Improving Spatial Reference
in American Sign Language Animation
through Data Collection from Native ASL Signers

Matt Huenerfauth

The City University of New York (CUNY)
Computer Science, CUNY Queens College
Computer Science, CUNY Graduate Center
65-30 Kissena Blvd, Flushing, NY 11367 USA
matt@cs.qc.cuny.edu

Abstract. Many deaf adults in the U.S. have difficulty reading written English text; computer animations of American Sign Language (ASL) can improve these individuals' access to information, communication, and services. Current ASL animation technology cannot automatically generate expressions in which the signer associates locations in space with entities under discussion, nor can it generate many ASL signs whose movements are modified based on these locations. To determine how important such phenomena are to user-satisfaction and the comprehension of animations by deaf individuals, we conducted a study in which native ASL signers evaluated ASL animations with and without entity-representing spatial phenomena. We found that the inclusion of these expressions in the repertoire of ASL animation systems led to a significant improvement in user comprehension of the animations, thereby motivating future research on automatically generating such ASL spatial expressions.

Keywords: American Sign Language, animation, natural language generation, evaluation, accessibility technology for people who are deaf.

1 Introduction

American Sign Language (ASL) is the primary means of communication for about one-half million deaf people in the United States [15], and it is a distinct language from English. In fact, the majority of deaf high school graduates in the U.S. have a fourth-grade (age 10) English reading level [9]; consequently, many deaf people find it difficult to read English text on computers, captioned television, or other applications. Software to display computer-generated animations of sign language can make more information and services accessible to these users [3] [8] [12] [14] [18]. Unfortunately, essential aspects of ASL are not yet modeled by modern computational linguistic software; ASL signers associate entities under discussion with 3D locations around their bodies during a conversation. For example,

determiners (e.g. "THE") or adverbial signs at the end of noun phrases (e.g. "THERE") involve a pointing movement that associates the entity referred to in the noun phrase with a location in the signing space. Signers remember these spatial associations, and the movements of later signs in the performance may change based on these locations. For example, when referring one of these entities later in the conversation, a signer may use a pronoun sign that points to the appropriate location in the signing space. Some ASL verbs also change their movement or hand orientation based on the 3D locations in the signing space associated with their subject or object [13] [17]. Signers may also aim their eye-gaze or head-tilt at locations in the signing space that correspond to entities under discussion to indicate the subject or object of a verb phrase – or other linguistic features [16].

We have begun a multi-year project at CUNY to use motion-capture technology to collect and analyze a corpus of ASL sentences. Our objective is to discover techniques for ASL generation that predict when to associate entities with 3D locations, where to place them, and how these locations affect sign movements. This technology will make ASL animations more understandable, enabling their use in accessibility applications for deaf individuals with low levels of English literacy.

2 Comprehension Effect of the Use of Space in ASL Animation

In order to determine whether spatial features of ASL animations are important for producing grammatical, understandable, and natural looking animations, we conducted experiments with ASL animations in which some animations contain spatially positioned referent points and spatially modified verbs (and other versions of the animations have these linguistic phenomena removed). We scripted ASL animations for the study, designed surveys and experimental questionnaires, and conducted experimental sessions with deaf research participants. Before discussing this latest study (section 2.2), we will first briefly survey an earlier experiment on speed and pause-insertion for ASL animations (section 2.1). The methods and results of this earlier study served as a foundation for the design of our most recent study.

2.1 Earlier Work on Experimental Evaluation of ASL Animations

As part of our on-going research on ASL generation, we conduct rounds of experiments in which native ASL signers evaluate ASL animations we generate. These experiments are used to guide the development of our ASL animation technology. In order for us to conduct studies in which we focus on specific linguistic phenomena, we first needed to ensure that our animations have a baseline level of understandability. To this end, we needed a model of the speed and timing of ASL animations in order to produce high quality ASL animations for future experiments.

After examining the ASL linguistics literature for data on speed and pausing in ASL, we designed algorithms for predicting speed and timing of signs and pauses in American Sign Language animations. Our algorithms predicted the proper speed at

which specific signs in a sentence should be performed (based on linguistic factors in the discourse); our algorithms also predicted the location and length of pauses that should occur in multi-sentence ASL animations. We had hypothesized that modifying the duration of signs and inserting pauses in linguistically appropriate locations would improve the understandability of these animations for deaf individuals. This hypothesis was supported by our data collected in a study with 40 deaf research participants [10] [11].

