The Analysis of Motion with the Goal of Developing a Model for Adverbs in American Sign Language

American Sign Language (ASL) is the primary language for the Deaf community, but due to a lack of a word-for-word translation between English and ASL, studies have suggested that the average literacy rate of the Deaf community is at a fourth grade level. The use of interpreters is considered to be the gold standard for translation, but this is something that has to be scheduled in advance for a minimum of two hours per session. This means that both written and spoken English are not easily accessible to members of the Deaf community and there are many scenarios, short unplanned interactions, where there will not be access to a translator. Even with technology advances in cochlear implants, studies have shown that the degree of speech perception and development of language skills is subject to the neural plasticity of the deaf individual, therefore this is not an option for everyone. In an effort to create an automatic English to ASL translator, we are developing a system to synthesize animations of ASL, with this research focusing on the synthesis of adverbial modifiers. The animation system will serve a target for an automatic translation process. The reason for using automatically–generated animation rather than pre-recorded video is that language is productive and it is not possible to have pre-recorded video for all possible utterances. A three-dimensional (3D) animation system can generate graphic animations in real time by converting linguistic data comprising ASL signs into time-based geometric data. It is important that the sequence of signs look smooth and natural for an ASL sentence, so as not to interfere with the information being translated. Also, due to the visual/gestural modality of ASL there is not always a word-for-word correspondence, especially in the case of adverbs where there are few lexical items that function for words such as “quickly” or “slowly”. These are considered to be expressed in the “nature of the motion”. At present, no work has addressed the problem of formally characterizing the “nature of the motion” in a way that makes it possible to implement adverbs as part of an automatic English-to-ASL translation system. For this, a mathematical model that incorporates both speed and affect will be developed based on observations of joint motion during signing. Several parameters will be calculated such as minimum, maximum, and overall speed, velocity, and acceleration, as well as displacement of joints and angular rotations. This model will portray a range of adjustable adverbial modifiers based on data gathered from motion studies. Incorporating a mathematically derived adverbial model to the current translation system will allow adverbs to be automated with incremental adjustments to speed and rotation accounting for increases in intensity. A survey will be conducted to determine if the model conveys adverbs convincingly by having participants judge the perceived intensity of inflection. Participants from the Deaf community will be shown animations incorporating the model and asked to rate it based on several parameters, including naturalness, clarity, grammar, and acceptability.

Section 1: Introduction and Overview

In North America, American Sign Language (ASL) is the primary language for the Deaf community and ranks third as the most commonly used language [Stern96]. ASL and English share some vocabulary, but it is a common misconception that ASL is a gestured version of English. ASL is linguistically different and there is not a one-to-one translation between the two, with them varying grammatically [Newell83, Valli93]. For most within the Deaf community, English is a second language where the average reading level is at a fourth grade ability [Holt94].

Currently the most widely accepted form of translation is to use interpreters. There are several drawbacks to this: advanced scheduling, scheduling minimum of two hours, cost, and invasion of privacy. In daily life there are several occurrences that need a means of communication where there will be no access to a translator. With technology advances in cochlear implants, the assumption could be made that the problem has been solved. For these individuals, there is the need to develop the language. This and their degree of speech perception are subject to the neural plasticity of the individual [Peterson10]. Solutions to this communication gap are still needed. Therefore in an effort to create an automatic English to ASL translator, we are developing a system to synthesize animations of ASL.

Research in developing avatars have progressed for the use several areas including automatic translation in English-to-ASL [Elliott00] and Deaf education [Efthimiou07]. ASL is a visual/gestural language that uses handshape, body movement, and facial expression. Due to this modality a word-for-word correspondence is not always an option, as in the case of most adverbs. Within ASL, adverbs are expressed as signs, facial expressions, or modifications to the sign itself. In the latter ASL uses “inflection” to modify the meaning of signs such as “very quickly” or “very slowly” in the place of lexical items. These adverbs are signed in an exaggerated fashion with changes in speed or affect. Currently no work is addressing the implementation of automating adverbs as a modifier of signs in an automated system, though similar work has been done to synthesize the motion during a role shift [Wolfe15].

