven phrasing for speech synthesis in low-resource languages. in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4013–4016. https://doi.org/10.1109/ICASSP.2012.6288798

[11] Jérémie Segouat and Annelies Braffort. 2009. Toward the study of sign language coarticulation: methodology proposal. In Proc. of the Second Inter. Conf. on Advances in Comput.-Hum. Interac., 2009. 369-374. https://doi.org/10.1109/ACHI.2009.25

[12] Carol Bloomquist Traxler. 2000. The Stanford Achievement Test, Ninth Edition: National norming and performance standards for deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education* 5, 4, 337-348. http://dx.doi.org/10.1093/deafed/5.4.337

[13] Vcom3D. 2019. Sign Smith Studio. Retrieved December 19, 2019 from http://www.vcom3d.com