Effect of Spatial Reference and Verb Inflection on the Usability of Sign Language Animations

Matt Huenerfauth

Department of Computer Science,
Queens College,
The City University of New York,
65-30 Kissena Boulevard,
Flushing, NY, 11367 USA

Pengfei Lu

Department of Computer Science,
Graduate Center,
The City University of New York,
365 Fifth Avenue,
New York, NY, 10016 USA

Author's copy of article published as:

Matt Huenerfauth, Pengfei Lu. 2012. Effect of spatial reference and verb inflection on the usability of sign language animations. *Universal Access in the Information Society*: Volume 11, Issue 2 (June 2012), pages 169-184. doi: 10.1007/s10209-011-0247-7

Abstract Computer-generated animations of American Sign Language (ASL) can improve the accessibility of information, communication, and services for the significant number of deaf adults in the US with difficulty in reading English text. Unfortunately, there are several linguistic aspects of ASL that current automatic generation or translation systems cannot produce (or are time-consuming for human animators to create). To determine how important such phenomena are to user satisfaction and the comprehension of ASL animations, studies were conducted in which native ASL signers evaluated ASL animations with and without: establishment of spatial reference points around the virtual human signer representing entities under discussion, pointing pronoun signs, contrastive role shift, and spatial inflection of ASL verbs. It was found that adding these phenomena to ASL animations led to a significant improvement in user comprehension of the animations, thereby motivating future research on automating the generation of these animations.

Keywords. American Sign Language, animation, evaluation, sign language, spatial reference, verb inflection, accessibility technology for people who are deaf.

Abbreviations. ASL: American Sign Language; HCI: human-computer interaction; MT: machine translation; BSL: British Sign Language.

1 Motivations and background

American Sign Language (ASL) is the primary means of communication for about one-half million deaf people in the US [2]. ASL is a full natural language with a distinct word order, syntax, and lexicon from English – it is not simply a representation of English sentences using the hands. While degree of hearing loss, educational experiences, and family background affect whether someone uses ASL, there are many deaf individuals for whom ASL is their first or preferred language. Although written-English reading is an important part of the curriculum for deaf students in the US, a lack of auditory exposure to English during the language-acquisition years of childhood often leads to lower literacy for many deaf adults. In fact, the majority of deaf high school graduates in the US (age 18 and above) have only a fourth-grade English reading level [3] (fourth-grade US elementary school students are typically age 10).

The information above has focused on ASL and the US, and it is important to note that sign languages are not universal internationally. Different countries have unique sign languages, with grammars and vocabularies that are distinct from each other. The sign language in a particular country is also typically independent from the local spoken/written language of that country; the sign language often has a distinct grammar and vocabulary from the spoken one. While this article focuses on ASL, various other sign languages have been the focus of animation and accessibility research, including Japanese Sign Language, British Sign Language, Sign Language of the Netherlands, Greek Sign Language, and others. While these sign languages are not mutually intelligible, they do share certain key linguistic aspects of ASL that are the focus of this article – the use of spatial reference and verb inflection (discussed in section 1.2). While the experimental results reported in sections 2 and 3 of this article are specific to ASL and deaf research participants from the US, the methodologies described in this article for measuring the usability of sign language animations could be replicated for these other sign languages and other users internationally. The data presented in Sect. 2.3 of this article were collected during a study that was first described in a presentation at the Universal Access in Human Computer Interaction conference [1]. This article presents a novel analysis of a subset of the data from this earlier study – focusing on contrastive role shift in American Sign Language (ASL) animations. Section 3 of this article presents the results of a newly conducted, unpublished, larger study focusing on verb inflection in ASL animations.

1.1 Sign language animation technologies

Most technology used by people who are deaf does not address the literacy issue discussed above; many deaf people find it difficult to read the text on a computer screen or on a television with closed-captioning. Software to present information in the form of animations of ASL could make more information and services accessible to deaf people in the US Instead of English text, which may be difficult to read for some users, these technologies would display an onscreen animated character performing sign language. In addition to making information sources more accessible to deaf users, sign language animation technology could be used to produce new types of software and services specifically for these users. For instance, ASL animation software could be used to build new educational software for deaf children to help them improve their English literacy skills; further applications of sign language technology are discussed in [4].

There are several reasons why it is desirable to present information in the form of ASL instead of written English text on software and Web sites for people who are deaf. As discussed above, many people who are deaf have lower levels of literacy in their local written language, but they have more sophisticated fluency and better comprehension of information in the form of sign language. These users may feel more confident and comfortable when interacting with computer technology using their preferred language. Many people who are deaf also feel a personal connection or shared identity with the community of other deaf signers, and the use of ASL serves as a unifying cultural feature of the Deaf community in the US [5]. Thus, Web sites and software specifically targeted to this community may find greater acceptance when using ASL instead of written English. There is also a trend in the US toward "bilingual/bicultural" education of deaf children [6]; in such contexts, bilingual or parallel presentation of information in the form of written English text and ASL is desirable.

Previous ASL animation research can be divided into two categories: scripting software and machine translation (MT) software. Scripting software allows a human to "word process" a set of sentences of ASL by placing individual signs from an animation dictionary onto a timeline; the software later synthesizes an animation of a virtual human character based on this timeline. For example, the European eSIGN project [7, 8] has produced technologies for content developers to build sign databases using a symbolic notation, build scripts of sign language sentences for use on web pages, and allow users to view animations using a plug-in for their web browser. Sign Smith Studio [9] is a commercially produced software system to allow users to script ASL animations. In addition to a dictionary of signs, this system includes a fingerspelling generator, approximately 50 facial expressions, and control of the virtual character's eye-gaze direction, head-tilt, and shoulder-tilt. To use the software, a human author (who knows ASL) arranges signs, facial expressions, and other movements on a set of parallel timelines. When the human author is happy with the result, the timelines can be saved (for future editing) or an animation can be synthesized in which a virtual human character performs the ASL script.

Sign language MT software differs from scripting software in that the computer must automatically produce an ASL sentence, when given an input English sentence to be translated. These systems must plan the sequence of signs and facial expressions for a sign language sentence, and then (like the scripting systems) they must synthesize the resulting sentence as an actual animation. Researchers have studied how to develop MT systems and component technologies for a variety of sign languages, e.g., [10-18]. Some researchers have built short-lived MT systems that translated only a few sample English input phrases into sign language [19], some more developed rule-based systems [12], and some statistical systems [14].

The scripting and MT technologies discussed above produce *animations* of a virtual human character performing sign language – not *videos* of human signing. They plan sign language content using a symbolic encoding of the movements to be performed and synthesize animations of a virtual character as their output. While, at first glance, it would seem that videos of human performances would be preferable, there are advantages to the use of animations in software or Web sites for people who are deaf. If MT or generation software instead merely concatenated signs from a video-recorded dictionary of human signers, it would be difficult to produce smooth transitions between signs (where two videos are spliced), create variations in the performances of individual signs based on grammatical features of the sentence, or produce correct combinations of required facial expressions with the signs. Computer-synthesized animations allow for greater control, blending, and modulation.

Even in "scripting" applications in which a human is authoring ASL sentences, there are still benefits to the use of animations over video. If the information content on a Web site is frequently updated, then the use of a video of a human signing may be inefficient. Each time a small detail of the information is changed, a human would need to be recorded re-performing the sign language content. Because of how ASL signers set up locations in the 3D space around their bodies to represent entities under discussion (see section 1.2), it can be difficult to splice a new sentence into the middle of a video of an ASL performance, i.e., the human would have to remember exactly how everything was set up in space at that point in the discourse. If the information content had been created in a scripting system, then the human author could simply edit the script to change the minor detail and then re-synthesize the animation.

This ease-of-editing is also important if ASL content is being produced in a collaborative manner, in which several authors are contributing to the information content to be posted on a Web site. For wiki applications like this, the use of video would be awkward – it would be incoherent for different humans to appear in spliced video segments – performing different signs and sentences of a single block of information. On the other hand, if multiple human authors collaboratively edit a script for a virtual human signer, then the animation can be synthesized to produce a seamless performance. In other online contexts in which human ASL signers want to contribute content in an anonymous manner, the use of videos is not possible. Eye-gaze direction, head movement, and facial expressions are an essential part of ASL, without which the meaning of sentences cannot be understood. Thus, videos of ASL must include the face of the human signer. Instead, humans can script animations of ASL to be performed by a virtual human signer without revealing their identity.