There are two aspects that can be used to modify a sign to add inflection, speed and affect. Therefore, the following research questions are asked:

- Is it possible to find relative parameters to express speed and affect of adverbs in ASL?
- Can these parameters be characterized in a way that can be implemented into an automated animation system?

The following hypothesis will be investigated:

The intensity of inflection of signed adverbs can be integrated into a sparse key-frame synthesized animation system because incrementally adjusting the speed of the motion or the angle of rotation for joints can alter the interpretation.

To examine the applicability of automating adverbs in ASL, a mathematical model based on analyzed motion data is investigated. To develop a model, we consider the question “what is needed for adding inflection to adverbial signs?” Distinct differences that relate to specific adverbs need to be found. Therefore we will analyze recording of signed sentences with the goals of:

- Determining the differences in affect of movement as well as velocity for selected adverbs, by measuring minimum, maximum, and overall speed of the motion, velocity, acceleration, displacement of joints for all three dimensions, and minimum and maximum distance traveled for each joint.
- Determining the commonalities within an adverb that can lead toward a way to generalize the synthesized motion as much as possible by comparing the consistent or incremental changes to the measures from the previous statement.
- Determining if modifying signs and posture will convey the adverb well enough for it to be perceived by adjusting the joint rotation and speed of motion based on the previously mentioned measures.

Section 2: Related Work

When considering the study of motion, the perceptual significance of body motion when displaying emotion has been studying in many different contexts. There is a noticeable difference between happiness and sadness [Coulson04]. Work has also shown the characteristics between emotion and posture and that emotion can be deduced from velocities, accelerations, and jerk [Wallbott98]. Motion capture data focusing on the posture and dynamics of the body have been applied to a character to study the perception of emotion with moderately successful results [Normoyle13].

There is also how editing motion in an animation changes its naturalness. Viewers are less tolerant to the automated decreases in velocity from time warping [Vicovaro12] and viewers are more sensitive to time warping where slow motions are made faster [Reitsma03].

Several methods to incorporate these perceptions of emotion into animation exist. One approach adds parameter adjustments directly to the joints rotations [Amaya96]. They developed a method to produce emotional animation from neutral states using signal processing. Transformations of speed and spatial amplitudes were developed. This study shows that it possible to convey emotions at different intensity through, but the method allows for the incorporation of the whole body to communicate the intensity of the emotion. The recordings used were examples of communicating emotion not in a sign language context. Our problem focuses on conveying emotion and intensity in the context of adverbial use in a sign. This will incorporate movement from primarily the upper body, where the intensity is conveyed with changes in speed and affect of the sign itself.

For sign synthesis, there are two approaches: the use of libraries of motion captured signs [Awad10], or sparse key-frame libraries [Delorme09]. These libraries are procedurally combined to create signed sentences. Synthesis using motion capture produces animations that are very natural in motion. This approach is able to reproduce the subtle details to create smooth, natural signs. Though visually this method realistically captures the motion, it is exceptionally difficult to modify the data. Current technology in processing and post-production of motion capture data is not advanced enough to allow it to be easily modified. There is a high amount of data being sampled to capture a movement during motion capture. This becomes impractical to modify because every angle at every frame needs to be reconsidered to maintain the subtleties of the motion. This restricts the data to be used for only the sign that was captured, without the ability to extend or modify it to add inflection.

