Improved Hand Animation for American Sign Language

Mary Jo Davidson, Karen Alkoby, Eric Sedgwick, 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
FAX: (312)-362-6116
E-mail: asl@cs.depaul.edu

Introduction

American Sign Language (ASL) is a natural language used by members of the North American Deaf community and is the third most widely used language in the United States [Ster96][Deaf00]. At present, deaf people rely on sign language interpreters for access to spoken English. The cost and limited availability of interpreters has contributed to isolation for many in the deaf community. A pioneering effort to make English accessible to the deaf community was closed captioning on television. This is an incomplete solution because closed-captioning requires reading skills that are beyond many in the deaf community. Since ASL is very different linguistically from English. [Klim79] [Vall93], most native adult ASL signers read English at the third or fourth grade level [Holt94]. We believe that a technology that translates English into ASL, especially for conversation, would provide greater freedom and privacy, and greater opportunity for the deaf. For example, medical and legal matters could be transacted doctor-to-patient or attorney-to-client without the need for an interpreter. A personal digital translator that would translate English into American Sign Language, would better bridge the gulf between deaf and hearing worlds.

While there are currently many different approaches to the digital presentation of ASL, from video clips of ASL signs to letter-for-letter translation of English to still images of fingerspelling handshapes, we believe that the best approach is animated three-dimensional computer graphics (CG). The nature of conversation requires the type of flexibility that CG can provide. This paper describes one of the crucial components of a future personal digital translator, which is hand animation.

The Nature of ASL

American Sign Language is a rich and varied natural language. While ASL shares some vocabulary with English, it is not a direct translation of English words and sentence structure. It presents many of the same challenges of any language translation process, but adds the complexity of changing modality from aural/oral to visual/gestural [Alko99a]. There are two major subsets of ASL - signs that express words, concepts and complex phrases, and fingerspelling.

Word/phrase signs can express an extraordinary range of meaning by using the natural

geography of the body and facial expression, in addition to the hands. As practiced by fluent signers, word/phrase signs are economical and of endless variety. Positioning and facial expression convey differences in sentence type (e.g. question vs. exclamation), as well as level of intensity. Word/phrase signs account for the vast majority of a typical ASL conversation [Tenn98].

Fingerspelling is the use of the hands to spell out English words and numbers letter-for-letter. Fingerspelling is used for proper nouns, technical terms, acronyms, and in situations where no word/phrase sign exists. Fingerspelling slows ASL conversation, but is necessary for complete communication.

Although word/phrase signs and fingerspelling have very different roles in ASL communication, they share a common physical building block, the handshape. All signs and fingerspelling are composed of one or more handshapes. Additional information is conveyed by facial expression, hand orientation and position, but the handshape is a key factor.

Digital Presentation Technologies

Technologies that can present ASL digitally already exist. They include: Video clips of fingerspelling [Mich99] or word/phrase signs expressing a specific concept [Ster96] A series of still images of fingerspelling presented in sequence to spell an English word [ASL99] Animated three-dimensional computer graphics (CG) [Su98] [Tomu00] Both video clip technology and still image technology are limited in their ability to create the full range of ASL ultimately necessary. In order to be useful in conversation, a presentation technology must be flexible enough to create new sentences from signs, taking into account such items as the correct conjugation of verbs.

Computer Graphics Technology

We believe that computer graphics is the most appropriate technology choice for presentation of ASL on a digital translator. Its support for "on the fly" creation of new animations based on both existing rules/conditions and input from outside sources provides the flexibility necessary for ASL sign translation. We believe that CG has the potential for:

Conveying the grammar of ASL more fully, e.g. questions, verb tenses Supporting translation beyond "phrase book" type, scripted applications Providing support for more combinations of signs and development of new signs based on combinations of handshapes and physical positioning While CG best supports a broad spectrum of the users' communication needs, it carries with it two major challenges, which are:

Development of representations of the fine motor movements of the hand. CG has been used frequently to emulate body physiology for movie work, where the appearance of gross motor movement has been the only requirement. Any acceptable representation of ASL requires small subtle movement of the hands and other parts of the body.

Lack of physicality - objects can pass through each other. This problem, called "collision

avoidance" is similar to the challenge presented to virtual reality applications when virtual objects must be grasped while the hand is constrained by the implied boundary of the object.

Fingerspelling as a Prototype Application

In order to find solutions to these problems, the team decided to choose the limited domain of fingerspelling for prototyping. This would be an opportunity to develop a scale model of the problem set and iteratively refine solutions that would apply to the general category of hand animation used in ASL. It would also provide the opportunity to get user feedback on the representation of the hand, including realism and recognizability of the resulting fingerspelling.

As discussed above, fingerspelling is a small but essential portion of ASL. The handshapes used in fingerspelling are derived from the same basic set of handshapes that form word/phrase signs. We believed that solving the problem of accurately portraying fine motor movement and the problem of collision avoidance for fingerspelling would lead to a breakthrough in animating entire sentences in ASL. For these reasons, fingerspelling was chosen as the prototype domain. When creating the handshapes, we developed a more accurate hand model than those previously available [McD00]. When animating the hand, we developed a simplified collision avoidance approach that capitalized on a data-driven solution instead relying on a general brute-force technique [Sedg01].

Usability Concerns

Usability is a central concern of the ASL project. In concept, the personal digital translator could become a constant resource for the deaf as they carry out day-to-day tasks. It is critical that the user spend most of the time it takes for an interaction with the digital translator observing the results of the translation, not in observing oddities of the representation. The user would also have to be able to recognize the signs at a high presentation speed. However appealing the animations might appear to a hearing person, we think it is imperative that our approach be tested with people who are likely to use the translator.