Animations of ASL have advantages over videos in other specific applications. The novelty and attention-getting aspect of the use of 3D animation technology can be useful in advertising targeted toward the deaf community, and the visual appeal of animated characters could make them useful for designing education software targeted toward deaf children or teens. For individuals who are learning sign language or individuals with limited visual acuity, animations can be preferable to videos because they can be viewed from different angles, at different speeds, or as performed by different virtual humans.

1.2 Spatial reference and verb inflection in ASL

In order to discuss the limits of current ASL animation technologies and to motivate the evaluation experiments discussed in sections 2 and 3, it is useful to define several types of linguistic phenomena that occur in human ASL signing. All of these phenomena involve the use of the 3D space around the signer (often called the "signing space") to represent entities under discussion. During a conversation, ASL signers often associate people, concepts, or other entities under discussion with 3D locations around their bodies. For example, by pointing at a location in the surrounding space at the beginning or at the end of a noun phrase mentioning a new entity, the human signer associates the entity referred to in the noun phrase with that location. Signers remember these spatial associations, and the movements of later signs in the performance may change based on these locations. When referring to one of these entities later in the conversation, a signer may use a pronoun sign (which also looks like a pointing gesture) aimed at the appropriate location in the signing space.

For example, a signer may discuss someone named "John." After mentioning John the first time, the signer may point to a location in space where a spatial reference point is created that represents him. On future occasions in which the signer wants to refer to John, the signer will simply point to this location. Often, a signer may not repeat the name of a person or concept again; all future references to that entity will consist of a pointing sign to a 3D location. This article will refer to the process of establishing 3D locations representing entities under discussion as ***spatial reference***, and the locations will be called ***spatial reference points***. Fig. 1 contains images of a native ASL signer performing pointing signs that establish spatial reference points. During these pointing signs, human signers may also aim their eye-gaze or head-tilt at the 3D location [20]. When signers refer later to an entity in the signing space, they will again perform a pointing sign.

Fig. 1. The images from videos of a native ASL signer depict two examples of a pointing pronoun sign. In the first example, the signer is pointing to a spatial reference point on his right side. In the second example, the signer is pointing to a spatial reference point on his left side.

In addition to the pronoun "pointing signs," spatial reference points also affect other types of ASL movements. For example, during ASL verb signs, human signers will sometimes aim their eye-gaze or head-tilt at the spatial reference point that corresponds to the subject or object of the verb [20]. Sometimes the subject/object is not mentioned in the sentence, and this use of eye-gaze and head-tilt is the only way in which the signer conveys the identity of the subject/object. On other occasions, a signer may set up a spatial reference point on the left side of the signing space and one on the right; later, when discussing these entities in a contrastive manner, a signer may twist the waist, so that his/her torso is aimed at the left or right side when discussing each of these entities in turn. This ***contrastive role shift*** emphasizes which entity is being discussed during the conversation, and it is usually used to emphasize the sense of contrast. Fig. 2 contains images of a virtual human character performing ASL; this character was used in the experimental studies discussed in Sects. 2 and 3. Fig. 2(c) contains an image of an animated virtual human character twisting at the waist in order to perform a contrastive role shift.

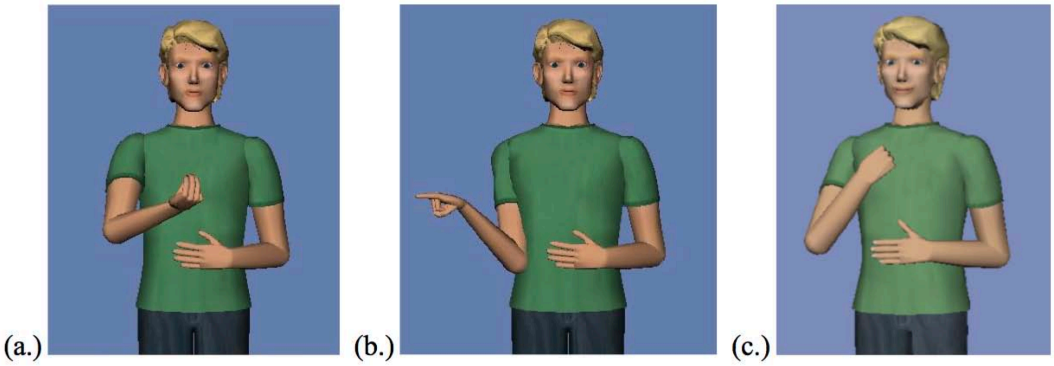

Fig. 2. Screenshots of an animated character performing ASL produced using Sign Smith Studio [9]. In **a**, the character faces front while performing a sign. In **b**, the character points to refer to a spatial reference point (this is glossed as "POINT(left)" in Fig. 4). In **c**, the character is facing to the left side while performing contrastive role shift; in Fig. 4., this is indicated by a span of text beginning with the notation: "*(facing-left)*."

Some ASL verb signs also change their motion paths or hand orientation to indicate the 3D location where a spatial reference point has been established for their subject, object, or both [21, 22]. Generally, the motion paths of these ***inflecting verbs*** change, so that their direction goes from the subject to the object; however, their paths can be more complex than this. Each ASL inflecting verb has a standard motion path that is affected by the subject's and the object's 3D locations – producing a motion path that is unique to the specific verb, the specific 3D location of the subject, and the specific 3D location of the object. When a verb is inflected in this manner to indicate its subject/object, the signer does not need to mention the subject/object in the sentence. Fig. 3 shows an example of a human performing a spatially inflected verb.

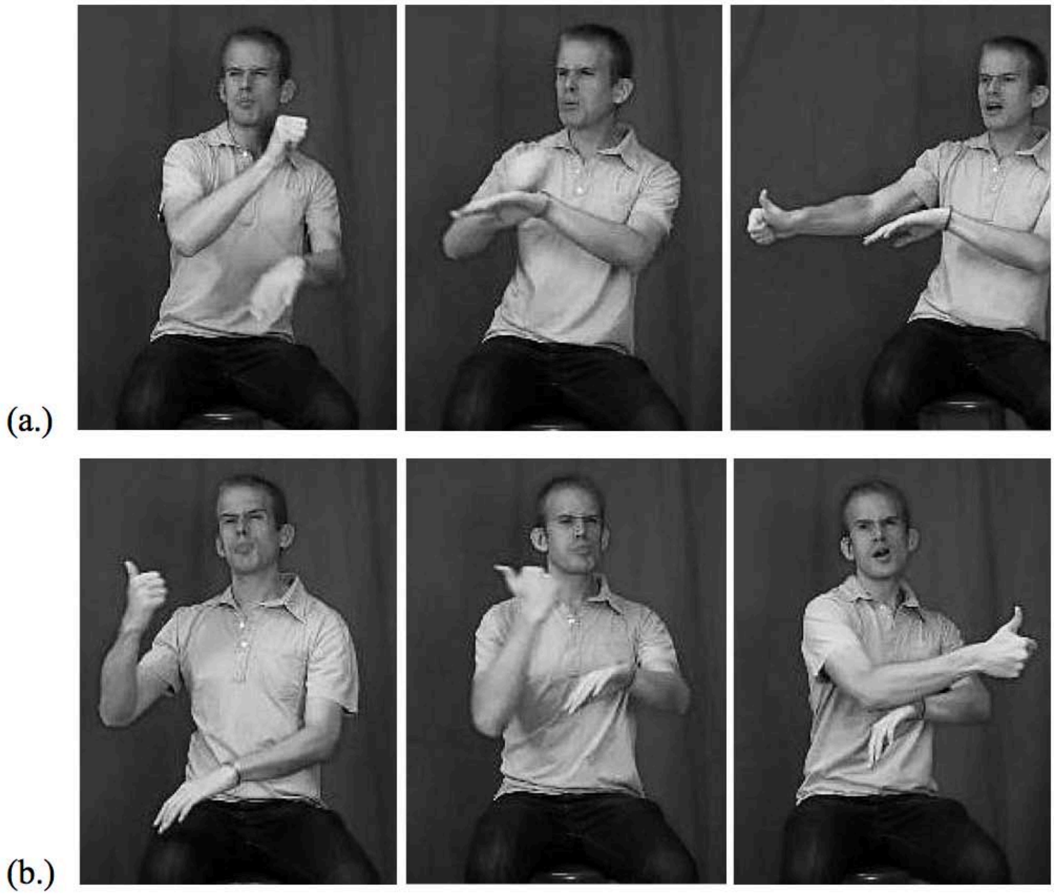

Fig. 3. Two different performances of the ASL inflecting verb "BLAME." In **a**, the signer is performing the verb BLAME with the subject of the verb on the signer's left and the object of the verb on the signer's right. In **b,** the signer is performing the verb BLAME with the subject of the verb on the signer's right, and the object, on the signer's left.

