

Linguistics as structure in computer animation: Toward a more effective synthesis of brow motion in American Sign Language

Rosalee Wolfe, DePaul University
Peter Cook, Columbia College Chicago
John C. McDonald, DePaul University
Jerry Schnepf, DePaul University

Abstract

Computer-generated three-dimensional animation holds great promise for synthesizing utterances in American Sign Language (ASL) that are not only grammatical, but well-tolerated by members of the Deaf community. Unfortunately, animation poses several challenges stemming from the necessity of grappling with massive amounts of data. However, the linguistics of ASL can aid in surmounting the challenge by providing structure and rules for organizing animation data. An exploration of the linguistic and extra linguistic behavior of the brows from an animator's viewpoint yields a new approach for synthesizing nonmanuals that differs from the conventional animation of anatomy and instead offers a different approach for animating the effects of interacting levels of linguistic function. Results of formal testing with Deaf users have indicated that this is a promising approach.

Computer animation, nonmanual signals, brows, American Sign Language

Background

A system to synthesize animations of American Sign Language (ASL) as three-dimensional (3D) animation would benefit efforts in automatic translation and interpreter education. For automatic translation, it could serve as the target for conversion of spoken or written English into ASL. At present, no such target exists, which precludes the development of an automated translator. An animation system capable of producing ASL would help overcome this. Although the prospect of a general translator from English to ASL is dim for the foreseeable future, systems for short interactions, where the conversation is highly predictable, are a current possibility (Cox 2002) (Furst 2002).

A sign synthesis system would also benefit interpreting students. For hearing students of sign language, recognition skills lag far behind production skills (Poor 1986). A system that allows students to view and identify signs and signed sentences would provide additional practice opportunities for those wishing to improve their recognition skills (Davidson 2005).

The top-level priority of a sign synthesis system is that it produce animations that are not only grammatical and understandable, but are natural and would be acceptable as examples of good signing. This design goal is critical for two reasons. Imperfections in the production of an utterance distracts from the message carried by the signing. If a viewer has the reaction, “Oh, that’s strange – no one would ever sign like that,” the awkwardness is diverting the viewer’s attention away from the message content. Poor visual quality adds stress to the viewer, and the viewer becomes fatigued and frustrated more rapidly (Ciaramello 2007). The situation is analogous to that of a hearing person who encounters a poor speech synthesis system, which can happen in automated phone answering systems.

Synthesizing natural, grammatical signed utterances would also help sign language learners. The animations need to mimic signing produced by actual people as closely as possible so that the students who view them are strengthening their recognition skills. Further, students would be using them for study purposes, the animations must be exemplars of good signing.

Animation basics

Animation is a series of images, called “frames”, shown in quick succession. When the rate of succession approaches a speed of 24 frames per second, viewers perceive the images as motion (Ehrenstein 2003).

Three-dimensional animation of ASL requires the creation of a 3D model, the development of data for posing the model, and a method for automating the transitions between poses (Foley 1990:1049). An artist begins this process by digitally sculpting a human form (Maestri 1996:11). At this point in the process, the form is rigid, as if it were made from stone. The next step is to build an articulated rig, which is a digital skeleton of jointed bones that can rotate to mimic the motion of human bones (Zeltzer 1982). The skeleton fits inside the sculpted form. The sculpture is replaced with a flexible rubber sheet that conforms to the surface of the sculpture (Barr 1984). The artist attaches the rubber sheet to skeleton via strategically-placed digital “muscles” (Magnenat-Thalmann 1988). With this step, the 3D model is complete.

Posing the model requires numeric data to specify the rotations of each joint (Catmull 1972). This data can be stored using a conventional database application. In addition to the joint rotation data, animation requires that the artist to specify the time at which the model strikes each pose. Figure 1 shows selected poses from an animation of the phrase, “Show me your ID and ticket”.



Figure 1: Poses from the animation I-D TICKET GIVE-ME.

Lastly, an interpolation system provides an automated means for creating transitions between the poses (Burtnyk 1976). Figure 2 shows a series of interpolated frames that occur between the highlighted poses in Figure 1. The intermediate frames were computed automatically by means of an interpolation system.



Figure 2: Interpolation between key poses.

Advantages of Animation

At first glance, it might seem that recording a signer with a video camera would be a reasonable approach, but video lacks the flexibility of 3D animation. Language is productive, capable of creating novel utterances, and with video, one is limited to the signs and sentences previously recorded. In contrast, an animation system has the capability for creating new utterances.

Depending on how a sign is used in a sentence, it can change form (Stokoe 1965:282). An animation system can accommodate these changes in ways that current video editing techniques cannot. The production of a sign may change, depending on the preceding and succeeding signs (Liddell 1984). Edited video can display a sequence of signs, but the transitions between signs will not be satisfactory because the motion between signs will contain abrupt changes that are not natural. On the other hand, an

animation system can take into account the previous and subsequent signs and produce a smoothly flowing animation.

Even with state of the art techniques for video editing such as (Chang 2005), the point of view is limited to the original positioning of the camera. In contrast, animation allows positioning of the camera in any location – on the left or right side of the signer, or even from above. These vantage points can be useful for sign language learners (Davidson 2006).

Animation also has the potential for being an addition tool for sign language study. In a testing situation, a researcher can validate an animation approach by presenting animations portraying a linguistic function to members of the Deaf community for feedback. If the animations are acceptable, then the underlying data representation and algorithms are deemed capable of producing depictions of the linguistic function (Sedgwick 2001).

Challenges of Animation

Accompanying the advantages of animation are several challenges. It is time consuming to set up the poses from which the animations are created. Further, each pose requires positioning between 50 and 100 joints, depending on the detail of the model. Each sign requires multiple poses, and each sentence consists of multiple signs.