Linguistic rules are more easily applied to modify animations that use sparse key-frame libraries. Sparse key-frame libraries are based on geometric and timing data for specific beginning and end poses, where the transition is interpolated between the two poses. Though this method easily correlates to the linguistic structure, it has an unfortunate lack of realism in the motion, lacking the subtle movements that are not being represented in the two poses, but should be implemented during the transition. No information is currently given through the linguistic structure to address inflection to modify signs. Combining data gathered from recordings with a sparse key-frame library would be the optimal system, with the best from both approaches

Some work has been done to develop models of generalized motion for an avatar in the case of role shifts [Wolfe15]. In ASL this is where a signer uses a body turn to assume a character role of protagonist. This work included similar motion studies as we propose, that tracks the motion of major joints during the role shift. Recordings were taken of role shifts at several degrees of rotation as well as speed of the rotation. This allowed for timing data to be gathered for different joints that could be applied to the avatar and show staggered timing that was more natural, less robotic, as well as adjust for the intensity of the rotation of the role shift.

Section 3: Research Design and Methodology

When considering 3D animation, signs are the combination of geometric poses and movement of an avatar. With adequate video recordings as reference, any sign can be animated. Viewing it in this sense does not incorporate the linguistic structure of ASL, because only timing and geometric data are given. For the synthesis of signing, the necessary combination is the result of co-occurring linguistic and extralinguistic processes applied to the animation [Wilbur00]. Another critical component of the development of an automated English-to-ASL translator is the flexibility of the systems ability to generate sentences [Sedgwick01]. This requires that the system be built with a set of parameters that can be easily adjusted given the context. With linguistic information, an animation system computes the avatar's motion based on a global orientation. Non-manuals are stacked on top of the linguistic track. Staggered timing is implemented by offsetting this initial global orientation by computing the transition of each joint in local coordinates. Parameters, that when combined allows for naturalness in movement and rotation, consists of start and end times of each joint rotation, acceleration and velocity of the motion, and angle of joint rotation for each joint in Cartesian coordinates.

To develop the parameters needed to automate adverbs we will first need record video references for analysis. We will focus on five different adverbs: slowly, quickly, happily, sadly, and angrily. For recording the different adverbs being used, we will have the aid of an experienced Deaf stage actor. During the recording the actor will wear strategically placed colored markings to be picked up by the motion tracking software. We will be using After Effects to track the markers. This will give us data on timing and displacement. The actor will be recorded from multiple angles. They will perform five versions of each of five adverbs at different intensities. For these five adverbs we will track the motion and rotation of joints of the actor during the sign. This will be used to determine the rotational speeds and angles for the joints.

The upper body will be the focus of tracking, specifically the root, spine, neck, shoulders, elbows, and wrists. Data from these points will be analyzed to determine commonalities, if any, in the different adverbs as well as the differences. Several parameters will be calculated for each joint: minimum and maximum speed, overall speed, velocity, acceleration, displacement for all three dimensions, and minimum and maximum distance traveled. This will give us range of motion and timing information that can be applied to the avatar per intensity needed. These developed parameters for each adverb, adjustments to start and end times for each joint, as well as acceleration, velocity, and angle of rotation, will be layered onto the geometric data of the sign at a normal condition. This will be connected to the animator interface where adjustments can be easily selected. This will modify the angle of rotation of joints as well as speed to account for the intensity of the adverb in use. The data collected from this will be generalized in a way that they can be interpolated across intensities to be applied to the avatar, giving speed and affect to the signs in a smooth and natural way.

Evaluation of our model will consist of a two part series of surveys to determine if the model produces readily understood animations and convey the adverbs. To conduct the survey we will be using SignQUOTE (Signed Qualitative Usability Online Testing Environment) [Schnepp11]. We will need to hire an interpreter to record the informed consent as well as the test instructions. All information included in the survey will be presented in ASL. Surveys will be completely confidential, with only emails being recorded for use of an e-gift card.

The survey will consist of ~50 participants. The participants will vary in experience of ASL fluency, from novice to expert signers. Participants will be asked to identify their skill level, how many years they have been signing, as well as if they are Deaf. Participants will be sent a link to the survey and will be allowed to view the animated clips multiple times. These animations will be presented in a random order for each participant. These animated sentences will vary in degrees of speed or affect and participants will be asked to specify what they perceive the adverb to be. After each clip, the participant will be given a list of adverbs to choose from. We have selected this approach because we want the results to be comparable to the previous studies done in this area. Though this has the potential to restrict the responses of participants and may help with recognition [Frank01].

