Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages, pages 151–156
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Extending a Model for Animating Adverbs of Manner in American Sign Language

Robyn Moncrief
School of Computing, DePaul University
Chicago, IL
rkelley5@mail.depaul.edu

Abstract
The goal of this work is to show that a model produced to characterize adverbs of manner can be applied to a variety of neutral animated signs to be used towards avatar sign language synthesis. This case study presents the extension of a new approach that was first presented at SLTAT 2019 in Hamburg for modeling language processes that manifest themselves as modifications to the manual channel. This work discusses additions to the model to be effective for one-handed and two-handed signs, repeating and non-repeating signs, and signs with contact.

Keywords: avatar technology, sign synthesis, adverbs in ASL

1. Introduction

A signing avatar is a necessary component of any sign translation system. It needs to produce signed utterances that are grammatically correct and easy to read. Although an avatar's appearance is important to its legibility, its motion has even greater impact (Malala et al., 2018). *Sign synthesis* is the computation of an avatar's movement. It is a combination of 1) path computation, which is related to motion planning (Barraquand & Latombe, 1991), 2) timing along that path, 3) the determination of joint participation in creating the path (McDonald, 2005; McDonald et al., 2016) and 4) ancillary motions required to support the clarity of the utterance (Schnepp et al, 2012). This case study presents the extension of a new approach to modeling sign language processes that manifest themselves as modifications to the manual channel.

2. Adverbs of Manner

American Sign Language (ASL) is an independent natural language (Valli & Lucas, 2000). Automatic translation between spoken language to sign have lagged behind spoken-to-spoken translation, due in part to the fact that there is no one-to-one mapping from ASL to English.
A case in point is the inclusion of adverbs of manner that express how the action of the verb takes place.
In English, the actions, states, and sensations of a verb can be modified through the application of an adverb of manner (Valli & Lucas, 2000). The following are two examples of adverbs of manner used in English. Each example is made up of two sentences. The first without the adverb; the second with the adverb:

The boy ran.
The boy ran quickly.

The couple danced.
The couple danced beautifully.

3. Related Work

Previous research of the use of adverbs of manner in ASL was limited. In contrast to English, ASL does not necessarily add an independent lexical item to express an adverb of manner. Instead, adverbs of manner are considered to be non-manual (Bickford, 2006) and modify the verb. Adverbs of manner occur through changes to the "quality of motion" as well as nonmanual signals (Baker & Cokely, 1980, Valli & Lucas, 2000, Padden, 2016).
Therefore, the starting point for related work is grounded in analysis of gesture motion. Two important examples of previous work in this area are the EMOTE (Expressive MOTion Engine) model (Chi et al, 2000) and GRETA (Hartmann, Mancini, and Pelachaud 2005). EMOTE stems from the Laban Movement Analysis (LMA) and the Effort-Shape model. The Effort-Shape model draws on LMA's classification of motion in the following way: Effort, qualitative descriptions of energy in motion and Shape, how the body changes forms during motion. EMOTE proposed a parameterization of Effort using qualitative descriptions of energy in motion through the following: 1) Space, 2) Weight, 3) Time, and 4) Flow.
GRETA's expressivity parameterization stems from psychology and expands on EMOTE's techniques for synthesis. It is comprised of six attributes: 1) Overall Activation, 2) Spatial Extent, 3) Temporal Extent, 4) Fluidity, 5) Power, and 6) Repetition.
Although researchers have examined the effects of affect on gesture (Kleinsmith & Bianchi-Berthouze, 2012), and Zhao et al. (2000) suggests that the EMOTE system would be useful in synthesizing sign languages, no one has reported on using such an approach for portraying adverbial modifiers in sign language.
Sign synthesis requires more specification than what is outlined in either EMOTE and GRETA. The characterizations proposed by EMOTE do not fully capture adverbial modifiers used in ASL. For example, EMOTE would characterize

<u>slowly</u>
WALK (1)

as *Bound: controlled or restrained*. As to be discussed in Section 8., data collected in this study demonstrates that the motion does not conform to this EMOTE descriptor. These systems do not take into account the importance of positionings of signs, as well as the animation techniques needed for building realism.
A more complete motion model is necessary to allow a 3D avatar to modify signs such as verbs, while supporting and respecting ASL's grammatical structure. Accuracy and

naturalness in the generated motion are necessary to make sentences as easy to understand as possible.