Working with this large quantity of data is a continual challenge. When awash in a sea of data, it is difficult to manage and manipulate the thousands of numbers. It is not simply a matter of acquiring computational power to work with the data; it is a question of understanding what the data mean and how to make decisions on how to change it to create new sentences in sign language.

Motion capture is a newer technology often used as an alternative to the time-consuming task of hand animation (Lee 2002) (Moeslund 2006). Motion capture records the movement of humans in real-time for later use in 3D animations. Motion capture technologies are either active or passive (Moeslund 2001). In active motion capture, an actor wears a specialized body suit and/or body gloves equipped with internal sensors, which connect to a data cable plugged into a computer. In a passive system, a set of cameras focus on the actor who is typically wearing small, highly visible markers. Specialized software processes the video from the camera to determine the 3D position of each marker. Figure 3 shows two examples of active sensing technology -- wired data gloves (Astrauskas 2008), and a body suit (Kirk 2005).

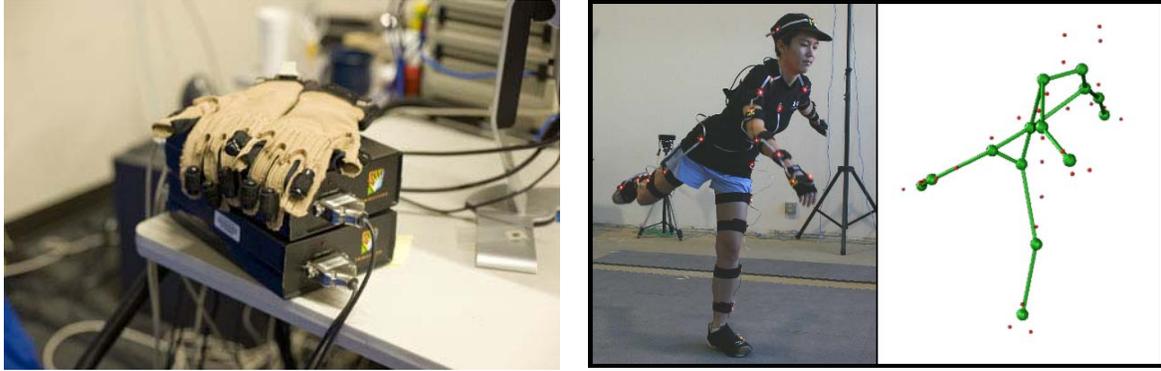


Figure 3: Examples of data gloves and a motion capture suit.

Although the technology can record motion as a person creates it, motion capture has several challenges that need to be surmounted before it becomes feasible for animating sign language. Both accuracy (Kovar 2002) (Miller 2006) and missing data (Aristidou 2008) continue to be a problem. In addition, motion capture generates massive amounts of data. Our current animation approach uses less than 5 keys per joint per second. This compares favorably to typical capture rates of 200 data points per marker per second (Goldman 2000) (Motion Analysis 2009). Due to these attributes, motion capture data requires a large time investment simply to “clean up” the data before it can be used.

The vastness of the data produced by motion capture make it even more difficult to edit than animation. As Kovar, Gleicher & Pighin (2002) observe:

Motion capture data has proven to be difficult to modify, and editing techniques are reliable only for small changes to a motion. This limits the utility of motion capture. If the data on hand isn't sufficiently similar to what is desired, then often there is little that can be done other than acquire more data, a time-consuming and expensive process. This in particular is a problem for applications that require motion to be synthesized dynamically.

Although there is ongoing research in using mocap data to support naturalistic motion synthesis (Lau 2009), motion capture still shares some of the same inflexibility as video recording. In an empirical study that compared the efficacy of motion capture and hand animation for portraying fingerspelling, viewers preferred the fingerspelling produced through hand animation (Sanders 2005). They found it easier to read and more realistic in appearance.

The appeal of motion capture's promise of real-time recording is not to be underestimated and may, in the future, be a viable option. However, the current technology has limited effectiveness as an alternative for driving sign language synthesis.

Linguistics as animation structure

Any system built to synthesize sign language must incorporate more than the conventional considerations required for animation because it must be able to manage and modify massive amounts of data. Linguistics can aid in making sense of the data by providing a framework to organize, codify and index it.

Phonemic information is a powerful resource. There is an intuitive mapping between the phonemes of sign language, and the posing and timing of a digital model. For example, the phonemic element of handshape (Wolfe 1999, McDonald 2000) can provide a means of posing a large number of hand joints in a concise and economic manner. Palm orientation, point of articulation and point of contact are sufficient to calculate the angles to pose a model's arms (McDonald 2002) without the need of an iterative IK solution.

In addition to providing data organization, linguistics provides insights into how to modify the data. Processes that occur in agreement verbs (Toro 2004) and classifier predicates (Huenerfauth 2006) affect the movement of a model's hands and can change the palm orientation. Syntax also affects timing (Huenerfauth 2009). For many simple declarative sentences, considering the phonemic elements of handshape, position, orientation and movement are sufficient to create recognizable and convincing animations. Figure 4 is a schematic diagram of a synthesis system that uses linguistics to create the animations, and Figure 5 is a link to an animation created with this approach.

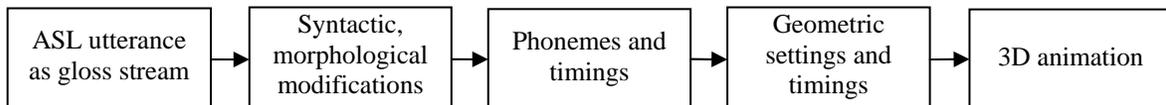


Figure 4: A sign synthesis system where linguistics drives the animation.