1.3 Lack of spatial phenomena in current ASL technology

Unfortunately, modern sign language scripting, generation, or MT software does not *automatically* handle the spatial aspects of ASL discussed above. Humans authoring ASL animations using state-of-the-art scripting systems can manually add pointing signs to the performance that aim at various locations around the virtual human signer – these signs can be used to establish spatial reference points during the ASL animation being created, as in Fig. 2(b). The human author of an ASL animation can also choose to manually modify where the eyes of the virtual human character are aimed – in ASL, the signer typically also uses eye-gaze direction to indicate spatial reference points in space. A human author scripting an ASL animation can also manually control the tilt of the shoulders of the virtual human character (left or right) to indicate contrastive discussion of entities established on the left or right side of the signing space, as in Fig. 2(c). However, none of these types of ASL animation performances are *automated* – the human author of the sign language sentences must control the animation to manually add these features. Researchers have proposed various approaches for modeling this use of space around a signer [23]. Current ASL scripting software does not automatically suggest locations in space where entities under discussion should be assigned, how to coordinate the use of pointing and eye-gaze for establishing spatial reference points, nor determine when the shoulders should tilt to indicate different sides of the signing space. Further, state-of-the-art generation and MT systems, which automatically plan sign language sentences, do not produce animations in which spatial reference points are established.

Further, no current ASL scripting, generation, or machine translation systems automatically modify the motion path of ASL verb signs based on the 3D locations in space at which the subject and object of the verb has been established. A human who is scripting an ASL sentence would need to manually modify the hand orientation and motion path of the virtual human in order to produce inflected verb performances – a meticulous and detailed process that can make producing ASL animations with inflected verbs extremely time-consuming. Further, not all ASL scripting systems give the human user the ability to make fine-grained decisions about the exact placement and manipulation of the hands: many simply allow standard signs from a dictionary to be used.

ASL generation and MT systems generally use uninflected sign language verb forms; their lexicon of signs do not include rules for how to modify the motion path of each type of verb based on locations established for the verb's subject/object. The sign language animation generation system with the most sophisticated handling of spatial reference and verb inflection thus far was part of the European eSIGN project, which included technology to translate limited types of English sentences into animations of British Sign Language (BSL) [24]. Their project included a full machine translation system from English to several European sign languages. Their sign language animation generator could automatically associate a noun with one of a finite number of locations around the animated signer's body; however, it had no principled way to select which nouns to assign to a spatial reference point, nor where to position each spatial reference point. The generator could also produce a small number of BSL verbs by moving their motion path in a line from one location in space to another; however, the system did not address verbs (or other BSL movements) with more intricate motion paths that are parameterized on the 3D locations of spatial reference points in more complex ways. This translation technology was not able to handle a wide variety of arbitrary English input sentences (it was a research prototype system), and these technologies have not been adapted for use in American Sign Language animation systems.

1.4 Measuring usability of sign language animations

The overall objective of the research presented in this paper is to improve the understandability and correctness of animations of ASL, which would make these animations more useful and effective at conveying information to deaf users in an accessible manner. Thus, it is important to determine whether the lack of spatial reference and verb inflection phenomena in current ASL animation systems has a significant effect on the quality of these animations. ASL animations are a relatively new technology, and they have only recently begun to appear in software and Web sites for deaf users [4]. The HCI research literature has not yet begun to define what aspects of these animations contribute to their *usability*. While professionals have debated the precise meanings of the terms "accessibility" and "usability" for technologies for people with disabilities [25], commonly accepted HCI definitions of "usability" reflect its multi-faceted nature. Nielsen's definition of usability includes five properties: "easy to learn, efficient to use, easy to remember, low error rate, and meets user satisfaction" [26]. The International Standards Organization defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [27].

To measure "effectiveness" or "efficiency," a specific task must be defined. The ASL animations to be evaluated would need to be embedded in some application, and researchers would measure the performance of users at some task. For instance, ASL animations could be used on a Web site that allows users to make airline reservations, and the accuracy and speed of users at a reservation-booking task could be evaluated. To compare alternative forms of ASL animations, then researchers would perform multiple trials: some using one version of an ASL animation, and some using another. While there is value to such *extrinsic* forms of evaluation, in order to evaluate the quality of ASL animations across a range of potential applications, a researcher would need to perform a variety of such studies with different applications/tasks.

As much as possible, it is of interest to evaluate the usability of the ASL animations themselves in a more *intrinsic* manner. In an application-based evaluation like the airline reservation scenario above, there are a variety of factors, beyond the ASL animations themselves, which may affect a user's success. It is, therefore, useful to isolate the sub-task for which the quality of the ASL animation is most important: the conveying of information content to a user. The primary motivation for research on ASL animation technologies is that they have the potential to more successfully convey information content to deaf users who may have lower levels of written-English literacy.

Thus, prior research of the authors on other aspects of ASL animations – not spatial reference and verb inflection – has addressed task-based evaluations of ASL animation that focus on how well the user of the animation has successfully comprehended (and remembered) the information content being conveyed [28-30]. In some studies, users watch ASL animations that describe a scene, and they later select a photograph that best matches the description [30]. In other studies, users watch an ASL animation and then watch comprehension questions (also presented in ASL) about the information in the original ASL animation; users are asked to select the best answer to each question by circling a clip-art picture that matches their choice [29]. (See Fig. 7 in section 2.2 for an example). The use of photographs or pictures in these tasks allows users who might have low levels of English literacy to participate successfully. To measure "efficiency" of conveying information, in some studies, the speed of presentation of the ASL animations was varied to measure how well users are able to understand the information content [29].

Just as in any HCI study, it is essential to recruit and screen for potential participants correctly and to control aspects of the experimental environment. Many deaf adults learn ASL later in life and possess varying degrees of fluency; these signers can be more "forgiving" of ASL that includes errors [20]. To recruit the most critical judges possible, a screening questionnaire was designed [28] to identify adult users of ASL who are "native signers," individuals who have grown up with ASL as a first language. It has also been observed that ASL signers in a setting surrounding by non-native signers or English users tend to modify their signing to become more English-like, and they tend to accept more English-like non-fluent ASL as being correct [20, 31]. Thus, a native ASL signer has conducted the ASL evaluation studies in the authors’ prior work [29], and the use of English in the laboratory environment was minimized.

To measure the more subjective “satisfaction” aspect of usability, deaf users who have viewed ASL animations were also asked to answer subjective questions concerning grammatical correctness, ease of understanding, and naturalness of movement of the virtual human character [29, 30]. These questions are explained in ASL, and users circle their answer choices on 1-to-10 Likert scales. The choice of “grammaticality,” “understandability,” and “naturalness” for the subjective questions used in the conducted evaluation studies is based on the use of similar types of questions by researchers who study the automated synthesis of speech audio from text. Such researchers commonly ask hearing users to listen to a computer-synthesized speech recording and to respond to similar sets of Likert scale questions (e.g., “accuracy,” “intelligibility,” and “naturalness”) [32, 33]. As it will be discussed in section 2.4, sometimes the results of subjective evaluation questions and objective comprehension questions do not agree: in prior studies, it has been found that there can be low correlation between a user’s subjective impression of the “understandability” of an ASL animation and his/her success at answering comprehension questions about that animation [30]. Thus, it was decided to include both types of evaluation in the conducted studies.

1.5 Research hypotheses and structure of this article

Section 1.2 explained how human ASL signers use linguistic phenomena like spatial reference and spatial verb inflection, but section 1.3 described how current ASL animation technology does not handle these phenomena. What is unknown is whether the lack of these phenomena has a significant impact on the usability of ASL animations for deaf users. Developing computational linguistic techniques to generate these phenomena is a non-trivial research goal, and before beginning such work, it would be useful to know what benefits there might be. Sections 2 and 3 of this article discuss two evaluation studies conducted to measure the degree to which spatial reference and verb inflection affects the quality of ASL animations – along the criteria outlined above. There are several hypotheses that have been evaluated by these two evaluation studies:

- **H1:** The addition of spatial reference (including pointing pronouns and contrastive role shift) in ASL animations will improve signer's success at answering comprehension questions about those animations.
- **H2:** The addition of spatial reference (including pointing pronouns and contrastive role shift) in ASL animations will lead signers to rate these animations higher on subjective evaluation scales of grammaticality, understandability, and naturalness.
- **H3:** The addition of spatial verb inflection in ASL animations will improve signer's success at answering comprehension questions about those animations.
- **H4:** The addition of spatial verb inflection in ASL animations will lead signers to rate these animations higher on subjective evaluation scales of grammaticality, understandability, and naturalness.