4. Procedure

Achieving an improved model required several steps, including the selection of adverbs, a motion study of the adverbs, animating and validating the exemplars, and data analysis (Moncrief, 2019).

4.1 Adverb Selection

Four commonly used adverbs of manner were chosen for this study. These adverbs represent different qualities of possible adverbial modifications to a sign, as well as having corresponding independent lexical items. The four adverbs contain two pairs of contrast: intensity and affect. For the contrast in intensity, the adverbs *quickly* and *slowly* were chosen. For the contrast in affect, the adverbs *happily* and *sadly*. All four adverbs were applied to the sign WALK, which is a two-handed noncontact sign with repeating motion that has few additional constraints to consider when applying motion modification.

4.2 Motion Study

Video recordings of a fluent signer are the basis for characterization of adverbial modifications and nonmanual signals. Based on the data from the video recordings, animations were generated. These animations then went through two revision cycles with a certified sign-language interpreter for grammaticality and naturalness.

4.3 Data Analysis

To create each animation, keyframe data was set by an animator. When the animations were generated, the in-betweens were interpolated. This interpolated motion data was then collected from the generated animations and used for analysis. This included joint positions for the wrists, elbows, and shoulders; timing; and joint velocities.

To determine the primary contributing variables for separating the adverbs of motion using the collected motion path data, a Linear Discriminant Analysis (Eisenbeis, 1972) was performed. This led to a high degree of separation of the adverbs, with the first linear discriminate accounting for over 98% of the separation of adverbs. The primary variables that accounted for the differences in the adverbs of manner included the wrist position and velocity. The separation based on the first two discriminant functions is shown in Figure 1.

Figures 2 and 3 show the distinct motions paths for the right wrist in the five animations on the transverse plane (x, y), as looking down at the signing space, and sagittal plane (y, z), as looking at the side of the signing space. The color variation through the motion path accounts for the change in velocity.