Figure 5: Sample animation: HOME YOU
<http://asl.cs.depaul.edu/SLnL/HomeYou.avi>

This technique works well for statements and for interrogatives that require no nonmanuals other than the wh- or yes/no marker, and is similar in approach to linguistically driven animation systems for BSL (Kennaway 2007, Elliot 2008), LSF (Rezzoug 2006) and GSL (Fotinea 2008). However, in this form, it does not have the extensibility to handle co-occurrences of nonmanual signals. We wanted to create a more extensible system, and used the eyebrows as a case study toward this goal.

A Case Study: Brows

Animators are familiar with the seminal work of Ekman & Friesen (1978), which codifies muscle movements of the face. Systems of codified muscle movements form the basis of many commercial products as well as an open standard (Pandzic 2002) to specify facial animation. These help animators in depicting affect and mouth poses.



Figure 6: Basic poses of anger, disgust, fear, sadness, happiness, and surprise which are often employed in conventional computer animation.

In fact, early research by Baker-Shenk (1983) made use of Ekman's Facial Action Coding System (FACS) to characterize brow position for wh- and yes/no questions. However, Weast (2008) notes that Ekman's system lacks the precision to characterize subtleties of brow behavior in ASL utterances. Understanding this behavior is crucial to the successful synthetic production of ASL involving the brows. The following is a review of linguistic literature from the perspective of an animator wishing to synthesize appropriate brow usage.

Early Linguistic Findings

In early efforts, researchers established that grammatical nonmanual signals differed from displays of affect (Baker & Padden 1978, Liddell 1978). Baker-Shenk showed that neutral brows occur with declarative sentences, lowered brows with wh-questions and raised brows with yes/no questions. Coulter (1979) first investigated each feature of

facial anatomy and explored their individual linguistic function rather than considering the face as a whole.

Investigation of the linguistic function of the brows ensued. Several researchers (Liddell 1986, Reilly, McIntire & Bellugi 1990) established that brows can contribute to the formation of a conditional. As part of the nonmanual markers used to distinguish a conditional, a signer maintains raised brows throughout the production of the constituent and marks its completion with an abrupt drop to neutral. Aarons (1994) and Boster (1996) observed that raised eyebrows can occur with sentence-initial topics, and analyzed the spread of wh-related nonmanual markers in their presence, as seen in (1).

_____br
_____bl
 BOOK YOU WANT WH-MANY (1)
 ‘How many books do you want?’

A Starting Point for Animation

For an animator, these results give an excellent characterization of the scope of nonneutral brow position as a syntactic indicator. From these findings, we can make an initial planning document or “dope sheet” for animation. Figure 7 shows a timeline with the manual glosses on the bottom track. A track in a dope sheet is analogous to a tier in the Elan annotation software (Crasborn et al 2006). The top track shows the brow position. Although this is a planning diagram for an animation, there are similarities with the symbols used to transcribe syntactic processes in ASL.

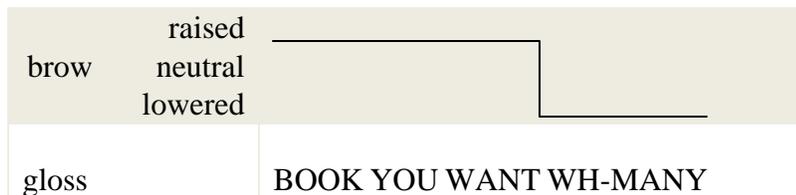


Figure 7: A two-track dope sheet including the scope of brow positions.

This animation notation shows scope and co-occurrence of brow positions, but does not include any information about intensity. Either the brow is completely raised, completely lowered or in a neutral position. Further, there is no information about the transitions among the three choices of brow position.

The role of brow motion in language

Subsequent studies examined the intensity of nonneutral brows. Bahan (1996) and Neidle et al. (2000) noted that the intensity of brow lowering decreases as a function of distance from the +wh feature. Contemporaneously, several researchers (Reilly, McIntire & Bellugi 1990, Wilbur 1991, Sandler & Lillo-Martin 2006:253) observed a prosodic

role for brows that has parallels to intonation in spoken languages. These studies also noted and explored changes of intensity over time.

Wilbur (2000, 2003) united the observations on roles of nonmanual signals in syntactic and prosodic processes in a theory of nonmanual layering in which multiple language functions can co-occur within the nonmanual channels. In this theory, three key design features help nonmanuals function separately.

Two of the features specifically address timing and coordination of nonmanual channels. The first distinguishes paralinguistic or affective facial expression from linguistic facial expressions. The appearance of a linguistic marker begins abruptly and concurrently with a given sign, while the appearance of an affective facial marker can begin gradually and may precede the sign. The second design feature is the abrupt and precise start and end timings of a nonmanual marker in coordination with a particular sign or constituent phrase. These findings give animators valuable help in disambiguating the lowered brows of a negative affect from the lowered brows asking a wh-question. The third design feature of Wilbur's theory considers the spread of potential nonmanual channels throughout the face, head and upper body. This permits independent channels to function separately.

Refining the animation model

Other resources that provide useful insights for an animator include previously-annotated utterances using SignStream (Neidle 2001) or Elan. Such software facilitates the annotation of manual and nonmanual markers that are synchronized with a digitized video that displays the utterance. In SignStream, brow poses are recorded as qualitative descriptions. Grossman's work (2001, Grossman & Kegl 2006) extended the annotations of SignStream by quantifying the descriptive labels, and using the resulting numeric quantities to plot the changes of brow position as a function of time. This created a contour portraying brow movement as seen in Figure 8. The contour visualizations facilitated analysis of dynamic changes in brow position.