There are several reasons why the above predictions have been divided into four hypotheses. The pointing movements needed to establish spatial reference points and perform pronouns referring to them are not difficult to synthesize from an animation perspective. While there are non-trivial computational linguistic issues to be solved (i.e., when to set up spatial reference points, which entities in a conversation should get spatial reference points, and where to put them in 3D space), the actual hand movements are not complex (i.e., finger pointing and eye-gaze). The production of animations of a virtual human character correctly performing the complex motion paths and hand orientations for inflected ASL verbs are much more difficult to synthesize. In this case, there are both computational linguistic and animation generation research issues to resolve. Thus, H1 and H2 focus on the benefits of spatial reference and pointing pronouns alone, and H3 and H4 focus on the benefits of ASL verb inflection.

Further rationale for the four research hypotheses arises from the alternative ways in which the usability of ASL animations can be measured (as discussed in section 1.4). Hypotheses H1 and H3 focus on whether ASL animations with or without these phenomena lead to differences in native signers' performance on a question-answering task. Hypotheses H2 and H4 focus on whether native signers viewing ASL animations with or without these phenomena are aware of any differences in the animation quality – such that they report those differences on Likert scale subjective evaluation questions.

Sections 2 and 3 will discuss two sets of experimental studies that evaluate the four hypotheses above. Afterward, section 4 will summarize the major conclusions of the conducted work, and section 4.1 will discuss some other contributions. Finally, section 4.2 will describe current research: a multi-year project to use motion-capture technology to collect and analyze a corpus of ASL sentences – to find ways to improve spatial reference and verb inflection in ASL animations.

2 Effect of spatial reference on ASL animations

As discussed in section 1.2, 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. When comparing two sets of entities that have been established in space, signers may also swivel their torso in the direction of one of each of these spatial reference points – while commenting on each. The goal of this study was to understand the degree to which these spatial phenomena in ASL lead to comprehension differences in native ASL signers (H1) or differences in subjective evaluation scores (H2). Therefore, experiments were conducted in which native ASL signers evaluated different types of ASL animations. Some animations contained spatial reference points, pointing pronouns, and contrastive role shift, and other versions of the animations did not. For this study, a set of ASL stories was scripted, animations of each were produced, surveys and questionnaires were designed, and experimental sessions were conducted with deaf research participants.

2.1 ASL animations in the spatial reference study

In the context of the experiment, 10 paragraph-length stories in ASL were designed and computer animations of them were scripted using Sign Smith Studio [9], the ASL animation software described in section 1.1. This software allows users to script ASL performances using a dictionary of signs, fingerspelling, eye-gaze, and facial expressions. Details regarding the creation of ASL animations for experiments using this tool and the motivation for choosing this software for the experimental studies are discussed in [29], which presents an earlier set of evaluation studies conducted on the speed and timing of ASL animations. The ASL stories produced for the study were on a variety of topics: short news stories, adaptations of encyclopedia articles, fictional narratives, and personal introductions. The stories contained sentences of a variety of lengths and complexity; some included topicalized noun phrases, condition/when clauses before a main clause, or rhetorical questions (which are common in ASL). The stories were an average of 119 signs in length. Fig. 2 contains screenshots of the ASL signing character used in the animations.

Each story was produced in two versions. In the *space* version of each story, the virtual human character occasionally performs contrastive role shift (tilting shoulders left or right to contrast two concepts) and also uses pointing pronouns to associate some of the entities under discussion with spatial reference points. In the *no-space* version of each story, the contrastive role shift was removed, and no spatial reference points were established.

(a.) *(facing-front):* TWO STUDENTS NORTHWEST UNIVERSITY. NAMED #LARRY #ATKEN POINT(right) AND #DAVID #SLOTE POINT(left). *(facing-right):* POINT(right) FROM ILLINOIS. STUDY COMPUTER SCIENCE. *(facing-left):* POINT(left) FROM KENTUCKY. STUDY MATHEMATICS. *(facing-front):* POINT(left) AND POINT(right) MAKE COMPUTER PROGRAM NAMED #CHESS. PROGRAM PLAY GAME CHESS. USE #SUPER COMPUTER. GO COMPUTER CHESS CONTEST, VARIOUS ALL-OVER AMERICA. THAT COMPUTER PROGRAM WIN ALL CONTEST. BOTH STUDENTS INTERVIEW NEWSPAPER FINISH JUST-NOW COMPUTER CHESS WIN. *(facing-right):* POINT(right) TRY MAKE COMPUTER PREDICT TRAFFIC. POINT(right) SKILL LOGIC AND TEACH COMPUTER THINK. *(facing-left):* POINT(left) GOOD CHESS PLAYER. POINT(left) PLAY UNIVERSITY CHESS TEAM. POINT(left) ALSO UNIVERSITY TENNIS TEAM. POINT(left) WILL WANT WORK ANALYZER INSURANCE COMPANY.

(b.) *(facing-front):* TWO STUDENTS NORTHWEST UNIVERSITY. NAMED #LARRY #ATKEN AND #DAVID #SLOTE. #LARRY FROM ILLINOIS. STUDY COMPUTER SCIENCE. #DAVID FROM KENTUCKY. STUDY MATHEMATICS. #DAVID AND #LARRY MAKE COMPUTER PROGRAM NAMED #CHESS. PROGRAM PLAY GAME CHESS. USE #SUPER COMPUTER. GO COMPUTER CHESS CONTEST, VARIOUS ALL-OVER AMERICA. THAT COMPUTER PROGRAM WIN ALL CONTEST. BOTH STUDENTS INTERVIEW NEWSPAPER FINISH JUST-NOW COMPUTER CHESS WIN. #LARRY TRY MAKE COMPUTER PREDICT TRAFFIC. #LARRY SKILL LOGIC AND TEACH COMPUTER THINK. #DAVID GOOD CHESS PLAYER. #DAVID PLAY UNIVERSITY CHESS TEAM. #DAVID ALSO UNIVERSITY TENNIS TEAM. #DAVID WILL WANT WORK ANALYZER INSURANCE COMPANY.

Fig. 4. Gloss transcripts of two different versions of the same ASL story. In **a** on the left, the signer associates locations in the signing space with characters in the story (using pronouns like "POINT(left)") and uses contrastive role shift (twisting the waist to aim the torso left or right). In **b** on the right, the signer does not associate characters with locations in space.

Figure 4 contains transcripts of two different versions of same story. Figure 4(a) is the version that uses 3D space to represent entities under discussion, and Fig. 4(b) is the version without this use of space. In Fig. 4(a), the signer uses the pronoun "POINT(left)" to establish a spatial reference point for the character "David." During the "POINT(left)" sign, the signer points to a location on the left side while aiming his eye-gaze and head-tilt at that location. Later, when the signer needs to refer to David, he will again point to this location to refer to him. (The signer also uses "POINT(right)" to set up a different location for "Larry.") In Fig. 4(b), the signer instead fingerspells the full name "David" (written in the transcript as "#DAVID") each time that he refers to David. In this version of the animation, David is not associated with a 3D location in the signing space.

While the glosses "POINT(left)" and "POINT(right)" are used in Fig. 4, it is important to note that signers can set up more than two spatial reference points for a story, and they may choose arbitrary locations in the signing space for each. They are not limited to "left" and "right." The transcripts in Fig. 4 also include annotation of the form "(facing-front)," "(facing-left)," and "(facing-right)" to indicate portions of the story during which the signer is performing contrastive role shift. Fig. 5 contains an English translation of the story.

There are two students from Northwest University named Larry Atken and David Slote. Larry is from Illinois and studies computer science. David is from Kentucky and studies math. They both made a computer program called "CHESS"; the program plays the game of chess. It uses a supercomputer. They have travelled to computer chess competitions all over America. The program has won all of the competitions. They were interviewed for a newspaper recently after a chess victory. Larry would like to create software to predict traffic. He's good at logic and at making computers think. David is a good chess player. He is on the university chess team, and he's also on the university tennis team. In the future, he wants to work as an analyst at an insurance company.

Fig. 5. Approximate English translation of the story shown in Fig. 4.