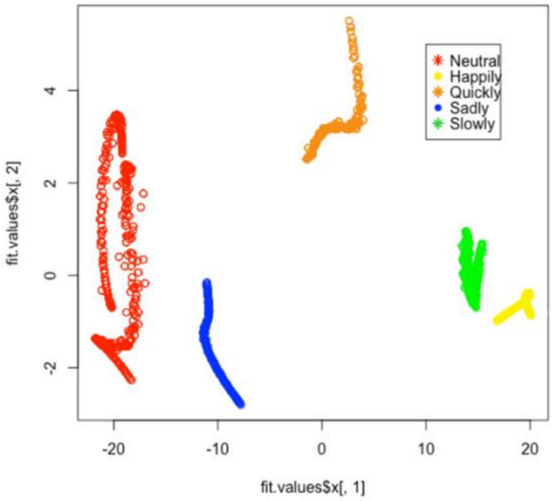


Figure 1: Separation of adverbs through the first two discriminate functions

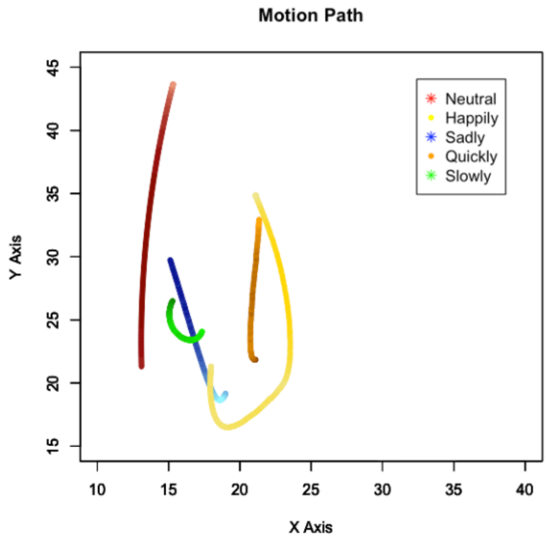


Figure 2: Motion path of wrist in the traverse plane

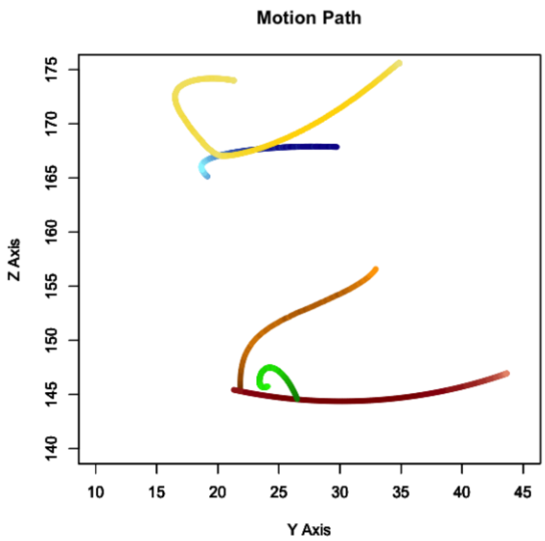


Figure 3: Motion path of wrist in the sagital plane

5. WALK Results

This model was then applied to the neutral animation of WALK for initial comparison and revision. Though wrist position and velocity were the significant contributors, those alone were not enough to convincingly convey the chosen adverbs. The model further incorporated changes to timing, ease-in and ease-out, joint positions, joint rotations, and adjustments to the spine. For the use of the *slowly* modifier, if only the wrist position and velocity were implemented, the sign would look Bound according to EMOTE parametric characterization, but based on the data and the visual observations, *slowly* required expressive joint rotations that could be described as Light.

The adjustments to the timing, joint rotations, and joint positions of the neutral animation consisted of developing a multiplier for each adverb. To account for the differences in speed along the motion path shown for each adverb, multipliers for timing were applied. The adjustments to timing would compress or lengthen the amount of time between keyframes. Arm joint positions were adjusted, changing the overall motion paths for the wrists. Rotational adjustments were made to the wrists to expand expressivity.

However, adjusting joint rotations and joint positions are not sufficient to convey the adverbial changes. Ease-in and ease-out were incorporated to increase the perceived naturalism (Thomas and Johnston 1995). Ease-in refers to the Slow In principle and ease-out refers to the Slow Out principle found in animation. A fast ease-in can model the sudden breaking on a car. A slow ease-out can simulate a gentle acceleration. Adding these allows for changes to the speed of the curve of an animation (Burtnyk & Wein, 1976). Ease-in gives a slow start to the transition and ease-out gives a slow end.

The model was extended to include the proximal joints of the spine to alleviate the need of requiring an unnatural exaggerated extension of the arms as the path of the wrist moved further from the. body. This aligns with the migration of motion between distal and proximal joints as described by Brentari (1998).

6. Validating the Model

To evaluate for generalizability, the model has been further tested on several other verbs. Testing the limitations of the model required a selection of a variety of verb signs. Selected signs included use of one or two hands, varying use of contact or noncontact, and repeating and nonrepeating motion. Selection is shown below in Table 1. Table 1 includes the sign used for model development, WALK, for comparison.