After the first round of viewing the animations and selecting the perceived adverb, a second round of clips will be shown that focuses on the intensity of the modification. For instance, to compare the adverbs “sad” or “happy”, shown in Figure 1, or “slowly” and “quickly”, shown in Figure 2, as series of animations will be shown and the participant will be asked to rate it. They will also be asked questions on the naturalness, clarity, grammar, and acceptability of the synthesized adverbial inflections. This will be given as a closed-ended question using a Likert scale. There will be no text labels associated with the survey.

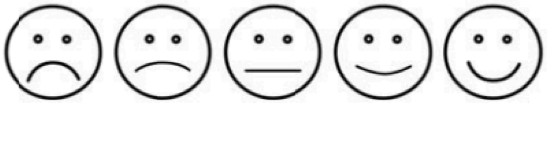

Figure 1 Sad to Happy Likert Scale

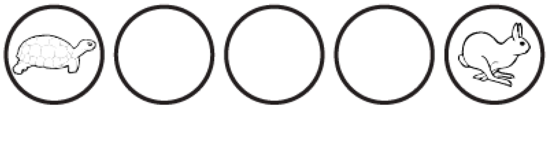

Figure 2 Slow to Fast Likert Scale

The collected data from the first round will be analyzed by percentage of correct responses per clip, as well as a confusion matrix that will show displayed adverbs versus selected adverbs. The data from the second round of viewing will be analyzed to confirm expected differences in amplitude of movement and inflection using repeated measure ANOVA. Our hypothesis will be evaluated on results from the first part of the survey showing whether the adverb is being correctly identified, as well as the second part, showing that incremental changes to the speed or angle of rotation will convey more intense uses of the adverbs.

Section 4: Plan of Work and Outcomes

Plan of Work

The major phases of the project include:

- Record ASL Deaf theater actors using the selected adverbs in sentences at different levels of speed and affect
- Apply motion tracking to recordings
- Analyze differences in speed and affect to determine specific parameters that can be applied per adverb
- Development of a model based on analyzed data comprised on noticeable incremental differences in speed or angle of rotation as intensity increases.
- Integration of the model into the current transcriber interface
- Generate series of animated sentences using the parameter settings
- Testing/survey
- Analyze survey results
- Reverse / repeat
- Documentation

Deliverables:

The target end results of this research are:

- A mathematical model that can be incorporated into a 3D avatar system
- A parameter file containing specific adjustments based on the adverb being applied
- An adjustment to the current transcriber interface to adjust levels of speed and affect

Section 5. Conclusions and Future Work

The implementation of an automated adverbial modifier would allow for a more capable English-to-ASL translator. This will facilitate a more natural animation system that will incorporate the inflection needed to convey adverbs. Moving forward we plan to extend this technique to other adverbs not covered in the initial recordings.

This proposal focuses on if “inflecting” the sign is an adequate representation of adverbs in ASL. Future work will introduce the use of facial expression and body posture as “nonmanual markers” to enhance the readability of the sign. Movements of the face help to modify the mood of the signs. Facial expressions are subtler than the inflection used to modify the sign and will require a more in depth study to synch these expressions to their corresponding adverb.

Section 6: References

[Amaya96] Amaya, K., Bruderlin, A., Calvert, T. 1996. Emotion from motion. In Graphics Interface, Canadian Information Processing Society, Toronto, Ont., Canada, Canada, GI ’96, 222–229.

[Awad10] Awad, C., Courty, N., Duarte, K., Le Naour, T., & Gibet, S. (2010). A combined semantic and motion capture database for real-time sign language synthesis. *Intelligent Virtual Agents,*, 432-438.