2.2 Participants and design of the spatial reference study

As part of the recruitment process, online advertisements were placed on local deaf community Web sites; prospective participants were asked whether: (1) they had grown up in a household in which ASL was used as a primary language or (2) if they had attended an ASL-based school as a young child. Section 1.4 has described why it is important to screen for "native signers" for ASL animation evaluation studies. For this study, 5 women and 2 men were recruited; the average age of participants was 37. Section 1.4 has discussed how important it is to create an ASL-centered environment (with few English influences) when conducting experimental evaluations of ASL animations with native signers; so, a native ASL signer conducted the experimental sessions in this study.

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 5 animations of each version (*space* vs. *no-space*). After viewing each story, participants answered 1-to-10 Likert scale questions about the grammaticality, understandability, and naturalness of the animations. These three subjective questions were discussed previously in section 1.4.

To measure participants' comprehension of the information content of the animations, a set of multiple-choice questions were used. The Sign Smith Studio software was used to produce a set of four animations containing comprehension questions – about basic facts that were specific to each of the 10 ASL stories. After asking each comprehension question, the signer would give a list of possible answers to the question. Fig. 6 contains a set of "gloss" transcripts of the sequence of signs for four of the comprehension questions used in the study. These questions correspond to the sample story shown in Fig. 4. While participants were allowed to replay the questions multiple times, they were not allowed to replay the ASL story. So, the comprehension questions not only tested comprehension – but also the signer's memory of the information content.

WHERE #LARRY POINT(right) GROW-UP?
    KANSAS. KENTUCKY. TENNESSEE. ILLINOIS. IOWA.
WHAT #DAVID POINT(left) STUDY UNIVERSITY?
    COMPUTER SCIENCE. MATH. TENNIS. FINANCE.
WHERE FUTURE #DAVID POINT(left) WANT WORK?
    INSURANCE COMPANY. TENNIS. TRAFFIC CONTROL. COMPUTER. CHESS.
WHICH STUDENT PLAY UNIVERSITY TENNIS?
    #LARRY POINT(right). #DAVID POINT(left).

Fig. 6. Transcripts of the comprehension questions and answer choices for stories in Fig. 4.

On the paper form on which participants were marking their responses, there appeared a set of clip-art images that corresponded to the answer choices for each question. The participant was asked to circle the appropriate clip-art picture to select their response. Clip-art images were used so that participants with limited English reading literacy would be able to participate in the study without having to read a list of answer choices on their paper response form. Fig. 7 contains sample of the clip-art pictures used for one of those questions.

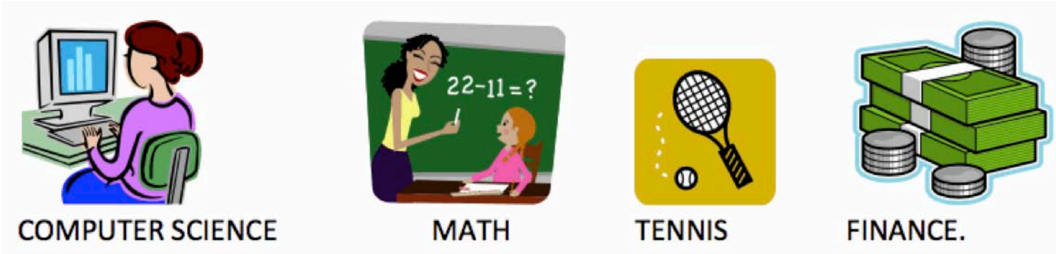

Fig. 7. Clip-art images appearing on the response form for the second question in Fig. 6.

The Spatial Reference study described above is a subset of a larger study described in [1]. The ASL stories in the Spatial Reference study were a subset of those shown to participants in the larger study. In the larger study, animations appeared in two versions: (1) some with spatial reference points and pointing pronouns (some of these animations contained contrastive role shift and some did not) and (2) animations in which the signer does not use any spatial reference points or pointing pronouns. In the Spatial Reference study (presented in this article), all of the *space* animations contained use of contrastive role shift.

2.3 Results of the spatial reference study

Fig. 8 displays the average comprehension question scores for participants in the study for the *space* vs. *no-space* animations. The overall score of the comprehension question responses for each story was calculated subtracting the number of correctly circled choices minus 25% of any incorrect circles or missing circles. There were an average of five multiple-choice options for each question; so, random guessing would produce a score of 0%.

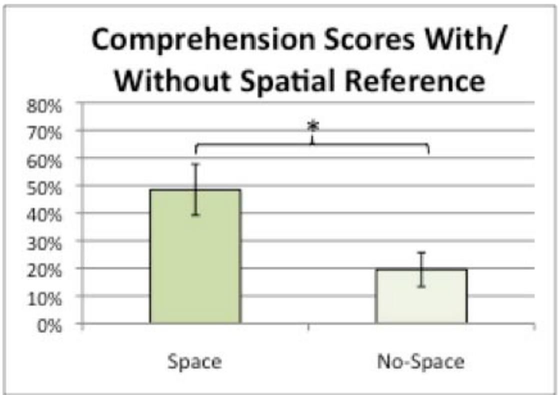

Fig. 8. Average comprehension question accuracy scores for Spatial Reference study.

As shown in Fig. 8, 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). In Fig. 8, Fig. 9, Fig. 13, and Fig. 14, statistically significant pairwise differences are marked with an asterisk. The error bars displayed in these graphs indicate the standard error of the mean for each group of data.

Fig. 9 shows the participants' average response scores for the 1-to-10 Likert scale subjective questions they were asked in the study. The difference between the *space* and *no-space* groups is statistically significant only for the "naturalness of movement" category; participants responded that the *space* animations appeared more natural. While the differences between the other scores were not significantly different, the grammatical correctness and understandability scores were slightly higher for the *space* animations.

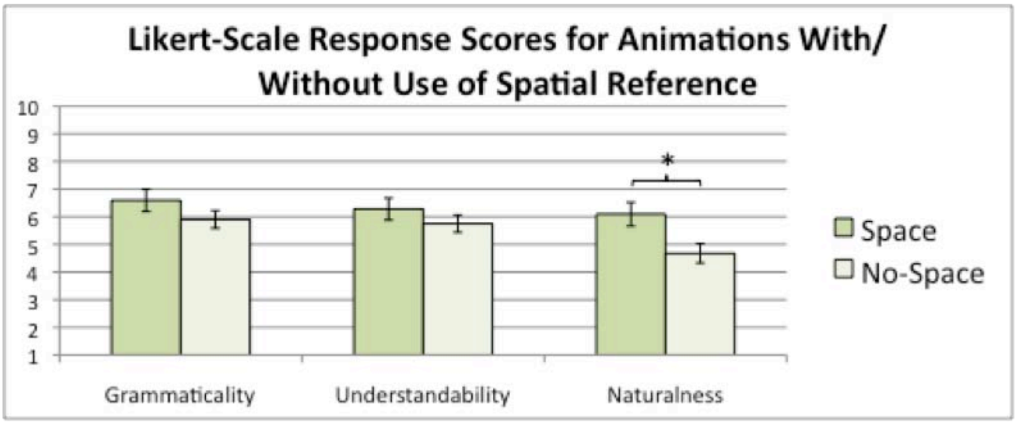

Fig. 9. Average scores for grammatical correctness, understandability, and naturalness of movement as reported on 1-to-10 Likert scales by participants in the Spatial Reference study.

2.4 Discussion of spatial reference study

It was discussed earlier how 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 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 experiment discussed above has demonstrated that there is a significant difference in deaf users' comprehension of ASL animations when they include spatial reference and contrastive role shift behaviors. This result empirically supports one hypothesis (H1) and only partially supports another hypothesis (H2) discussed in section 1.5.

Hypothesis H2 is only partially supported because a statistically significant difference was observed only for the "naturalness" Likert scale question. Considering the significant difference observed in the percentage of correct responses to the comprehension questions, 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.

There are several possible explanations for the comprehension benefit observed in the participants' success at the multiple-choice comprehension questions. One 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. Thus, the improved quality of the animation allowed for better understanding. An alternative explanation for these results is that the use of space facilitated the participants' *remembering* the information from the story; associating information content with locations in space may allow viewers to more easily encode or chunk information, which improved their recall of this information when later answering the comprehension questions (participants were not allowed to replay the ASL stories to look for the answers to the comprehension questions). The use of contrastive role shift may have emphasized the association of elements of information with the spatial reference points on opposite sides of the signer's body. This spatial association may have further aided participants' recall of elements of information as they were answering the comprehension questions.

