TOWARD THE EFFECTIVE ANIMATION OF AMERICAN SIGN LANGUAGE

Eric Sedgwick, 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

School of Computer Science, Telecommunications and Information Systems
DePaul University
243 South Wabash Avenue Chicago, IL 60604-2301
asl@cs.depaul.edu

ABSTRACT

American Sign Language (ASL) is the primary language used by the North American Deaf Community. We present our method for producing natural animations of fingerspelling, a functionally important subset of ASL. User testing demonstrates that our animations are readily identified by members of the deaf community.

Keywords: Animation, American Sign Language, Fingerspelling.

1. INTRODUCTION

American Sign Language (ASL) is the third most commonly used language in North America [Deaf00] [Stern96]. Its use of handshape, body movement, and facial expression is both efficient and elegant, making ASL the primary language used within the North American Deaf Community. Outside the deaf community, however, knowledge of ASL is rare [Baker80]. For the deaf, lack of access to spoken English is a significant barrier to participation in the hearing world.

While it may appear that the closed-captioning technology used on television would bridge the communication gap, it offers only a partial solution at best. A common misunderstanding is that ASL is merely a gestured form of English. In fact, ASL is a natural and living language. While it shares some vocabulary in common with English, it is linguistically distinct, with a grammar radically different from English [Baker80] [Valli93]. To most members of the deaf community, English is a second language; the average English reading level of deaf adults in the US is between the third and fourth grade level [Holt94]. Closed captions are *not* as effective for a deaf person as subtitles in a foreign movie are for a hearing person since the closed captions are not in the deaf person's native language.

A more effective means of translation is the use of an ASL interpreter. Certified ASL interpreters are skilled in quick and accurate translation between English and ASL. Though interpreters are effective, they are not appropriate for many common situations. Interpreters are trained professionals who demand (and deserve) a rate of compensation that is prohibitive for everyday use. Interpreters are privy to otherwise private communication regarding medical, legal and other personal matters. Also, appointments must be scheduled in advance, a difficulty in emergency situations.

We believe that a *personal digital translator*, a system that would translate written or spoken English into ASL, would better bridge the deaf and hearing worlds. The presentation would be in the deaf person's native language, and access would be affordable, private, and available on demand.

A critical component of such a technology is the ability to generate animations of ASL in a flexible and natural manner. To be useful these animations will need to be quickly and reliably recognized by ASL signers. To create such animations, we have built a digital human model with a set of controls that allow for the intuitive entry of ASL signs, and are developing a method for animating ASL sentences.

2. ASL SIGNS

The ASL vocabulary consists of *signs*. While additional elements may be present, there is a consensus among ASL linguists that handshape, location and movement are essential elements of a sign [Lidde89]. *Handshapes* are particular configurations of the hand; a relatively small set (40) generates the majority of signs in ASL [Tenna98]. Comprehension of a sign depends on recognizing the handshape. For example, the ASL signs for "year" and "world" have the same pattern of movement but differing handshapes.

To communicate proper nouns, acronyms and technical terms the deaf use *fingerspelling*. Fingerspelling is the process of spelling English words with handshapes representing the letters of the English alphabet. Movement is restricted to the hand and is independent of the position and movement of the body beyond the wrist. Fingerspelling is too inefficient for general communication and accounts for a small portion of typical conversation.

Although fingerspelling comprises a small portion of ASL [Padde98], it is an excellent test of any approach for several reasons. First, fingerspelling contains most of the handshapes that appear in the signs of ASL. Solving problems in animating fingerspelling will solve difficulties in animating handshapes when creating signs. It also allows us to restrict our focus to the hand and still produce animations with content. Of the approximately 40 joints necessary for creating signs, 30 of them are in the hands. Animating the hand is a significant portion of the process of animating a complete ASL sign. Moreover, we can test comprehension among ASL signers and determine whether our approach is well directed. For these reasons, fingerspelling was chosen as the first task for our digital human model.

3. HAND MODEL

Modeling ASL handshapes poses unique challenges. In many applications, for example grasping [Rijpk91], the hand itself is largely in an open position. In contrast, many ASL handshapes require the fingers to be in close proximity to each other and to the palm. In some handshapes the hand is in a very compact position. See Figure 1. We have taken great care that our hand's appearance and movement is natural in a bent position [McDon00].

The human hand is a complex articulated system consisting of 27 bones [Caill82]. The user is referred to [Lands55], or [Nette87] for accurate descriptions. Our hand simplifies this anatomy, see [McDon00] for details. Using the scripting facility available in a commercial animation package we have built a set of slider controls that enable the user to readily generate handshapes. The movement of each finger is controlled by fundamental motions identified by ASL linguists [Brent98] [Lidde89] [Sandl89].