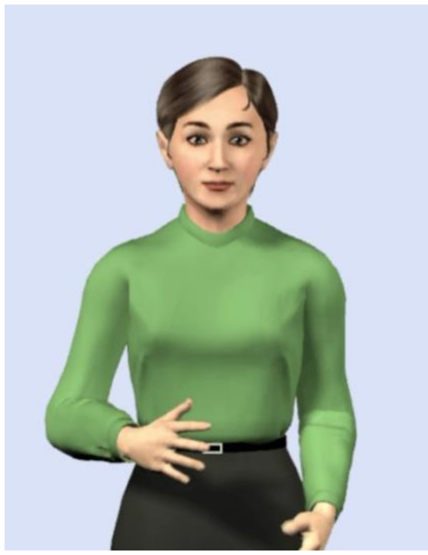
Sign	*Two- Handed*	*Contact*	*Repeating*
WALK	X		X
ASK			
GIVE		X	
THINK			X
BREATHE	X	X	X
INFORM	X		
CLEAN	X	X	
SEARCH	X		X

Table 1: Shows chosen signs and their characteristics for selection.

It was also important to evaluate signs that displayed a different motion path than that used for the model. A neutral version of WALK alternates the extension and retraction of the arm, led by the hand, in a flat motion path that does not vary up or down. The signs chosen for further consideration of the model were again chosen for their variety of motion paths in comparison to WALK.

7. Adverbs of Manner - Speed

The adverbial modification for *slowly* proved to be the easiest to transfer to the new signs, followed by *quickly*. The changes to this set of contrasting adverbs of manner relied less on a change in motion path to convey the modification and even less on nonmanual signals, though wrist rotations were a contributing part of the model for both. Whereas the model for *slowly* was perceived across all evaluated signs, *quickly* was perceived on signs that had repeating motion, such as BREATHE, and was harder to perceive on non-repeating signs such as ASK and GIVE. This is primarily due to the short duration for the neutral version of the nonrepeating signs. When the model for *quickly* was applied, the overall timing of the sign was not changed significantly and further increasing the speed only made the animation come across as less natural. To account for this, the model was further extended to the surrounding signs. In the case that the sign where the adverbial modifier for *quickly* was applied, the prior and following signs were also modified. Figure 4 below shows an overlay of the neutral animation for BREATHE and the applied *slowly* model. Figure 5 below shows the comparison with *quickly*.