For this speed and timing study, we designed comprehension questions and response materials, and we designed a screening questionnaire to determine whether potential research subjects had native-level fluency in American Sign Language. During the study, subjects saw some ASL animations with default settings for the speed and pausing and some animations that had been processed by our speed and timing algorithms. We found that the use of our speed and timing algorithms led to a significant increase in comprehension of ASL animations. The results indicated that there is a benefit to processing ASL animations with our pause-insertion and sign-duration algorithms – a benefit that goes beyond merely inserting additional time into the animations. Another result of this study was that we determined an overall speed at which computer-generated ASL animations should be displayed to maximize user-satisfaction with the speed and their performance on comprehension questions: approximately 1.1 signs per second [10] [11].

2.2   Current Experiments Evaluating Use of Space in ASL Animations

We conducted a new study to determine whether there is a comprehension benefit when ASL animations include signs that associate locations in 3D space with entities under discussion. As discussed in section 1, ASL signers often point to locations in 3D space around their bodies to represent people or concepts under discussion; they will point to these locations again to refer to these entities later in the conversation. Certain ASL signs will also modify their movement path to indicate these locations in space. Our goal with this study was to understand the degree to which spatial phenomena in ASL lead to comprehension differences in native ASL signers.

ASL Animations in the Study. We designed 16 paragraph-length stories in ASL, and we scripted computer animations of these stories using SignSmith Studio, commercial sign language animation software by VCom3D, Inc. [19] This product allows users to script an ASL performance using a dictionary of signs, a fingerspelling generator, and limited control of eye-gaze and facial expression. Details about how and why ASL animations were created in this manner for our earlier experimental study on speed and timing are discussed in [11]. The ASL passages were on a variety of topics: short news stories, adaptations of encyclopedia articles, fictional narratives, and 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). A second version of each story was produced in which any associations of entities with

Fig. 1. Screenshot of the animated character from the animations used in the study

locations in space were removed. Fig. 1 contains a screen image of the ASL signing character used in the animations.

Fig. 2 contains transcripts of two different versions of same story. Fig. 2(a) is the version that uses 3D space to represent entities under discussion, and Fig. 2(b) is the version without this use of space. In Fig. 2(a), the signer uses the pronoun "HE(x)" to point to a location in 3D space that will represent the character Simon. During the "HE(x)" sign, the signer points to a location "x" while aiming his eye-gaze and head-tilt at that location. Later, when the signer needs to refer to Simon, he will again point to this location "x" to refer to him. (The signer later uses "HE(y)" to set up a different location "y" in space to represent the character Jeff in the story.) In Fig. 2(b), the signer instead performs the full name "Simon" using fingerspelling (written in the

MY 2 COUSIN BOTH WORK FACTORY. HE(x) fs-SIMON WORK CAR FACTORY SINCE LONG 20 YEARS. HE(y) fs-JEFF WORK ELECTRICITY FACTORY SINCE 10 YEARS. HE(x) WORK GOOD BENEFIT FREE DOCTOR FREE DENTIST GOOD RETIRE MONEY. HE(x) WORK BOSS NOT NICE ALWAYS YELL YELL. HE(y) WORK BOSS NICE WOMAN ALWAYS HELP BUT HE(y) GET NONE BENEFIT MUST PAY DOCTOR DENTIST OTHER OTHER. HE(y) WORK NIGHT. HE(x) WORK DAY. HE(x) WANT SAME WORK FOREVER.   HIS(y) WIFE TELL HE(y) NEED EARN MORE MONEY, I THINK MAYBE HE(y) WILL CHANGE JOB.

MY 2 COUSIN BOTH WORK FACTORY. fs-SIMON WORK CAR FACTORY SINCE LONG 20 YEARS.   fs-JEFF WORK ELECTRICITY FACTORY SINCE 10 YEARS.   fs-SIMON WORK GOOD BENEFIT FREE DOCTOR FREE DENTIST GOOD RETIRE MONEY. fs-SIMON WORK BOSS NOT NICE ALWAYS YELL YELL. fs-JEFF WORK BOSS   NICE WOMAN ALWAYS HELP BUT fs-JEFF GET NONE BENEFIT MUST PAY DOCTOR DENTIST OTHER OTHER.   fs-JEFF WORK NIGHT. fs-SIMON WORK DAY. fs-SIMON WANT SAME WORK FOREVER.   fs-JEFF HIS WIFE TELL fs-JEFF NEED EARN MORE MONEY, I THINK MAYBE fs-JEFF WILL CHANGE JOB.