3 Effect of verb inflection on ASL animations

The above study has demonstrated that there are significant benefits to ASL animations in which the virtual human character associates entities under discussion with spatial reference points in the signing space, uses pointing pronoun signs, and uses contrastive role shift. As discussed in section 1.2, there are other ASL phenomena that interact with the spatial reference points established around the signer. Specifically, some ASL verbs also change their

Author's Copy. Official version available online at http://dx.doi.org/10.1007/s10209-011-0247-7

movement or hand orientation based on the 3D locations in the signing space associated with their subject or object. To determine the importance of the use of these "spatially inflected verbs" to the overall understandability and quality of ASL animations, another study was conducted in which native signers evaluated two versions of ASL animations: some with spatially inflected verbs and some in which the standard dictionary form of each verb was used (in which the motion path has not been spatially inflected to indicate the location of subject and object).

3.1 ASL animations in the verb inflection study

In the second study, a new set of 10 stories was produced using the VCom3D Sign Smith Studio software; the stories consisted of similar genres to those in the prior Spatial Reference study: short narratives, short encyclopedia-style passages, etc. For this study, the average length of each story was 53 signs. The stories were engineered so that each included a set of 3-5 main characters, each of which was established at a different spatial reference point around the signer. In several of the stories, the virtual human character introduces himself as "Charlie," and in some of the stories, "Charlie" is a participant in the events described. Each story was also engineered, so that it included several ASL verbs that can be spatially inflected (as described in section 1.2). Fig. 10 shows an example of an ASL story from this Verb Inflection study (with an English translation provided). In this example, three spatial reference points were established around the signer; these are indicated by the numbers "1," "2," and "3" in Fig. 10(a). Inflected verb forms with a motion path moving from one of these locations to another are indicated with an arrow; for example, "ASK(3→2)" is an inflected ASL verb "ASK" whose motion path goes from location "3" to location "2."

(a.) HI. ME #CHARLIE. ME HAVE THREE BROTHER SISTER. #JOHN POINT(2), #SALLY POINT(1), #BILL POINT(3). POINT(2) WIN #LOTTERY 1000 DOLLARS. SOME MONEY, POINT(2) LOAN(2→1) POINT(1). POINT(1) ASK(1→3) POINT(3), WHERE BANK. POINT(3) DON'T-KNOW. POINT-10 ASK(3→2) POINT(2), WHERE BANK. POINT(2) INFORM(2→3) POINT(3), BANK ON EIGHTH STREET. POINT(1) INFORM(1→2) POINT(2), BANK CLOSED. MONEY, POINT(1) GIVE(1→3) POINT(3). POINT(3) NEVER USE BANK.

(b.) Hi! My name is Charlie. I have three siblings: John, Sally, and Bill. John won $1000 in the lottery, and he loaned some money to Sally. She asked Bill where a bank was. He didn't know. Bill asked John where a bank was. John informed him that there's a bank on Eighth Street. Sally informed John that that bank was closed. Sally gave the money to Bill. He never uses banks.

Fig 10. Story used in the verb inflection study: **a** contains a transcript of the sequence of ASL signs in the story; **b** contains an approximate English translation of the story.

Each of the stories was produced in two versions: *inflected* and *uninflected*. In the *inflected* version of each story, all of the ASL verbs whose motion path can be modified to indicate the location of their subject/object were inflected. While the VCom3D Sign Smith Studio software does not have the ability to automatically inflect ASL verb forms for arbitrary locations of spatial reference points in the signing space, the VCom3D company produces a complimentary piece of software called Gesture Builder. Using this program, a human animator is able to manually specify the handshape, hand location, hand orientation, and motion path of the two hands of the virtual human signing character to produce a wide variety of movements. The result can be saved as a "gesture" computer file, and it can be imported into the VCom3D program as a new "sign." It should be noted that producing novel forms of verb performances for different arrangements of spatial reference points in the 3D signing space is a time-consuming and meticulous process that requires the expertise of a native ASL signer and a careful attention to 3D movements of the virtual human character (see Fig. 11 for an example of two versions of the ASL verb "GIVE" that were created in this manner and imported into the Sign Smith Studio software when producing the *inflected* version of the stories for this study).

Fig. 11. Four still images are shown from animations of two different performances of an inflected ASL verb "GIVE." In **a**, the subject of the verb is associated with a spatial reference point on the signer's right side, and the object is positioned on the signer's left side. In **b**, the subject is positioned on the signer's left, and the object, on the signer's right.

In the *uninflected* version of each story, the ASL verbs were replaced with the "normal" or "dictionary citation" form of the sign. This is the standard version of the performance of an ASL sign that one might expect to see depicted in a sign language dictionary. In this version of the sign, the motion path has not been modified to indicate the location of its subject or object; see Fig. 12 for an example of the uninflected version of the ASL verb sign "GIVE." As discussed in section 1.3, in current ASL scripting or MT systems, it is generally an uninflected version of ASL verbs that are included in the animations produced.

Fig. 12. Two still images (*beginning* and *end*) of an animation of the performance of the ASL verb "GIVE." This is the "dictionary" version of the sign, which appears in the *uninflected* ASL animations in this study. The hand moves forward during the performance of the sign.

While ASL signers can optionally omit the subject and/or object of a sentence if the verb inflection indicates their identity, in both the *inflected* and *uninflected* animations for this study, signs for the subject and the object for every sentence were included; so, even without the verb inflection, it should be possible for an ASL signer to identify the subject and object of each verb.

3.2 Participants and design of the verb inflection study

As in the previous Spatial Reference study, advertisements and screening questions were used to recruit native ASL signers from the local deaf community in New York City. Four women and eight men (average age 29) participated in this study. All had learned ASL prior to age 6, and eight had been raised in an ASL-signing household since birth. As before, a fully factorial within-subjects design was used such that: (1) no participant saw the same story twice, (2) the order of presentation of stories was randomized, and (3) each participant saw five animations of each version (*inflected* vs. *uninflected*). After watching each animation, participants answered Likert scale questions (grammaticality, understandability, naturalness) and multiple-choice comprehension questions (similar to those used in the Spatial Reference study). Comprehension questions asked about basic facts from each story; some questions asked the participant to identify which person in a story performed some action – which was conveyed using a spatially inflected verb in the *inflected* version of the story.

3.3 Results of the verb inflection study

Fig. 13 displays average comprehension question scores for participants in the study for the *inflected* versus *uninflected* animations. As in the earlier Spatial Reference study, the score was the number of correctly circled choices minus 25% of any incorrect circles or missing circles. As shown in Fig. 13, there was a statistically significant difference in comprehension scores between the *inflected* and *uninflected* versions of the animations ($p < 0.05$, Mann Whitney U-Test).

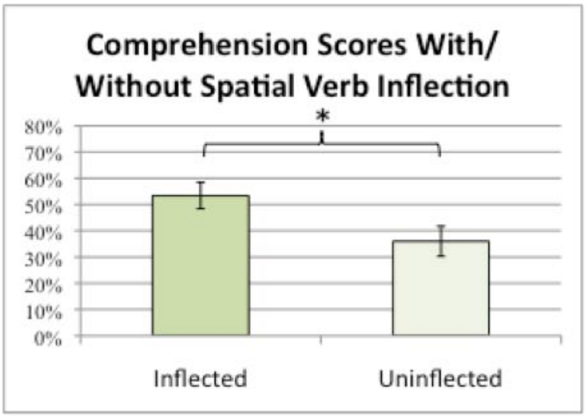

Fig. 13. Average comprehension question accuracy scores for verb inflection study.

Fig. 14 shows participants' average response scores for the 1-to-10 Likert scale subjective questions signers were asked in the study. There was almost no difference observed between the *inflected* and *uninflected* animations; participants assigned similar scores for grammatical correctness, understandability, and naturalness of movement for the space animations in this study.