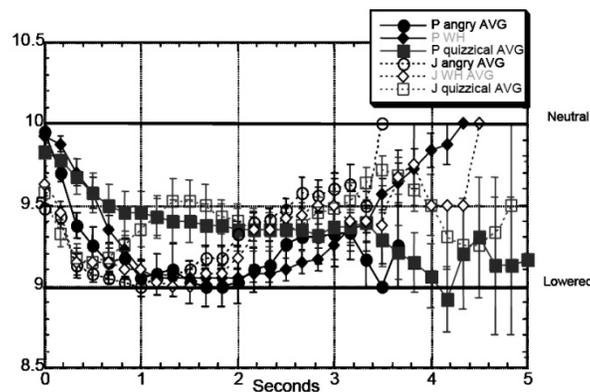


Figure 8: Brow level for two signers: angry, quizzical, and wh-question (Grossman & Kegl 2006)

Grossman & Kegl examined brow usage as part of their research on nonmanual productions. The productions included wh- and yes/no-questions, declarative emotional statements, and non-emotional, non-grammatical expressions. They found that in general, both surprise statements and yes/no questions had raised eyebrows while angry statements and wh-questions had lowered eyebrows.

The information captured by Grossman and Kegl includes a rich set of timing information, which is useful for refining the dope sheet first sketched in Figure 7. Figure 9 shows a proposed refinement, based on Grossman and Kegl’s findings.

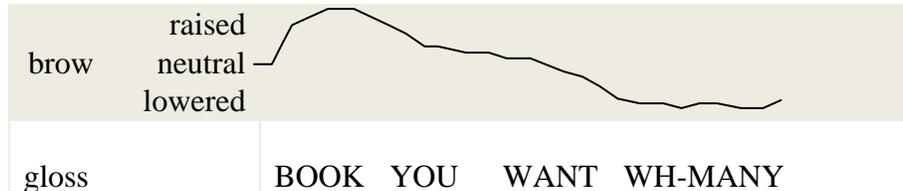


Figure 9: Revised dope sheet.

Recent investigations in co-occurring processes

Weast (2008) extended the work of Grossman and Kegl by adding greater precision to the brow measurements by treating brow position as a numeric quantity and calibrating the measurements to assure consistent annotations across utterances. From this precisely-measured data, Weast was able to demonstrate that the brows can be used to express affect and linguistic processes simultaneously, which supports Wilbur’s layering theory. Thus brows can simultaneously depict syntactic and prosodic processes as well as affect.

Weast’s findings suggest an alternative approach to traditional animation techniques. Conventional animation focuses on the position and manipulation of digital anatomy and provides controls to position brows as a function of time. However, instead of focusing on brow movement as the primary method of creating an animation, we can focus on the linguistic and affect processes that result in brow movement. For example, we can characterize mathematically the effect of affect on brow position, as well as the effect of syntactic processes, and store these as rules. With this information, we can improve on the representation used in Figure 9. Figure 10 shows a dope sheet where the anatomical feature (brows) track has been replaced two new tracks – one for affect and one for syntax.

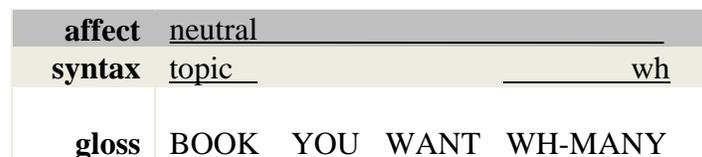


Figure 10: Planning an animation using affect and linguistic content, not geometric poses.

This shows the linguistic and extra-linguistic processes clearly, something that had been lost in the approach shown in Figure 9. Now, instead of storing poses for the production of a syntactic marker in the context of each and every possible affect, this approach first looks up rules for the producing the syntactic marker, then looks up the rules for producing the affect, and applies them. Instead of manipulating the geometric data directly, the animator works with the linguistic rules, which are more powerful because they carry within them the information to manipulate a great deal of geometric data. It is far easier to implement the following as linguistic-based rules:

1. When in the presence of neutral affect, brows for the entire sentence raise or lower, with maximums elevating 21% for yes/no questions and lowering 30% for wh-questions.
2. When in the presence of affect, the range of brow positions used to convey yes/no and wh-questions becomes more compressed (Weast 2008).

An animation model for co-occurring processes

It would be impossible to work directly with numeric data to implement these, because without the structure and goals of the linguistic-based rules, it is not possible to discern the context and intent of an isolated numeric value.

This approach lends extensibility and flexibility to animation planning as well as streamlining the amount of data needed to synthesize the motion. This is an advantageous improvement for automated synthesis, because the rules governing affect and syntax, as well as their interaction, can be automatically applied to create new utterances.

The implementation of the model begins with an artist referring to anatomical data. In a custom application driving a commercially-available animation package, the artist creates a pose for “brows up” and “brows down”. This only needs to be done once.

Based on linguistic findings in the literature, the artist creates the motion for a single process by choosing an appropriate pose and setting up an envelope depicting the intensity of the brow movement. For example, for a wh-question marker, the artist would choose the “brows down” pose and then create the envelope.

The shape of the envelope determines the pose’s influence on the brow over time. The envelope has values ranging from zero to one, where zero is the neutral position of the brows, one is the “brows down” position, and 0.5 would be a setting half way between the two. Again, this only needs to be done once.

Envelopes are not constrained as to length. Figure 11 shows the envelopes and associated poses for the wh- and yes/no question markers as well as envelopes for happy and angry affect.