Fig. 2. Gloss transcripts of two different versions of the same ASL story. In Fig. 2(a) on the left, the signer associates locations in the signing space with characters in the story. In Fig. 2(b) on the right, the signer does not associate characters with locations in space.

transcript as "fs-SIMON") each time that he refers to Simon.   In this version of the animation, Simon is not associated with a 3D location in the signing space.

Participants and Experiment Design. We have previously discussed how to recruit and screen signers for experiments evaluating animations [10] [11]. Advertisements for the study were posted on Deaf community websites in New York City, and potential participants were asked if they had grown up using ASL at home or attended an ASL-based school as a young child. Eight ASL signers in the New York City area were recruited for the study, and the experimental sessions were conducted by a deaf native ASL signer (maintaining an ASL-centric environment for the study is also important for soliciting accurate evaluations [11]). The 3 men and 5 woman were ages 23-52 (median age 37). Four had ASL-signing parents, and five attended schools for the deaf using ASL. Six learned ASL before age 6, and two learned ASL later in life but attended college with instruction in ASL. Five subjects had a deaf partner with whom they used ASL, and all used ASL in daily work or social contexts.

Participants in the study viewed 16 ASL stories of two versions: "*space*" and "*no-space*" animations. The *space* animations included ASL grammatical constructions that associated entities under discussion with locations in the signing space; the *no-space* animations did not. The animations were displayed at a speed of 1.1 signs per second – based on the results of our earlier study [11]. A fully-factorial within-subjects design was used such that: (1) no participant saw the same story twice, (2) the order of presentation was randomized, and (3) each participant saw eight animations of each version (*space* vs. *no-space*). After viewing each story, subjects were asked 1-to-10 Likert-scale questions about the grammatical correctness, understandability, and naturalness of movement of the animations.

Participants' comprehension of the animations was also measured using multiple-choice comprehension questions. After each story, the animated character performed four multiple choice questions in ASL, and after each question, the character would sign a set of possible multiple-choice answers to that question. The participants in the study had a paper response form with clip-art pictures corresponding to the answer choices for each comprehension question, and they had to circle the appropriate picture to indicate their response. The comprehension questions asked about basic facts from the passage. Questions actually measured a mixture of a participant's recall and comprehension of the passage: Participants were not allowed to watch the story more than once; so, they could not replay it to look for an answer. Fig. 3 contains transcripts of some multiple choice questions, and Fig. 4 contains some

---

WHAT BENEFITS fs-SIMON HE(x) GET FREE?
   DOCTOR. DENTIST. RADIO. TV. RETIREMENT MONEY. COMPANY CAR.
HOW-MANY YEAR fs-JEFF HE(y) WORK TV FACTORY?
   5. 10. 15. 20.
WHICH COUSIN MEAN BOSS?
   fs-SIMON HE(x). fs-JEFF HE(y).
WHEN fs-JEFF HE(y) WORK?
   MORNING. AFTERNOON. EVENING. NIGHT.

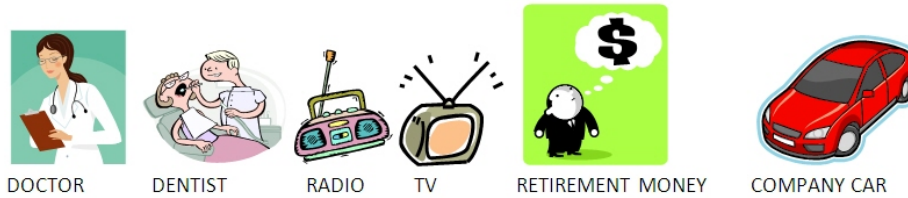**Fig. 3.** Transcripts of the comprehension questions and answer choices for stories in Fig. 2

Fig. 4. Clip-art images appearing on the response form for the first question shown in Fig. 3

clip-art images from the response form used in the study. The question types used in this study were based on our earlier experimental study discussed in section 2.1; additional details about the question types and use of clip-art are discussed in [11].