4. FINGERSPELLING

We aim to produce animations of ASL that appear natural and can be quickly and reliably recognized by members of the deaf community. Using our hand controls, we can easily generate handshapes. We create fingerspelling animations by using handshapes to set key frames, and interpolating the rotations of the joints. Because the hand is complex, we found that using inverse kinematics for control yielded unpredictable and unnatural movement.

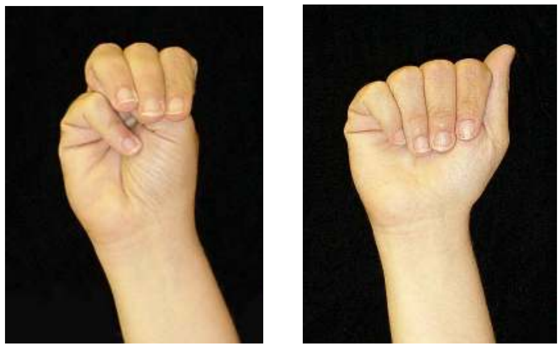

Figure 1: "M" and "A"

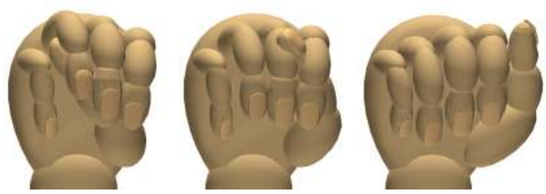

Figure 2: Naïve interpolation between "M" and "A," as in spelling the word "MAD."

Unfortunately, between many pairs of letters, straightforward interpolation leads to collisions of the fingers. Particularly troublesome are letters for which the handshape is in a "tight" or "entangled" position. For example the letters "M," "N" and "T" have the hand closed like a fist, with the thumb tucked underneath one or more fingers. See Figure 1. Naïve interpolation from these letters passes the thumb through these fingers. See Figure 2. In the handshape "R" the index and middle fingers are crossed, they must be uncrossed before interpolating to any other letter. See Figure 3. Other collisions are less catastrophic, involving collisions only between fingertips, and some pairs cause no collisions at all.

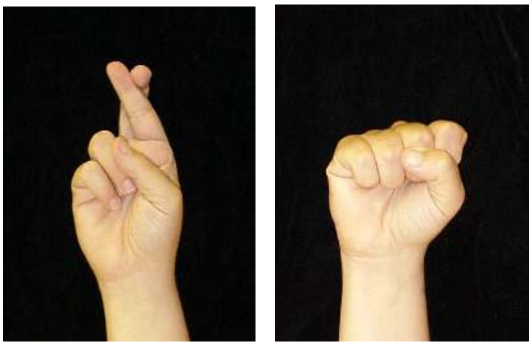

Figure 3: "R" and "S"

It is possible to use a conventional collision avoidance algorithm to protect the fingers. However, this will produce minimal solutions, not necessarily natural ones. The hand has complicated musculature controlling its motion, and the resulting constraints are not accounted for by a general algorithm.

Since ASL is almost entirely dependent on a relatively small set of handshapes, we have chosen a data-driven solution. For each colliding pair of handshapes we have designed *interpolation handshapes* which mimic natural intermediate positions made by the human hand while fingerspelling.

Our data-driven approach capitalizes on a restricted set of possible collisions. To represent the 26 letters of the alphabet, we need only 22 handshapes. Four letters repeat a handshape and either change palm orientation or add simple movement: "U" is a reoriented "H," "J" is a moving "I," "P" is a reoriented "K" and "Q" is a reoriented "G." As transitions between handshapes are symmetric, we were in the fortunate situation to have at most $22^2/2 = 242$ pairs to consider.

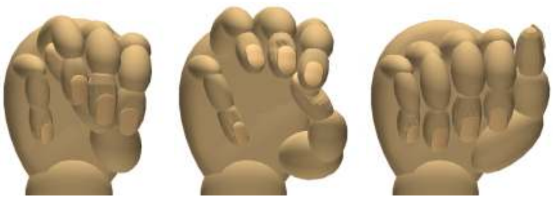

Figure 4: Interpolation between "M" and "A" using an intermediate handshape.

For the initial cost of data entry we avoid the computational overhead associated with a general solution. We also obtain more realistic movement of the fingers. For example, when a human hand transitions between two "tight" handshapes, inactive fingers will naturally open slightly, whereas a minimal solution leaves them fixed. See Figure 4. As the reading of a written English word is more than a sequence of recognizing letters, the pattern of movement in a fingerspelled word indicates the relationships between the letters and thus contributes to its comprehension.