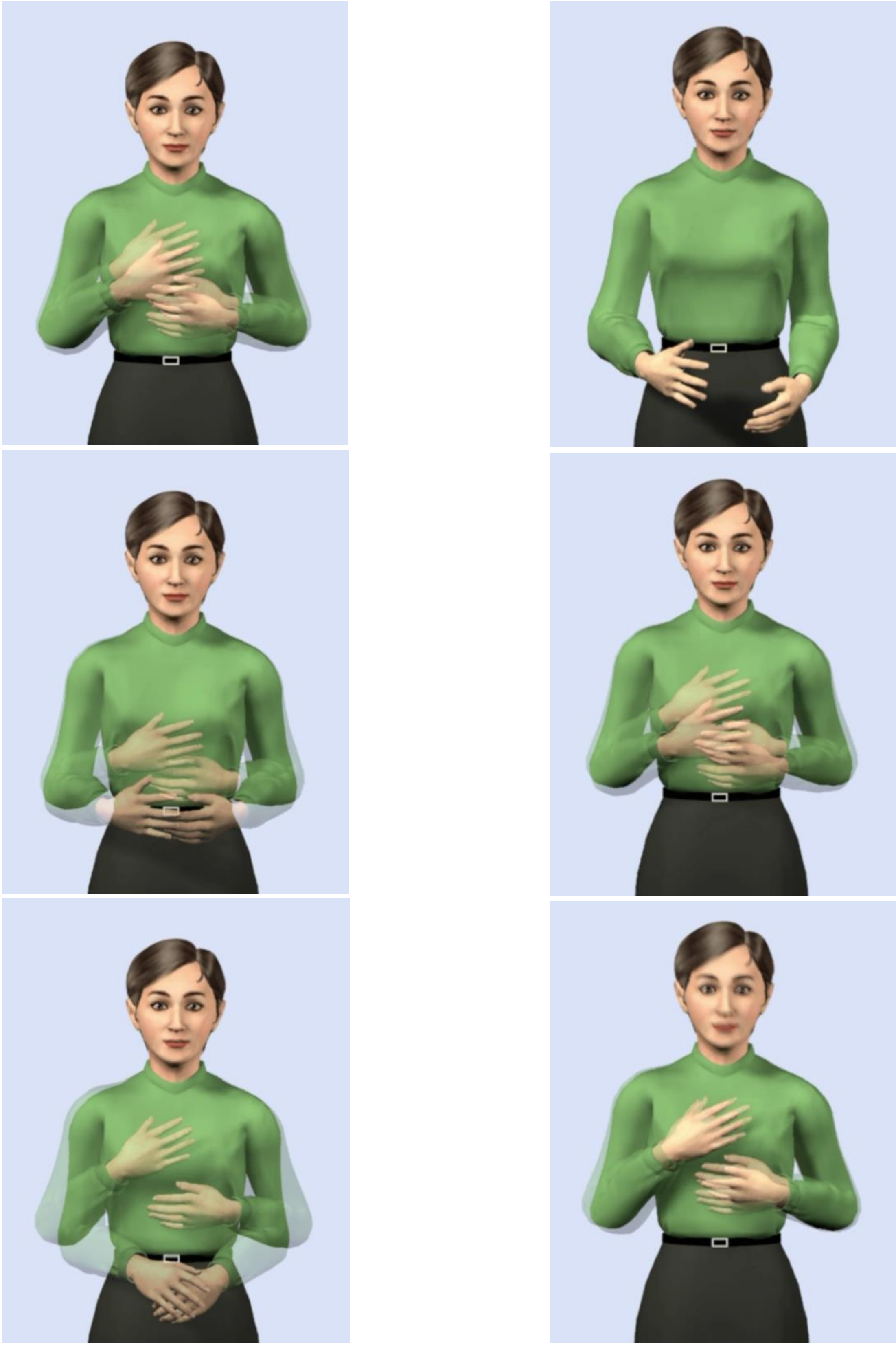


Figure 4: Shows an overlay with the neutral animation of BREATHE and BREATHE slowly.

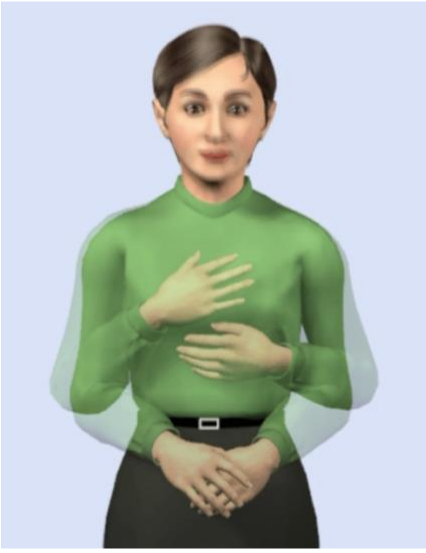


Figure 5: Shows an overlay with the neutral animation of BREATHE and BREATHE quickly.

8. Adverbs of Manner – Affect

The changes to this set of contrasting adverbs of manner relied heavily on a change in motion path to convey the modification and even more on nonmanual signals. For both *happily* and *sadly*, the timings were extended and ease-in and ease-out was used. The significant difference between the two came from the changes to the motion path. For *happily*, the model lifted and expanded the motion path. Figure 6 shows an overlay of the neutral GIVE and GIVE with the *happily* modification. For *sadly*, the model lowered and compressed the motion path. Based on the initial data collection, *sadly* showed to have a continuous lowering effect on the signing space in signs with repeating motion. This is demonstrated in Figure 7. To incorporate this into the model, keyframe data was compared in the neutral animation to see if the sign comes back near the starting position, any keys after would have an increased drop in their position and an even slower timing adjustment..