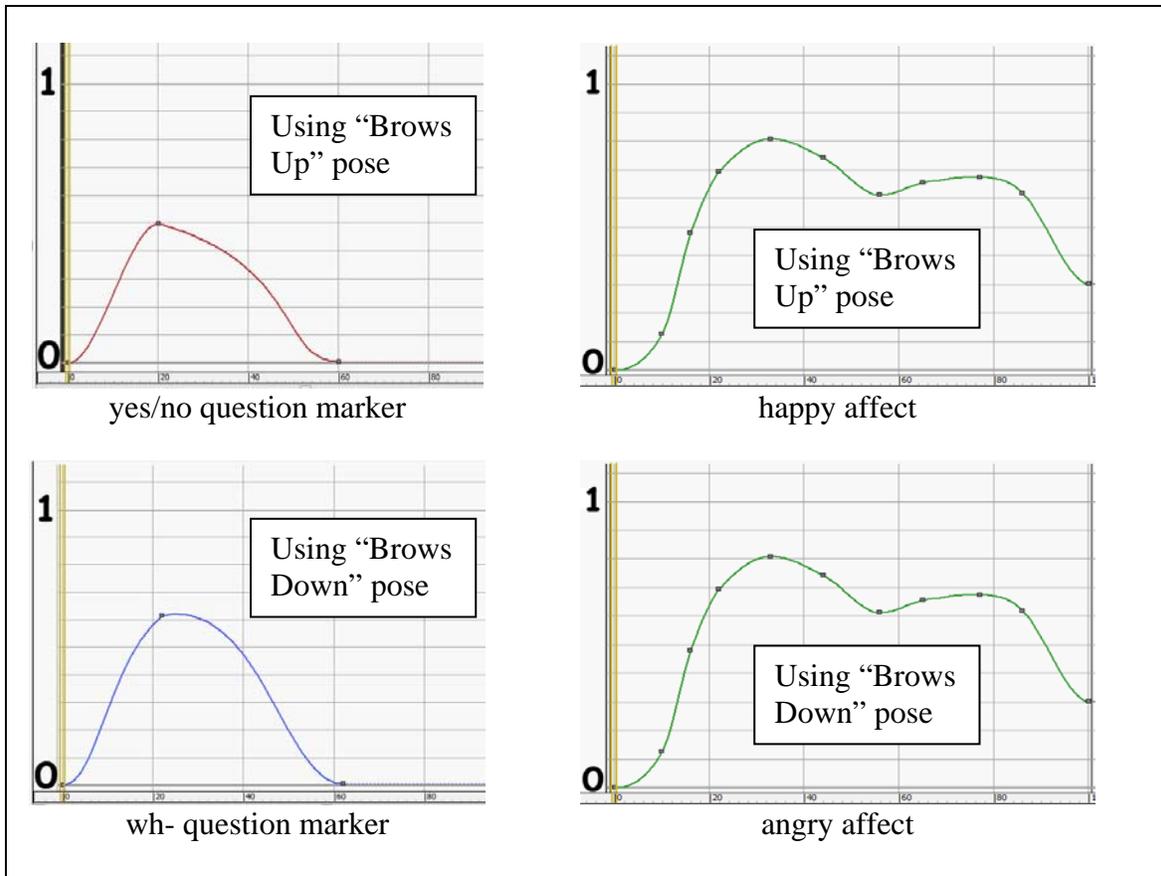


Figure 11: Envelopes and poses for brow behavior

The algorithm that creates the co-occurring brow behavior is implemented in script. Figure 12 gives an informal outline of the algorithm.

```

Preprocessing step:
  For each co-occurring process i
    select /
    {
      case Affect:
        compressionFactor i = 1
      case QuestionMarker:
        if no Affect specified
          compressionFactor i = 1
        else
          compressionFactor i = 0.75
    }

Creating the animation:
  for an instance in time t
  {
    for each co-occurring process i
    {
      Let PoseRotation = Rotation setting of Pose associated with process i
      Let intensity = value of envelope i at time t
      Let Rotation i = PoseRotation * compressionFactor i * intensity
      Apply RotationContribution i to brow
    }
  }

```

Figure 12: Algorithm for creating co-occurring brow behavior.

The role of artist

While the rules provided by linguistics are an essential piece of the solution, animations produced by this method are not quite satisfactory, and the way in which they are not satisfactory has an interesting link to the third design feature of Wilbur's layering theory.

Paradoxically, for a person to easily perceive the brow position on a face requires more information than just the brow position itself. For example, in Figure 13 the brows on the digital model are raised close to their maximal position, yet most viewers perceive this face as having brows that are neutral or close to neutral. An animator could use this model to replicate the brow motion as described by Kegl & Grossman and Weast, but the resulting motion will not be perceived by viewers as matching the original video as annotated in a linguistic analysis. Intensely raised brows are perceived as neutral or slightly raised, and intensely lowered brows are perceived as only slightly lowered.