[Delorme09] Delorme, M., Filhol, M., & Braffort, A. (2009). Animation generation process for Sign Language synthesis. *International Conference on Advances in Computer-Human Interaction (ACHI '09)* (pp. 386-390). Cancun, Mexico: IEEE.

[Efthimiou07] Efthimiou, E., & Fotinea, S.-E. (2007). An envrionment for deaf accessibility to education content. *International Conference on ICT & Accessibility*, (pp. GSRT, M3. 3, id 35). Hammamet, Tunisia.

[Elliott00] Elliott, R., Glauert, J. R., Kennaway, J. R., & Marshall, I. (2000). The development of language processing support for the ViSiCAST project. *Proceedings of the fourth international ACM conference on Assistive technologies (ASSETS 2000)* (pp. 101-108). Arlington, VA: ACM.

[Frank01] Frank, M. G., Stennett, J. 2001. The forced-choice paradigm and the perception of facial expressions of emotion. Journal of Personality and Social Psychology 80(1), 75–85.

[Holt94] Holt, J., "Stanford Achievement Test - 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results". http://www.gallaudet.edu/~cadsweb/sat-read.html Accessed April 5, 2014.

[Newell83] Newell, W., & National Technical Institute for the Deaf. (1983). Basic Sign Communication. Silver Spring, Md: National Association of the Deaf.

[Normoyle13] Normoyle, Aline, et al. "The effect of posture and dynamics on the perception of emotion." *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2013.

[Peterson10] Peterson, N., Pisoni, D., Miyamoto, R., “Cochlear implants and spoken language processing abilities: Review and assessment of the literature" http://iospress.metapress.com/content/X66L17L8118MM4N3

[Reitsma03] Reitsma, P. S. A., Pollard, N. S. 2003. Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. In ACM Transactions on Graphics, vol. 22, 537–542.

[Sedgwick01] Sedgwick, Eric, et al. "Toward the effective animation of American Sign Language." (2001).

[Schnepp11] Schnepp, Jerry, et al. "SignQUOTE: A remote testing facility for eliciting signed qualitative feedback." *Proceedings of the 2nd International Workshop on Sign Language Translation and Avatar Technology*. 2011.

[Stern96] Sternberg, M.: The American Sign Language Dictionary, Multicom (CD-ROM), 1996.

[Valli95] Valli, Clayton, Linguistics of American Sign Language: An Introduction. Gallaudet Univ Pr, 1995.

[Vicovaro12] Vicovaro, M., Hoyet, L., Burigana, L., O'Sullivan, C. 2012. Evaluating the plausibility of edited throwing animations. In SCA, 175–182.

[Wallbott98] WALLBOTT, H. G. 1998. Bodily expression of emotion. European Journal of Social Psychology 28, 6 (December), 879–896.

[Wilbur00] Wilbur R.B. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In Lane, H. & K. Emmorey (eds.), The signs of language revisited: Festschrift for Ursula Bellugi and Edward Klima, (pp. 213-241) Hillsdale, NJ: Lawrence Erlbaum.

[Wolfe15] Wolfe, R., McDonald, J. C., Moncrief, R., Baowidan, S., Stumbo, M., "Inferring biomechanical kinematics from linguistic data: a case study for role shift", Sign Language Translation and Avatar Technology Converence, 2015. https://perso.limsi.fr/filhol/misc/SLTAT2015-programme.pdf

Appendix: Budget and Budget Narrative

Budget Item	Amount
Survey Participant Incentive ($10.00 x 30 people)	**$300.00**
Deaf Theater Actor ($300.00 x 3 hours)	**$900.00**
Translator ($100.00 x 2 hours)	**$200.00**

The requested budget for this proposal is $1400.00. This will cover the incentive of $10.00 gift cards for those who participate in the survey, the recording of a Deaf theater actor for motion analysis at $300.00 an hour for three hours, and recording an interpreter for the informed consent and test instructions of the survey at $100.00 an hour for two hours.