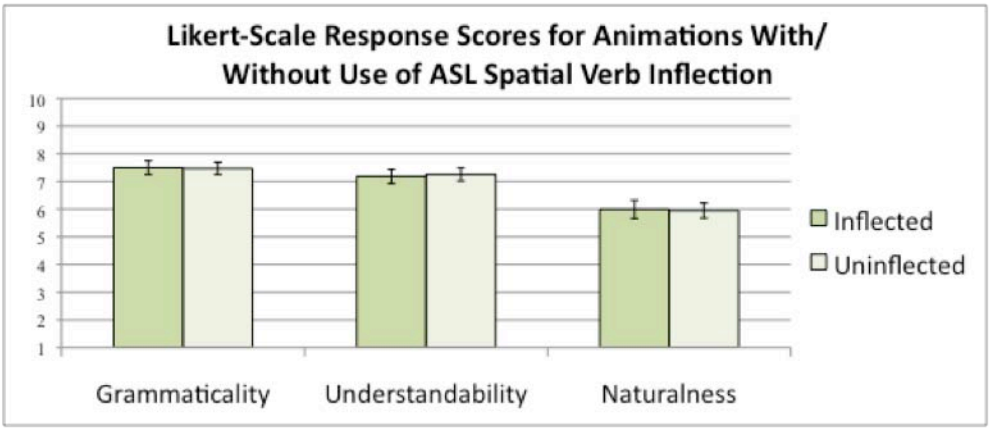

Fig. 14. Average scores for grammatical correctness, understandability, and naturalness of movement as reported on 1-to-10 Likert scales by participants in the verb inflection study.

3.4 Discussion of the verb inflection study

In both the Spatial Reference and Verb Inflection studies, a significant difference was observed in the comprehension question scores; however, in the Verb Inflection study, there were no significant differences in any of the Likert scale subjective evaluation questions. These results support hypotheses (H3), but fail to support hypotheses (H4) from section 1.5.

The participants in this study did not appear to *overtly* notice the impact of the inflected verbs on their comprehension. While they reported nearly identical scores for the grammaticality, understandability, and naturalness of the animations on the subjective Likert scale scores, it was clear that the inclusion of inflected verbs in the animations had a significant impact on their comprehension task success. There are several explanations we have considered for the higher comprehension question scores for the *inflected* animations in this study:

- When viewing animations without verb inflection, signers may have been confused by the performance of a verb that appeared in its standard "dictionary" form. The participant may have tried to interpret the movement direction of the verb as if it *had been* inflected spatially, and this may have conflicted with the subject and object expressed in the sentence. For example, the standard dictionary form of the verb "GIVE" shown in Fig. 12 is similar to how a signer would perform the verb if the subject were the signer himself and the object were the audience, i.e., "I GIVE YOU." If this form of the verb appeared in the middle of a sentence in which the subject and object are other entities in the conversation, then this may have caused confusion for the person viewing the animation.

- The use of space may have enabled the signer to keep the set of entities under discussion better "organized" in their mind while they watched the animation. Thus, while watching the story, they were better able to process and synthesize the information they were being presented with. The use of space around the signer and the way in which the verb motion paths changed allowed the viewer to more easily keep track of "who was doing what to whom" during the story.
- Another possible explanation for participant's higher comprehension question scores is that there is a memory benefit – i.e., when they were later presented with a comprehension question, participants may have recalled the actual visual appearance of how the verb was performed (the direction the verb moved through space). Perhaps this allowed the research participant to more easily recall the subject/object of the verb. Thus, the verb inflection may have facilitated participants' use of visual memory of the performance they had just watched.

Comparing the overall comprehension scores from the Spatial Reference study (section 2) and the Verb Inflection study (section 3), one can notice that the scores in the second study seem higher; however, the comprehension scores between the two studies are not directly comparable. A different set of stories and a different set of corresponding comprehension questions were used in each of these studies; so, it is possible that the differences observed between the two studies arose from the different story/question content. Another possibility is that the specific participants recruited for the Verb Inflection study may have been somewhat more skilled at answering comprehension questions. Thus, intra-study comparisons of the two versions of animations are possible, but inter-study comparisons may not be reliable. The Likert scale response results may be more directly comparable between studies since an identical set of questions was used in both studies; however, the set of participants and the ASL stories differed.

4 Conclusions and future work

Software to generate animations of ASL has the potential to provide important accessibility benefits for deaf users with low English literacy; however, current software cannot produce many ASL signs or constructions. This article has described two experimental studies designed to measure how much the lack of spatial reference points, pointing pronouns, contrastive role shift, or spatial verb inflection affects the usability of current ASL animation technologies.

The Spatial Reference study demonstrated that there are significant benefits to ASL animations in which the virtual human character associates entities under discussion with spatial reference points in the signing space, uses pointing pronoun signs, and uses contrastive role shift. Since current ASL animation generation systems cannot automatically determine which entities in a conversation should be associated with space, when to set up these spatial reference points, and where to position them in 3D space, it is up to a human user of ASL scripting software to make these determinations when producing ASL animations. Further, not all sign language scripting systems actually give the human user enough control of the hand movements, eye-gaze, head-tilt, or torso-movement of the virtual human character to produce fluent-looking phenomena listed above. Based on the results of this evaluation study, it is clear that additional research and development is needed to improve these spatial reference capabilities of sign language animation systems.

The Verb Inflection study focused on a set of ASL phenomena that are more challenging to accurately synthesize in an animation system – the complex movements of ASL verbs whose motion paths and hand orientation are parameterized on the 3D locations of their subject and object. While these ASL phenomena also require spatial reference points to have been established in space, they require mathematical models of the motion paths of individual ASL verbs. The pointing pronoun and contrastive role shift phenomena examined in the earlier study were relatively more straightforward to synthesize as virtual human animations. This study demonstrated that the inclusion of properly inflected ASL verb forms produced ASL animations that were more understandable to native signers (as measured by comprehension questions).

4.1 Other contributions of this research

While differences in comprehension question scores were observed in both the Spatial Reference and Verb Inflection studies, the addition of these spatial phenomena did not appear to draw the overt attention of users (as measured by their response to Likert scale subjective evaluation judgments). This is an important finding because it indicates that researchers studying ASL animation technologies cannot merely rely on their users to articulate for them which aspects of the animation need improvement. Some enhancements to ASL animations go unnoticed by native signers, but they yield significant usability benefits.

Specifically, it is important for ASL animation researchers to employ comprehension questions in studies that evaluate animations of sign language. If the two studies presented in this article had only included answers to the Likert scale evaluation questions , then differences between the animations might have been missed. In prior studies on ASL animations [29, 30], it was noticed that there is often no significant correlation between a participant's perception of how well they have understood an animation and their actual performance on a comprehension question about that animation. When conducting evaluations of ASL animations, Likert scale subjective questions alone are not sufficient for measuring subtle differences in the quality of the animations. Task-based performance measures are necessary.

In addition to validating the usefulness of comprehension questions in evaluations of ASL animations, another contribution of this research is the refinement of the adopted protocols for recruiting/screening native ASL signers, conducting studies in which signers evaluate ASL animations and answer questions, and using of multiple-choice questions with clip-art images presented as answer choices. These protocols will be useful in future research (see section 4.2 below). Further, a bank of over 20 paragraph-length stories in ASL has been designed during the two studies described in this article, and computer animations and corresponding comprehension questions for each have been produced. These stories are useful for testing alternative settings for improving the understandability and naturalness of ASL animations. Having a library of ASL story scripts (which can be easily modified, and then ASL animations re-synthesized) is an important resource for this research. These studies have also established baseline comprehension scores for these stories and questions.

4.2 Current and future work

After measuring the importance of spatial reference and verb inflection through experimental studies, the current research focus of the authors is on developing ASL animation technologies that can generate these phenomena automatically. This research is being pursued in a data-driven manner, beginning by collecting recordings of sign language sentences performed by human signers. These ASL motion-capture data are collected using a body suit with sensors, a head-tracker, and eye-tracker, and motion-capture gloves, and this corpus is being annotated with linguistic features related to the establishment of spatial reference points. Details of the collected types of ASL sentences , the used motion-capture equipment, and the added types of linguistic information are described in [34]. Previous researchers have collected video-based sign language corpora [20, 35-37] – some annotating how signers make use of spatial reference [23]; other researchers have collected short samples of sign language using motion-capture [38-40]. Instead, the corpus under development records multi-sentence passages, in which human signers set up several spatial reference points.

The collected motion-capture data will be analyzed using machine learning techniques to learn when/where human signers establish spatial reference points. A variety of ASL inflected verb performances are also being collected in order to produce animation software that can automatically interpolate an appropriate motion path for the virtual human character's hands during ASL verbs that are inflected for various 3D locations of the verb's subject or object. This new ASL animation software will be evaluated through studies with native ASL signers (using a methodology similar to that described in sections 2 and 3) to guide the development of future research.