Results of the Study. Fig. 5 displays the average comprehension question scores for participants in the study for the *space* vs. *no-space* animations. To score the comprehension question responses for each story, we subtracted the number of correctly circled choices minus 25% of any incorrect circles or missing circles. There were an average of four multiple-choice options for each question; so, random guessing would produce a score of 0%. As shown in Fig. 5, there was a statistically significant difference in comprehension scores between the *space* and *no-space* versions of the animations (p < 0.05, Mann Whitney U-Test).

There are several possible explanations for the comprehension benefit observed in this study. One explanation is that the use of space facilitated the participants' remembering information in the story; associating information content with locations in space may allow viewers to more easily encode or chunk information, which
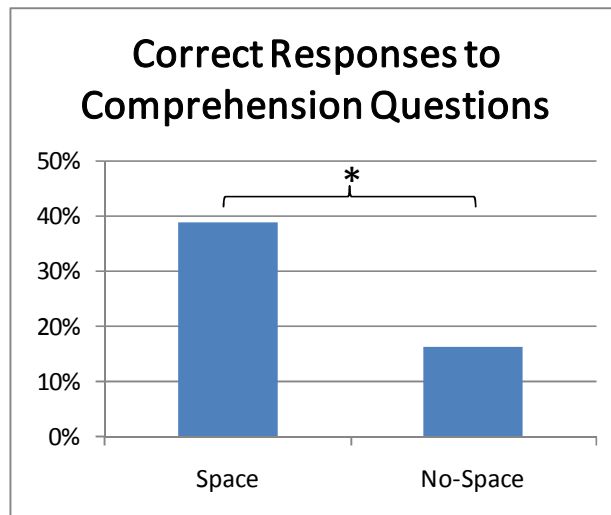

Fig. 5. Average comprehension question scores for the *space* and *no-space* animations

improved their recall of this information when later answering the comprehension questions. Another explanation for this benefit is that the use of 3D space around the signer made the ASL animations easier to follow, and this allowed the viewers to understand a larger portion of the animation.

Fig. 6 shows the participants' average response scores for the 1-to-10 Likert-scale subjective questions signers were asked in the study. While the differences between the *space* and *no-space* groups are not statistically significant, participants did give slightly higher scores for the grammatical correctness, understandability, and naturalness of movement for the *space* animations in this study. Considering the significant difference in the comprehension question scores, it is interesting how little effect was observed on the Likert-scale scores. While the use of space in the animations did not appear to overtly draw the attention of participants in the study (as reflected in the small impact on their subjective Likert-scale responses), it did have a significant impact on their comprehension scores for the animations.
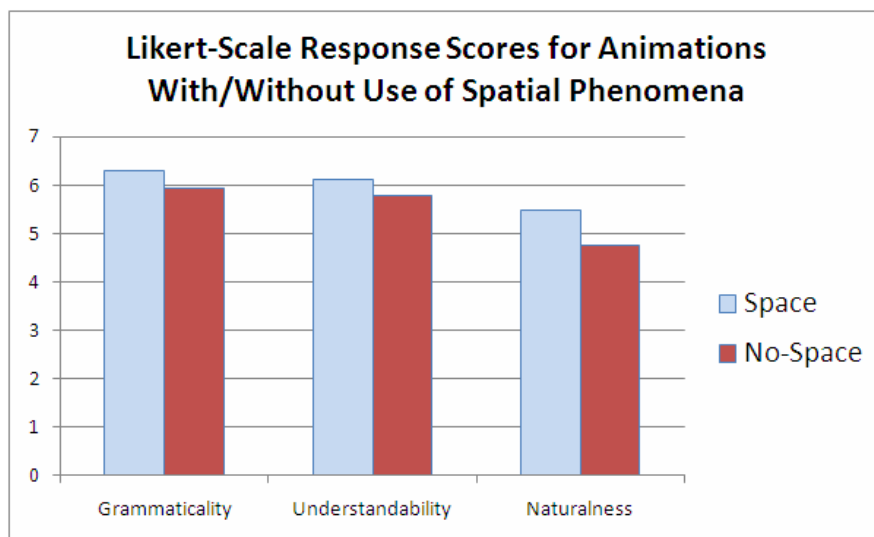