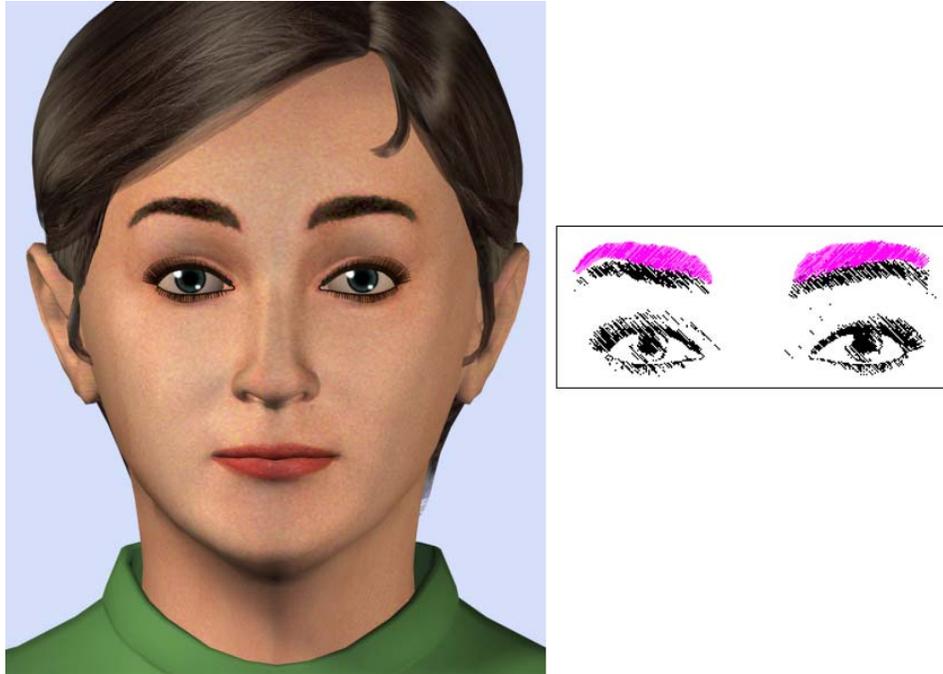


Figure 13: Raised eyebrows in comparison to neutral. The fuchsia areas in the right diagram show the brows from the image as displacements from neutral.

However, animators can help. Animators are skilled in making visual abstractions to aid a viewer's ability to perceive a visual message. In this case, the additional information required is facial wrinkles. When most adults raise their brows, a set of horizontal wrinkles appear on their foreheads, and when they lower their brows sufficiently, a set of vertical wrinkles appear between the brows as demonstrated in Figure 14.



Figure 14: Forehead appearance for lowered, neutral and raised brows.

In fact, the presence of these wrinkles are such a strong indicator of brow position that it is possible to create the impression of a brow raise simply by adding horizontal wrinkles without making any adjustments to the brow position itself. Compare the three images in Figure 15. In each case, the eyebrows are in the exact same position – the only thing that has changed is the presence of facial wrinkles. However the image on the left, when viewed independently of the other two, is perceived as having lowered brows while the image on the right is perceived as having raised brows.



Figure 15: Wrinkles can affect perception of brow position.

Most likely these sets of wrinkles do not qualify as a fully independent nonmanual channel as described in the third design feature of Wilbur's layering theory, but these wrinkles, which are spatially separated from the brows, are at least capable of reinforcing the signal being produced by the brows.

The resulting animations are a synergy of linguistics and computer animation. The linguistics dictates timing and position of the brows, while animation adds visual reinforcement to make the brow position and timing more apparent. **Error! Reference source not found.** is a link to an animation of (1), and **Error! Reference source not found.** is a link to an animation of the same utterance, but produced with negative affect.



Figure 16: Animation of (1), neutral affect.
<http://asl.cs.depaul.edu/SLnL/neutral.avi>



Figure 17: Animation of (1), negative affect.
<http://asl.cs.depaul.edu/SLnL/negative.avi>

Conclusions and next steps

User tests have indicated that this is a promising approach. In a formal study (Schnepp 2009), twenty members of the Chicago Deaf community viewed animations created with this technique. After viewing the animation, users were able to correctly repeat the sentence or phrase including syntax markers in 100% of the cases. Users correctly identified the intended emotional state in 95% of the cases. In terms of preference, all animations were rated “Very Clear” or “Clear” more than 70% of the time.

Using linguistics as a basis for creating animation has several advantages. Linguistics affords a framework to organize massive amounts of geometric data. Linguistic rules are compact, which makes them more tractable for automated application in computer software. Lastly, linguistic rules are flexible, which facilitates the production of novel utterances.

The current system described in this paper can create sentences having simple syntactic structures and offers the option of adding affect. The next step is to extend the system to include exploration of three areas: incorporating additional nonmanual channels, investigating additional syntactic processes and incorporating morphological markers.

Acknowledgments

The authors want to express their appreciation to all the team members of the ASL Synthesizer Project at DePaul University. The authors also thank the reviewers for their thoughtful and constructive comments.

References

- Aarons, Debra. 1994. *Aspects of the syntax of American Sign Language*. Boston, MA: Boston University dissertation.
- Aristidou, Andreas, Jonathan Cameron & Joan Lasenby. 2008. Real-time estimation of missing markers in human motion capture. *Bioinformatics and Biomedical Engineering (ICBBE)*. 2(1) 1343–1346.
- Astrauskas, Michael. 2008. CyberTouch gloves. <http://cubic.asu.edu/resources/index.php> (April 27, 2009.)
- Bahan, Benjamin. 1996. *Nonmanual realization of agreement in American Sign Language*. Boston, MA: Boston University dissertation.
- Baker, Charlotte & Carol Padden. 1978. Focusing on the nonmanual components of American Sign Language. In Patricia Siple (ed.), *Understanding language through sign language research*, 27–58, New York: Academic Press.
- Baker-Shenk, Charlotte. 1983. *A microanalysis of the nonmanual components of questions in American Sign Language*. Berkeley, CA: University of California dissertation.
- Barr, Alan. 1984. Global and local deformations of solid primitives. *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 21–30.
- Boster, Carole. 1996. On the quantifier-noun phrase split in American Sign Language and the structure of quantified noun phrases. In William Edmondson & Ronnie Wilbur (eds.), *International Review of Sign Linguistics*. 159–208. Mahwah, NJ: Lawrence Erlbaum Associates.
- Burtnyk, Nicolas & Marcell Wein, M. 1976. Interactive skeleton techniques for enhancing motion dynamics in key frame animation. *Communications of the ACM [Association for Computing Machinery]*. 564–569.
- Catmull, Edwin. 1972. A system for computer generated movies. *ACM [Association for Computing Machinery] Annual Conference*. 422-431.
- Chang, Yao-Jen & Tony Ezzat. 2005. Transferable videorealistic speech animation. *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation - SCA '05*. 143-151.