5. USER TESTING

Our goal is to produce animations that are readily understood by members of the deaf population. To evaluate the effectiveness of our work, we conducted two user tests of our fingerspelling animations: a preliminary test with deaf students at a local area high school, and a more thorough test with attendees at DeafExpo 1999, a national exhibition for and by the deaf community [David00]. Our participants had varying levels of ASL fluency. While many of the students were novice ASL signers, the typical DeafExpo attendee was proficient in ASL and several were expert signers. Each participant was asked to identify still images of the alphabetic handshapes and a sequence of fingerspelling animations shown at varying speeds. The feedback that we received included:

- With the exception of the word "Coke," our animations were recognized by 85% of the DeafExpo participants on the first attempt.
- Some participants preferred to view the animations at a high rate of speed (2.5 letters per second).
- We obtained useful feedback on the appearance of our hand, and the formation of our handshapes. Still images of the letters "C," "O" and "E" were often confused, which undoubtedly contributed to the low rate of recognition, 47%, for our animation of the word "Coke." This is likely a fundamental difficulty in ASL. Stungis conducted studies where participants viewed videotapes of ASL and noted that the handshapes for the letters "C," "O" and "E" were often confused [Stung81].
- The participants were universally enthusiastic about our project.

Our emphasis on natural movement was well founded. Interestingly, several participants had more success identifying animations than still images of handshapes, which leads us to believe that motion is an important factor in comprehension of ASL. We have also been pleasantly surprised that visitors to our website, `asl.cs.depaul.edu`, have been able to understand fingerspelled words that are stored as small (66 by 50 pixels) animated gifs.

6. FUTURE WORK

The high recognition rate by deaf users gives us confidence that our overall approach is effective. We can now fine tune our system to address the user feedback without having to alter our fundamental design. In addition to making these improvements, we have begun work on a *sentence generator*, capable of generating animations of complete ASL sentences. We have attached our hand, exaggerated in proportion, to its body. While handshapes are still essential, we are now concerned with their context in respect to the body as whole. In fact, manipulating the body allows us to address the issue raised most during user testing: the letters "C," "O" and "E" were sometimes confused. We suspect that this was due, in large part, to our straight-on angle of presentation, from which these letters have a similar appearance. Now that we have attached the hand to a body, rotation in the wrist will allow the hand to adjust to a more natural angle to view these handshapes.

While our work with the hand required manipulation of the vast majority of the joints that are needed to create ASL sentences, working with the body introduces some challenging issues for our future work. First, in contrast to the small set of handshapes used in ASL, general signs involve more arbitrary configurations and movements of the arms, head and torso. To accomplish this, we will need a more general method of collision avoidance.

A second issue concerns the use of facial expressions. In ASL facial expressions do more than modify mood. They are essential, for example, for expressing interrogative and imperative sentences. This is a challenge; movements of the face are more subtle than those of rotating limbs and will need to be easily understood by a deaf audience.

REFERENCES

[Baker80] Baker-Shenk, S. and Cockely, D.: American Sign Language, A Teacher's Resource Text on Grammar and Culture (the Green Book), *Gallaudet University Press*, 1980.

[Brent98] Brentari, D.: A Prosodic Model of Sign Language Phonology, *MIT Press,* Cambridge, Massachusetts, 1998.

[Caill82] Cailliet, R.: Hand Pain and Impairment, Third Edition, *F.A. Davis Company*, 1982.

[David00] Davidson, M.: Improved Hand Animation for American Sign Language, *Submitted to*, 2000.

[Deaf00] Deafworld web site: `dww.deafworldweb.org/int/us/hm` as of October 4, 2000.

[Holt94] Holt, J.: Demographic, Stanford Achievement Test – 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results. `http://www.gallaudet.edu/~cadsweb/sat-read.html`, October 2000.

[Lands58] Landsmeer, J.: A Report on the Coordination of the Interphalangeal Joints of the Human Finger and its Disturbances, *Acta Morphologica Neerlandop-Scandinavica*, 2, pp.59-84, 1958.

[Lidde89] Liddell, S., Johnson, R.: American Sign Language: The Phonological Base, *Sign Language Studies*, 64, 195-277, 1989.

[McDon00] McDonald, J., et.al.: An Improved Articulated Model of the Human Hand, *Proceedings of the 8th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*, 2000.

[Padde98] Padden, C. A.: The ASL Lexicon, *Sign Language and Linguistics*, Vol.1, No.1, pp.36-64, 1998.

[Rijpk91] Rijpkema, H., Girard, M.: Computer Animation of Knowledge-Based Human Grasping, *Computers Graphics*, Vol.25, No.4, pp.339-348, 1991.

[Sandl89] Sandler, W.: Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language, *Foris Publications,* Dordrecht, Holland, 1989.

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

[Stung81] Stungis, J.: Identification and Discrmination of Handshape in American Sign Language: Negative Incorporation, *Sign Language Studies,* Vol.3. No.5., pp.20-30, 1981.

[Tenna98] Tennant, R. and Brown, M.: The American Sign Language Handshape Dictionary, *Clerc Books*, Washington D.C., 1980.

[Valli93] Valli, C. and Luca, C.: Linguistics of American Sign Language, *Galluadet University Press*, 1993.