Fig. 6. Average scores for grammatical correctness, understandability, and naturalness of movement as reported on 1-to-10 Likert scales by participants in the study

3   Discussion

Software to generate animations of ASL has the potential to provide important accessibility benefits for deaf users with low English literacy; however, current computational linguistic software cannot produce many ASL signs or constructions. Difficult phenomena arise from the way that signers associate entities under discussion with locations in space around their bodies. Pronouns involve pointing to these locations and movements of many other signs change to show relationships with the entities located in space (e.g. a verb sign's motion path may change to indicate its

subject or object location). Current ASL generation technology cannot predict when a signer should associate an entity with a location in space or how to dynamically modify the movement of signs based on these spatial associations. The question we have addressed in this study is whether this limitation has an impact on the degree to which users understand and remember information from ASL animations.

The experiments we have conducted with deaf participants have demonstrated that there is a significant difference in deaf users' comprehension of ASL animations when they include these spatial phenomena. This work empirically supports the claim that these aspects of automatically generated ASL animations will have measurable benefits for deaf users who view these animations.

Another contribution of this research is that we have developed a successful protocol for conducting experiments involving deaf research subjects in which they formally critique computer animations of American Sign Language. We have also designed an approach for using multiple-choice questions with a paper answer sheet that includes clip-art images for answer choices. This experiment design has allowed us to successfully measure user-comprehension of our animations, and it is a useful resource for future research in this area. For use in the experiments discussed in sections 2.1 and 2.2, we have designed approximately 25 paragraph-length stories in ASL, and we have produced computer animations of these stories. These animation scripts are useful for testing alternative settings and algorithms to improve the understandability and naturalness of ASL animations. Having a library of such stories (which can be easily modified to experimentally test alternative ASL generation technologies) is an important resource for conducting research in this area. This experimental study has also established baseline levels of user comprehension scores when using these experimental protocols and this set of comprehension questions.

4 Broader Research Objectives and Progress

In additional to the experimental evaluation studies of ASL animations that have been discussed in this paper, there are several other important research objectives for our ASL generation research project. We are creating a corpus of ASL movement data collected from native signers (in a motion-capture suit and gloves), and we will annotate this corpus with linguistic features relating to the establishment of entity-representing locations in space. This corpus will be analyzed using machine learning techniques to learn when/where these entity-representing locations are established and how 3D motion paths of signs are parameterized on those locations. As we incorporate these findings into our ASL generation software, we will conduct additional evaluation studies with native ASL signers to evaluate the 3D animations that result and guide the development of our technology. Earlier researchers have collected video-based sign language corpora [2] [5] [6] [7] [16] or collected samples of sign language using motion-capture [1] [4] [20]. For this project, we are collecting a corpus of ASL using a motion-capture suit and gloves, and we are recording multi-sentence passages in which signers associate entities with locations in space. This will allow us to study spatial-association phenomena using machine learning techniques applied to this linguistically annotated motion-capture corpus.

For our motion-capture data collection, two 22-sensor Ascension Cybergloves will record the signer's handshapes, and an Intersense IS-900 acoustical/intertial motion-capture system will record the location and orientation of the signer's head, hands, and middle of torso. An Animazoo IGS-190 body suit will record joint angles for the wrists, elbows, shoulders, clavicle, and waist; small sensors on the suit record angles of the bones. A lightweight head-mounted eye-tracker (compensating for head motion) will record a 3D vector of the signer's eye gaze direction. Two digital high-definition video cameras will film front and side views of the signer. This particular combination of equipment being used for simultaneous recording of a human performance is unique to our project. For this reason, configuring the hardware and software for this project has been an essential activity during the first year of the project. During this time, we have also designed and written protocols for efficiently calibrating various motion-capture technologies in an accessible manner when recording data from research participants who are deaf. Motion-capture data collection and linguistic annotation will begin in the summer of 2009.

This ASL motion-capture corpus will enable exciting new research in ASL generation and recognition of ASL from human-recorded movements. Many other sign languages internationally have similar linguistic phenomena to ASL and may also benefit from research on ASL models of entity-establishment and placement.