- Ciaramello, Francis M. & Sheila Hemami. 2007. 'Can you see me now?' An objective metric for predicting intelligibility of compressed American Sign Language video. *Human Vision and Electronic Imaging XII (SPIE)* 6492
<http://foulard.ece.cornell.edu/publications/Ciaramello - HVEI07.pdf> (January 3, 2008.)
- Coulter, Geoffrey. 1979. *American Sign Language Typology*. San Diego: University of California, San Diego dissertation.
- Cox, Stephen, Michael Lincoln, Judy Tryggvason, Melanie Nakisa, Mark Wells, Marcus Tutt & Sanja Abbott. 2002. TESSA, a system to aid communication with deaf people. *International ACM [Association for Computing Machinery] Conference on Assistive Technologies (ASSETS)*. 205–212.
- Crasborn, Onno, Han Sloetjes, Eric Auer & Peter Wittenburg. 2006. Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. In Chiara Vettori (ed.), *Second Workshop on the Representation and Processing of Sign Languages*. 82–87.
- Davidson, Mary Jo. 2005 *A practical approach to sign language learning for institutional staff*. Chicago, IL: DePaul University dissertation.
- Davidson, Mary Jo. 2006. PAULA: A computer-based sign language tutor for hearing adults. *Intelligent Tutoring Systems 2006 Workshop on Teaching with Robots, Agents, and Natural Language Processing*.
<http://facweb.cs.depaul.edu/elulis/Davidson.pdf> (May 14, 2009.)
- Ehrenstein, Walter H. 2003. Basics of seeing motion. *Arquivos brasileiros de oftalmologia*. 66(5) 44–52.
- Ekman, Paul & Wallace Friesen. 1978. *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- Foley, James, Andries van Dam, Steven Feiner & John Hughs. 1990. *Computer Graphics: Principles and Practice*. 2nd ed. Reading, MA: Addison-Wesley.
- Fotinea, Stavroula-Evita, Eleni Efthimiou, George Caridakis, & Kostas Karpouzis. 2008. A knowledge-based sign synthesis architecture. *Universal Access in the Information Society*. 6(4) 405-418.
- Furst, Jacob, Karen Alkoby, Nedjla Ougouag-Tiouririne, Roymieco Carter, Juliet Christopher, Mary Jo Davidson, Dan Ethridge, Damien Hinkle, Glenn Lancaster, John McDonald, Lori Smallwood, Jorge Toro, Shuang Xu & Rosalee Wolfe. 2002. Making Airport Security Accessible to the Deaf. *International Conference on Computer Graphics and Imaging (IASTAD)*. 38–43.

- Goldman, Michael. 2000. Real-time optical motion capture: Will it change how you make TV animation? *Millimeter*.
http://digitalcontentproducer.com/dcc/revfeat/video_realtime_optical_motion
(November 15, 2008.)
- Grossman, Ruth. 2001. *Dynamic facial expressions in American Sign Language: Behavioral, neuroimaging, and facial-coding analyses for Deaf and hearing subjects*. Boston MA: Boston University dissertation.
- Grossman, Ruth & Judy Kegl. 2006. To capture a face: a novel technique for the analysis and quantification of facial expressions in American Sign Language. *Sign Language Studies* 6(3) 273–305.
- Huenerfauth, Matthew. 2006. *Generating American Sign Language classifier predicates for English-to-ASL machine translation*. Philadelphia, PA: University of Pennsylvania dissertation.
- Matt Huenerfauth. 2009. A linguistically motivated model for speed and pausing in animations of American Sign Language. *ACM Transactions on Accessible Computing*. 2(2) 1-31.
- Kirk, Adam, James O'Brien & David Forsyth. 2005. Skeletal Parameter Estimation from Optical Motion Capture Data. *IEEE [Institute of Electrical and Electronics Engineers] Computer Vision and Pattern Recognition (CVPR)*. (2) 782–788.
- Kovar, Lucas, Michael Gleicher & Frédéric Pighin. 2002. Motion graphs. *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 473–482.
- Kovar, Lucas, Michael Gleicher & John Schreiner. 2002. Footskate Cleanup For Motion Capture Editing *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)* 97–104.
- Manfred Lau, Ziv Bar-Joseph, and James Kuffner. 2009. Modeling spatial and temporal variation in motion data. *ACM Transactions on Graphics (SIGGRAPH ASIA 2009)*, 28(5), 171.
- Lee, Jehee, Jinxiang Chai, Paul S. A. Reitsma, Jessica K. Hodgins & Nancy S. Pollard. 2002. Interactive control of avatars animated with human motion data. *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 491–500.
- Liddell, Scott K. 1978. Nonmanual signals and relative clauses in American Sign Language. In Patricia Siple (ed.), *Understanding language through sign language research*, 59-90, New York: Academic Press.

- Liddell, Scott K. 1986. Head thrust in ASL conditional marking. *Sign Language Studies*. 52: 243-363.
- Liddell, Scott K. & Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies*. 64 195–277.
- Maestri, George. 1996. *Digital character animation*. Indianapolis, IN: New Riders.
- Magenat-Thalmann, Nadia, Richard Laperriere & Daniel Thalmann. 1988. Joint dependent local deformations for hand animation and object grasping. *Graphics Interface*. 26–33.
- McDonald , John, Jorge Toro, Karen Alkoby, Andre Berthiaume, Pattaraporn Chomwong, Juliet Christopher, Mary Jo Davidson, Jacob Furst, Brian Konie, Glenn Lancaster, Steven Lytinen, Lopa Roychoudhuri, Eric Sedgwick, Noriko Tomuro & Rosalee Wolfe. 2000. An improved articulated model of the human hand. *International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*. 306–313.
- McDonald, John, Karen Alkoby, Roymieco Carter, Juliet Christopher, Mary Jo Davidson, Dan Ethridge, Jacob Furst, Damien Hinkle, Glenn Lancaster, Lori Smallwood, Nedjla Ougouag-Tiouririne, Jorge Toro, Shuang Xu, & Rosalee Wolfe. 2002. A direct method for positioning the arms of a human model, *Graphics Interface*. 99–106.
- Miller, Iain, Stephen McGlinchey & Benoit Chaperot. 2006. Anomaly detection in magnetic motion capture using a 2-layer SOM network. *IEEE Symposium on Computational Intelligence and Games*. 236–242.
- Moeslund, Thomas B. & Erik Granum. 2001. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding* 81(1) 231–268.
- Moeslund, Thomas B., Adrian Hilton & Volker Krüger. 2006. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding* 104(2-3) 90–126.
- Motion Analysis. 2009. Digital realtime system.
<http://www.motionanalysis.com/html/animation/raptor4.html> (April 27, 2009.)
- Neidle, Carol, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan & Robert Lee. 2000. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, MA: The MIT Press.
- Neidle, Carol, Stan Sclaroff, & Vassilis Athitsos. 2001. SignStream™: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior research methods, instruments, and computers* 33(3) 311–320.