The motion-capture data collection started in the summer of 2009, and so far 58 ASL passages from 6 unique signers (about 40 min of annotated ASL motion-capture data) have been collected and linguistically annotated. A team of native ASL signers at the laboratory has labeled the signs, the part-of-speech of each sign, and other essential grammatical information on a timeline synchronized to the motion-capture recordings. As discussed in section 1, other sign languages internationally have similar linguistic phenomena to ASL and may also benefit from research on the use of space in ASL animations. The long-term goal is to learn how ASL signers set up spatial reference points and inflect ASL verbs. The experimental studies presented in sections 2 and 3 suggest that these improvements in ASL technology will produce animations that are more understandable – and, ultimately, more usable – for deaf signers in accessibility applications. In addition to planned work on spatial reference and verb inflection, this ASL motion-capture corpus will enable exciting new research in ASL generation and recognition of ASL from human-recorded movements.

Acknowledgements. This research was supported in part by the US National Science Foundation under award number 0746556, by The City University of New York PSC-CUNY Research Award Program, by Siemens A&D UGS PLM Software through a Go PLM Academic Grant, and by Visage Technologies AB through a free academic license for character animation software. Jonathan Lamberton prepared experimental materials and organized data collection for the ASL animation studies discussed in sections 2 and 3.

References

[1] Huenerfauth M (2009) Improving spatial reference in American Sign Language animation through data collection from native ASL signers. In: *Proceedings of the Universal Access in Human Computer Interaction conference (UAHCI'09)*, pp 530-539. doi: 10.1007/978-3-642-02713-0_56

[2] Mitchell R, Young TA, Bachleda B, Karchmer MA (2006) How Many People Use ASL in the United States? Why Estimates Need Updating. *Sign Language Studies* 6(4):306-335

[3] Traxler C (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. doi: 10.1093/deafed/5.4.337

[4] Huenerfauth M, Hanson VL (2009) Sign Language in the Interface: Access for Deaf Signers. In: Stephanidis C (ed.) *The Universal Access Handbook*. Lawrence Erlbaum Associates, Mahwah, NJ

[5] Lane H, Hoffmeister R, Bahan B (1996) *A Journey into the Deaf World*. DawnSign Press, San Diego

[6] Padden C, Humphries T (2005) *Inside Deaf Culture*. Harvard University Press, Cambridge, MA

[7] Elliott R, Glauert JRW, Kennaway JR, Marshall I, Safar E (2006) Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society* 6(4):375-391. doi: 10.1007/s10209-007-0102-z

[8] Kennaway J, Glauert J, Zwitserlood I (2007) Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction* 14(3):1-29. doi: 10.1145/1279700.1279705

[9] VCom3D. (2010) Sign Smith Studio. http://www.vcom3d.com/signsmith.php. Accessed 11 March 2010

[10] Chiu YH, Wu CH, Su HY, Cheng CJ (2007) Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 29(1):28-39. IEEE Press, New York. doi: 10.1109/TPAMI.2007.15

[11] Fotinea SE, Efthimiou E, Caridakis G, Karpouzis K (2008) A knowledge-based sign synthesis architecture. *Universal Access in the Information Society* 6(4):405-418. doi: 10.1007/s10209-007-0094-8

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

[13] Karpouzis K, Caridakis G, Fotinea SE, Efthimiou E (2007) Educational resources and implementation of a Greek sign language synthesis architecture. *Computers & Education* 49(1):54-74. doi: 10.1016/j.compedu.2005.06.004

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

[15] Morrissey S, Way A (2005) An example-based approach to translating sign language. In: *Proceedings of the Workshop on Example-Based Machine Translation*, pp 109-116.

[16] Shionome T, Kamata K, Yamamoto H, Fischer S (2005) Effects of display size on perception of Japanese sign language - Mobile access in signed language. In: *Proceedings of the Human-Computer Interaction Conference*, pp 22-27

[17] Sumihiro K, Yoshihisa S, Takao K (2000) Synthesis of sign animation with facial expression and its effects on understanding of sign language. *IEIC Technical Report* 100(331):31-36

[18] van Zijl L, Barker D (2003) South African sign language MT system. In: *Proceedings of AFRIGRAPH*, pp 49-52. doi: 10.1145/602330.602339

[19] Zhao L, Kipper K, Schuler W, Vogler C, Badler N, Palmer M (2000) A machine translation system from English to American Sign Language. In: *Proceedings of the 4th conference of the association for machine translation in the Americas on envisioning machine translation in the information future, Lecture Notes in Computer Science* 1934, Springer, Heidelberg, pp. 54-67. doi: 10.1007/3-540-39965-8_6

[20] Neidle C, Kegl J, MacLaughlin D, Bahan B, Lee R (2000) *The syntax of American Sign Language: functional categories and hierarchical structure*. MIT Press, Cambridge.

[21] Liddell S (2003) *Grammar gesture and meaning in American Sign Language*. Cambridge University Press, Cambridge

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

[23] Braffort, A, Dalle, P (2008) Sign language applications: preliminary modeling. *Universal Access in the Information Society* 6(4):393-404. doi: 10.1007/s10209-007-0103-y

[24] Marshall I, Safar E (2003) A Prototype Text to British Sign Language (BSL) Translation System. In: *Companion Volume to the Proceedings of the Association for Computational Linguistics Conference*, pp 113-116. doi: 10.3115/1075178.1075194

[25] Iwarsson S, Stahl A (2003) Accessibility, usability and universal design-positioning and definition of concepts describing person-environment relationships. *Disability and Rehabilitation* 25(2):57-66. doi: 10.1080/0963828021000007969

[26] Nielsen J (1993) *Usability Engineering*. Academic Press, Boston.

[27] International Organization for Standardization (1998) *ISO 9241-11: Guidance on Usability*. International Organization for Standardization. http://www.iso.org/iso/en/CatalogueListPage

[28] Huenerfauth M (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*, pp. 129-136, ACM Press, New York. doi:10.1145/1530064.1530067

[29] Huenerfauth M (2009) A Linguistically Motivated Model for Speed and Pausing in Animations of American Sign Language. *ACM Transactions on Accessible Computing* 2(2):1-31. ACM Press, New York. doi: 10.1145/1530064.1530067

[30] Huenerfauth M, Zhao L, Gu E, Allbeck J (2008) Evaluation of American Sign Language generation by native ASL signers. *ACM Transactions on Accessible Computing* 1(1):1-27. ACM Press, New York. doi: 10.1145/1361203.1361206

[31] Lucas C, Valli C (1992) *Language Contact in the American Deaf Community*. Academic Press, San Diego, CA

[32] Campbell N (2005) Speech Synthesis Evaluation, In: *Human Language Technologies (HLT) Evaluation Workshop, European Language Resources Association (ELRA),* http://www.elra.info/hltevaluationworkshop/img/pdf/Nick%20Campbell.ATR.Speech%20Synthesis%20Evaluation.pdf. Accessed 13 March 2010

[33] Van Bezooijen R, Pols L (1990) Evaluating text-to-speech systems: Some methodological aspects*, Speech Communication* 9(4):263-270, doi: 10.1016/0167-6393(90)90002-Q

[34] Huenerfauth M, Lu P (2010) Annotating spatial reference in a motion-capture corpus of American sign language discourse. In: *Proceedings of the Fourth Workshop on the Representation and Processing of Signed Languages: Corpora and Sign Language Technologies, The 7th International Conference on Language Resources and Evaluation (LREC 2010).* ELRA, Paris

[35] Bungeroth J, Stein D, Dreuw P, Zahedi M, Ney H (2006) 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

[36] Crasborn O, Sloetjes H, Auer E, Wittenburg P (2006) 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, The 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 82-87. ELRA, Paris

[37] Efthimiou E, Fotinea SE (2007) GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. In: *Universal Access in Human Computer Interaction*. *Lecture Notes in Computer Science* 4554, pp. 657-666. Springer, Heidelberg

[38] Brashear H, Starner T, Lukowicz P, Junker H (2003) Using multiple sensors for mobile sign language recognition. *IEEE International Symposium on Wearable Computers*, pp 45, IEEE Press, New York. doi: 10.1109/ISWC.2003.1241392

[39] Cox S, Lincoln M, Tryggvason J, Nakisa M, Wells M, Tutt M, Abbott S (2002) Tessa, a system to aid communication with deaf people. In: *5th International ACM Conference on Assistive Technologies*, pp. 205-212. ACM Press, New York. doi: 10.1145/638249.638287

[40] Vogler C, Metaxas D (2004) Handshapes and movements: Multiple-channel ASL recognition. *Lecture Notes in Artificial Intelligence* 2915, pp. 247-258, Springer-Verlag, Heidelberg. doi: 10.1007/11678816