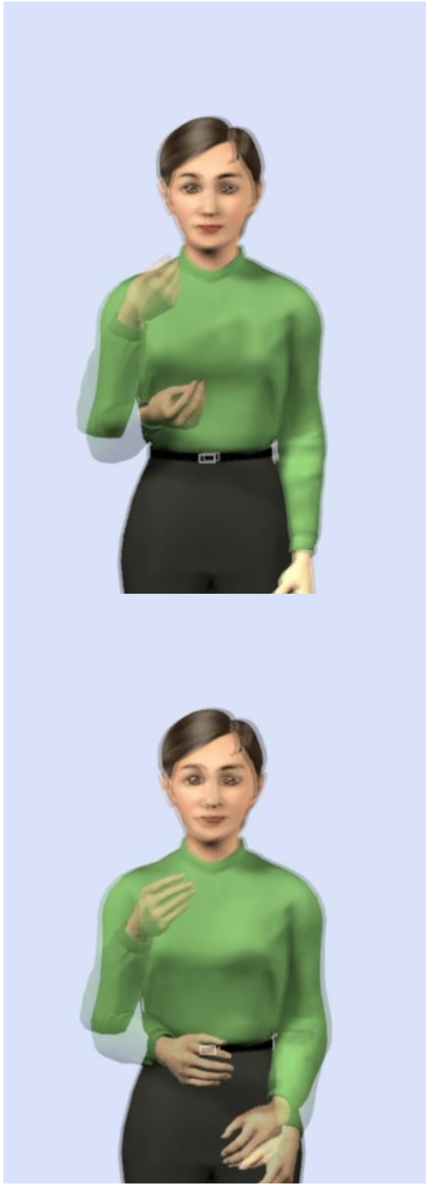


Figure 6: Shows an overlay with the neutral animation of GIVE and GIVE happily.

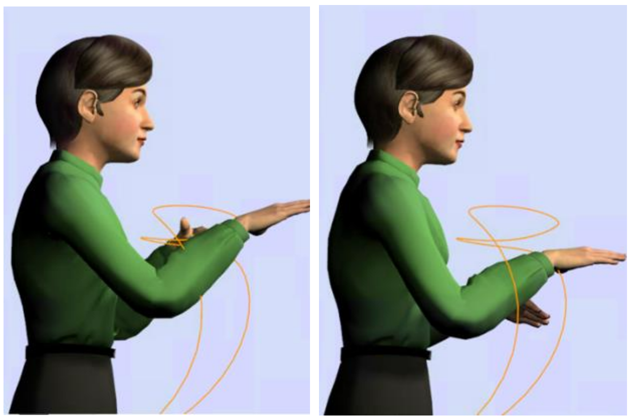


Figure 7: Shows the motion path the right wrist for WALK *slowly*. The first extension is shown higher, with the repition of the extension shown lower.

9.

9. One Handed vs Two Handed Signs

The model needed to recognize if a sign was one or two handed. Since the original model was built based on a two-handed sign, initially both arms were being modified when applied to one-handed signs. This would cause the left hand to move when there should not be movement. The model now recognizes if the sign is one-handed or two-handed based on comparison between keyframes throughout the sign. It then recognizes which hand to apply the modification to.

10. Single vs Repeating Motion Signs

When applying the model to single motion signs, there is less overall movement to aid in the perception of the adverbial modification. To help with the perception of the adverbial modifiers of intensity, surrounding signs can be adjusted. In the case of affect, the use of nonmanual signals (facial expressions) will play an important role.

11. Contact Signs

When confronted with signs that had some form of contact, GIVE, BREATHE, and CLEAN, the model did not initially take this contact into account. In several cases this resulted in the hands overshooting and ending in a collision with the body at the point of contact. To account for this, the neutral animations needed to be adjusted to include a tag on the keyframe with the contact. With this tag in place, the model would negate any positional/rotational motion path modifications applied to that keyframe. This would allow for the contact to occur as originally animated, with the surrounding keyframes being adjusted.