The long-term goal of our research is to learn how ASL signers set up locations in space to represent entities under discussion and how they modify the movement paths of ASL signs based on this use of space. The goal is to produce computer animations of ASL that are more grammatical, easier to understand, and appear more natural. Our experimental studies suggest that improvements in spatial-association and sign-modification will make ASL animations more useful in accessibility applications.

References

1. Brashear, H., Starner, T., Lukowicz, P., Junker, H.: Using multiple sensors for mobile sign language recognition. In: IEEE International Symposium on Wearable Computers, p. 45. IEEE Press, New York (2003)
2. Bungeroth, J., Stein, D., Dreuw, P., Zahedi, M., Ney, H.: A German Sign Language Corpus of the DomainWeather Report. In: Vettori, C. (ed.) 2nd Workshop on the Representation and Processing of Sign Languages, pp. 2000–2003. ELRA, Paris (2006)
3. Chiu, Y.-H., Wu, C.-H., Su, H.-Y., Cheng, C.-J.: Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. IEEE Trans. on Pattern Analysis & Machine Intelligence 29(1), 28–39 (2007)
4. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: Tessa, a system to aid communication with deaf people. In: 5th International ACM Conference on Assistive Technologies, pp. 205–212. ACM Press, New York (2002)

5. Crasborn, O., van der Kooij, E., Broeder, D., Brugman, H.: Sharing sign language corpora online: proposals for transcription and metadata categories. In: Streiter, O., Vettori, C. (eds.) LREC 2004 Satellite Workshop on Representation and Processing of Sign Languages, pp. 20–23. ELRA, Paris (2004)

6. Crasborn, O., Sloetjes, H., Auer, E., Wittenburg, P.: Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. In: Vettori, C. (ed.) 2nd Workshop on the Representation and Processing of Sign Languages, pp. 82–87. ELRA, Paris (2006)

7. Efthimiou, E., Fotinea, S.-E.: GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 657–666. Springer, Heidelberg (2007)

8. Fotinea, S.-E., Efthimiou, E., Caridakis, G., Karpouzis, K.: A knowledge-based sign synthesis architecture. Universal Access in the Information Society 6(4), 405–418 (2008)

9. Holt, J.A.: Stanford Achievement Test - 8th Edition: Reading comprehension subgroup results. American Annals of the Deaf 138, 172–175 (1993)

10. Huenerfauth, M.: Evaluation of a Psycholinguistically Motivated Timing Model for Animations of American Sign Language. In: 10th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 129–136. ACM Press, New York (2008)

11. Huenerfauth, M.: A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. ACM Transactions on Accessible Computing (in press)

12. Kennaway, J., Glauert, J., Zwitserlood, I.: Providing signed content on the Internet by synthesized animation. ACM Transactions on Computer-Human Interaction 14(3), Article 15, 1–29 (2007)

13. Liddell, S.: Grammar, Gesture, and Meaning in American Sign Language. Cambridge University Press, Cambridge (2003)

14. Marshall, I., Safar, E.: Grammar development for sign language avatar-based synthesis. In: Stephanidis, C. (ed.) Universal Access in HCI: Exploring New Dimensions of Diversity. Proceedings of the 11th International Conference on Human-Computer Interaction (CD-ROM), vol. 8. Lawrence Erlbaum Associates, Mahwah (2005)

15. Mitchell, R., Young, T.A., Bachleda, B., Karchmer, M.A.: How Many People Use ASL in the United States? Why Estimates Need Updating. Sign Language Studies 6(4), 306–335 (2006)

16. Neidle, C., Kegl, D., MacLaughlin, D., Bahan, B., Lee, R.G.: The Syntax of American Sign Language: Functional Categories and Hierarchical Structure. The MIT Press, Cambridge (2000)

17. Padden, C.: Interaction of Morphology and Syntax in American Sign Language. Outstanding Dissertations in Linguistics, Series IV. Garland Press, New York (1988)

18. Stein, D., Bungeroth, J., Ney, H.: Morpho-syntax based statistical methods for sign language translation. In: Proceedings of the European Association for Machine Translation, Allschwil, Switzerland, pp. 169–177 (2006)

19. VCom3D Company website, http://www.vcom3d.com

20. Vogler, C., Metaxas, D.: Handshapes and movements: Multiple-channel ASL recognition. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 247–258. Springer, Heidelberg (2004)