- Pandzic, Igor & Robert Forchheimer, (eds). 2002. *MPEG-4 Facial Animation: The standard, implementation and application*. Hoboken, NJ: John Wiley & Sons.
- Poor, Geoff., & D. Wilkins, 1986. Sign language instruction with a pre-production phase. In Carol Padden (ed.), *Fourth National Symposium on Sign Language Research and Teaching*. 134–144.
- Reilly, Judy, Marina McIntire & Ursula Bellugi. 1990. Faces: The relationship between language and affect. In Virginia Volterra & Carol Erting (eds.), *From Gesture to Language in Hearing and Deaf Children*, New York, NY: Springer-Verlag. 128–141.
- Rezzoug, N., P. Gorce, A. Heloir, S. Gibet, N. Courty, J.F. Kamp, F. Multon & C. Pelachaud. 2006. Virtual humanoids endowed with expressive communication gestures: the HuGEx project. *IEEE International Conference on Systems, Man and Cybernetics*. 4445 - 4450.
- Sanders, Matthew, Tim Dale, Andrew Ryzner, Laurie Ide, Joe Humbert. 2005. A comparative analysis of keyframing to motion capture in American Sign Language animations.
[http://www2.tech.purdue.edu/cg/Courses/cgt411/data/CGT%20411%20Fall%202005 Papers.pdf#page=23](http://www2.tech.purdue.edu/cg/Courses/cgt411/data/CGT%20411%20Fall%202005%20Papers.pdf#page=23) (April 27, 2009.)
- Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign Language and Linguistic Universals*. Cambridge, UK: Cambridge University Press.
- Schnepf, Jerry, Rosalee Wolfe, John McDonald. 2009. Synthetic corpora: A synergy of linguistics and computer animation. *Proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies Language Resources and Evaluation Conference (LREC)* Valletta, Malta, May 2010 Sedgwick, Eric, Karen Alkoby, Mary Jo Davidson, Roymieco Carter, Juliet Christopher, Brock Craft, Jacob Furst, Damien Hinkle, Brian Konie, Glenn Lancaster, Steve Luecking, Ashley Morris, John McDonald, Noriko Tomuro, Jorge Toro & Rosalee Wolfe. 2001. Toward the effective animation of American Sign Language. *International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*. 375–378.
- Stokoe, William C. Jr., Dorothy Casterline & Carl Croneberg. 1965. *A dictionary of American Sign Language*. Washington, DC: Gallaudet College Press.
- Toro, Jorge. 2004. Automated 3D Animation System to Inflect Agreement Verbs. Paper presented at the sixth High Desert Linguistics Conference, Albuquerque, NM.
- Weast, Traci. 2008. *Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows*. Arlington, TX: University of Texas at Arlington dissertation.

- Wilbur, Ronnie. 1991. Intonation and focus in American Sign Language. In Yongkyoon No & Mark Libucha (eds.), *ESCOL-90: Seventh Eastern States Conference on Linguistics*, 320–331, Columbus, OH: Ohio State University Press.
- Wilbur, Ronnie. 2000. Phonological and prosodic layering of nonmanuals in American Sign Language. In Karen Emmorey & Harlan Lane (eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 215–241, Philadelphia, PA: Lawrence Erlbaum Associates.
- Wilbur, Ronnie. 2003. Modality and the structure of language. In Marc Marschark & Patricia Spencer, (eds.), *Oxford handbook of Deaf studies, language and education*. Oxford, UK: Oxford University Press. 332–346.
- Wolfe, Rosalee, Karen Alkoby, Jamita Barnett, Pattaraporn Chomwong, Jacob Furst, Gary Honda, Glenn Lancaster, Frank "Tony" Lavoie, Steven Lytinen, John McDonald, Lopa Roychoudhuri, Carolyn Taylor, Noriko Tomuro & Jorge Toro. 1999. An interface for transcribing American Sign Language. *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH) Sketches*, 229–230.
- Zeltzer, David. 1982. Motor control techniques for figure animation. *Computer Graphics and Applications* 2(9) 53–59.

Contact Information

Rosalee Wolfe
School of Computing
DePaul University CDM 401
243 S. Wabash Ave.
Chicago, IL 60604
312 362 6248 (voice)
rwolfe@depaul.edu

Peter Cook
Department of ASL-English Interpretation
Columbia College Chicago
33 E. Congress
Chicago, IL 60605
pcook@colum.edu

John C. McDonald
School of Computing
DePaul University CDM 401
243 S. Wabash Ave.
Chicago, IL 60604
312 362 5142 (voice)
jmcdonald@cti.depaul.edu

Jerry Schnepf
School of Computing
DePaul University CDM 401
243 S. Wabash Ave.
Chicago, IL 60604
312 362 8264 (voice)
jschnepf@cdm.depaul.edu