12. Conclusion and Future Work

To generalize the model for adverbial modifications, it did require extension for application to various signs to adjust for contact, expanding the model to the prior and trailing signs based on the adverbial modification for *quickly*, and further lowering of motion path for repeating motion based on the adverbial modification for *sadly*.

The next step in this work is to conduct a user study to test whether using multiple channels will increase the intensity of the perceived adverb. In other words, is the adverb *happily* portrayed by motion modification and nonmanual signal perceived as *more happy* than when the adverb is portrayed by motion modification alone.

13. Bibliographical References

Baker, C. & Cokely, D. (1980) American Sign Language: A teacher's resource text on Grammar and Culture. Silver Spring Md.: T.J. Publishers. Reprint, Washington, D.C.: Gallaudet University Press

Barraquand, J. & Latombe, J. (1991) Robot motion planning: A distributed representation approach. The International Journal of Robotics Research. 10.6. pp. 628-649.

Brentari, D. (1998) A prosodic model of sign language phonology. MIT Press.

Burtnyk, N. & Wein, M. (1976). Interactive Skeleton Techniques for Enhancing Motion Dynamics in Key Frame Animation. Communications of the Association for Computing Machinery, Vol 19. No. 10. pp. 564-569.

Chi, D., Costa, M., Zhao, Z. & Badler, N. (2000) The EMOTE model for effort and shape. Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing. pp. 173-182.

Eisenbeis, R.A. & Avery, R.B. (1972) Discriminant analysis and classification procedures: theory and applications. Lexington, MA: DC Heath.

Hartmann, B., Mancini, M. and Pelachaud. C. (2005) Implementing expressive gesture synthesis for embodied conversational agents. International Gesture Workshop. Berlin: Springer. pp. 188-199

Kleinsmith, A. & Bianchi-Berthouze, N. (2012) Affective body expression perception and recognition: A Survey. IEEE Transactions on Affective Computing 4, 1: pp.15-33.

Malala, V.D., Prigent, E., Braffort, A. & Berret, B. (2018) Which Picture? A Methoology for the Evaluation of Sign Language Animation Understandability.

McDonald, J., Wolfe, R., Alkoby, K., Carter, R., Davidson, M. J., Furst, J., Hinkle, D., Knoll, B., Lancaster, G., Smallwood, L., & Ougouag, N. (2005) Achieving consistency in an FK/IK interface for a seven degree of freedom kinematic chain. Proceedings of the 13th international conference in central Europe on computer graphics, vizualization and interactive digital media. pp. 171-179.

McDonald, J., Wolfe, R., Wilbur, R. B., Moncrief, R., Malaia, E., Fujimoto, S., Baowidan, S., & Stec, J. (2016). A new tool to facilitate prosodic analysis of motion capture data and a data-driven technique for the improvement of avatar motion. In Proceedings of Language Resources and Evaluation Conference (LREC). pp. 153-159.

Moncrief, R. (2019) A Model for Animating Adverbs of Manner in American Sign Language. Talk presented at the 6th Workshop on Sign Language Translation and Avatar Technology. Hamburg, Germany

Padden, C. (2016) Interaction of morphology and syntax in American Sign Language. Routledge.

Schnepp, J.C., Wolfe, R.J., McDonald, J.C., & Toro, J.A. (2012) Combining emotion and facial nonmanual signals in synthesized American Sign Language. Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility.

Thomas, F. & Johnston, O. (1995). The illusion of life: Disney animation. Disney Editions.

Valli, C. & Lucas, C. (2000) Linguistics of American sign language: an introduction. Gallaudet University Press, 2000.

Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N. & Palmer, M. (2000) A machine translation system from English to American Sign Language. In *Conference of the* Association for Machine Translation in the Americas, pp. 54-67. Springer, Berlin, Heidelberg