Usability Testing - Methodology

In order to get early feedback on the sign images and animation created using the team's CG approach, we conducted an exploratory usability test to get feedback on our fingerspelling animations, speed of presentation, and general appearance of the hand. We conducted our usability tests with two groups of users: deaf high school students and participants at Deaf Expo, an annual conference that explores many of the issues and needs of the deaf community. Both tests took place in November 1999. All participants were proficient or moderately proficient in ASL. In both test sessions the protocol was the same, specifically each participant was: Shown a series of CG animations which presented fingerspelling of a word at three different speeds, highest speed first. The participant was asked to identify the word. If the participant could not recognize the word the next slower speed animation was used. The participant was asked which speed they preferred. This was repeated for a number of words

Usability Testing - Findings for High School Students

The deaf high school students tested the images and animation first. While their ASL skill varied, most students could recognize the fingerspelled word on first or second presentation at the fastest speed. Based on this result, an animation that fingerspelled at an even faster speed was created and used during the tests at Deaf Expo. The fastest speed presented at Deaf Expo was 2.5 letters per second. This compares well with studies of recognition rates for fingerspelling [Blad98].

Usability Testing - Findings for Both Groups

The following items were findings from both the high school students and Deaf Expo test participants.

Most users were able to recognize the fingerspelled words in one or two presentations. Most were able to recognize the words at the highest speed of presentation. All participants were enthusiastic about the potential of the personal digital translator.

Results and Future Work

We will do some fine tuning of the handshapes, as users commented on the appearance of the thumb in some handshapes and sometimes confused the letters "C", "O" and "E".

The results of the usability test and user comments gave a clear indication that the approach we are using for hand animation is sound and we will be using it in our future work. We are using these results as we work on a sentence animator.

References

[Alko99a] Alkoby, K. A Survey of ASL Tenses. <>Proceedings of the 2nd Annual CTI Research Symposium. Chicago, Illinois, November 4, 1999.

[ASL99] ASL Fingerspelling Dictionary, http://where.com/scott.net/asl/

[Blad98] Blades, F. and Kyle, J. Video Conference and Sign Transmission: Studies carried out as part of the FORUM work on video conferencing. 1998. http://www.sign-lang.uni-hamburg.de/Forum/docs/vidconf.htm

[Deaf00] Deafworld web site, http://dww.deafworldweb.org/int/us/

[Holt94] Holt, J., "Demographic, Stanford Achievement Test – 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results".

[Klim79] Klima, E. and Bellugi,U., The Signs of Language. Harvard University Press, 1979.

[McD00] 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. May 2000.

[Mich99] Personal Communicator CD, Michigan State University Communication Technology Laboratory, 1999.

[Sedg01] Sedgwick,E.,et.al., Toward the effective animation of ASL. Submitted to WSCG 2001.

[Ster96] Sternberg, M., The American Sign Language Dictionary, Multicom, 1996. (CD ROM)

[Su98] Su, S.A., "VRML-based Representations of ASL - Fingerspelling on the World-Wide Web.

[Tenn98] Tennant, R. and Brown, M. The American Sign Language Handshape Dictionary. Washington, DC: Clerc Books, 1998.

[Tomu99] Tomuro, N.,et.al, An Alternative Method for Building a Database for American Sign Language. Presented at the Technology and Persons with Disabilities Conference 2000. California State University at Northridge, Los Angeles, CA March 20-25, 2000.

[Vall93] Valli, C. and Luca, C., Linguistics of American Sign Language, Gallaudet University Press, 1993.

ABSTRACT Developing successful sign language recognition, generation, and translation systems requires expertise in a wide range of felds, including computer vision, computer graphics, natural language processing, human-computer interaction, linguistics, and Deaf culture. Despite the need for deep interdisciplinary knowledge, existing research occurs in separate disciplinary silos, and tackles separate portions of the sign language pro-cessing pipeline. This leads to three key questions: 1) What does an interdisciplinary view of the current landscape reveal? 2) What are the biggest challenges Authentic American Sign Language (ASL) lessons on YouTube taught by deaf instructors. Many early lessons intended for beginners have subtitles. Some use grap...Â Be mindful of Indigenous sign languages still existing in United States (of America), Canada and abroad since Langue des Signes Française (LSF, EN: French Sign Language) and British Sign Language (BSL) were brought to the colonies by white European settlers, which later became American Sign Language and displacing other languages. Using the Sign Language MNIST dataset from Kaggle, we evaluated models to classify hand gestures for each letter of the alphabet. Due to the motion involved in the letters J and Z, these letters were not included in the dataset. However, the data includes approximately 35,000 28x28 pixel images of the remaining 24 letters of the alphabet.Â While this analysis sets a solid baseline for American Sign Language recognition, more work needs to be done to apply this concept in real-time. This would require further work with the LeapMotion API to enable real-time generation of data, feeding through the model, and identification of the word and/or numbers. This would also require the model to be able to handle more than the 60 class labels it currently deals with. The tech giant is sharing its hand-tracking software to allow others to develop translation apps.Â Google has developed software that could pave the way for smartphones to interpret sign language. The tech firm has not made a product of its own but has published algorithms which it hopes developers will use to make their own apps. Until now, this type of software has only worked on PCs. But campaigners from the hearing-impaired community have welcomed the move, but say the tech might struggle to fully grasp some conversations. The Nature of ASL American Sign Language is a rich and varied natural language. While ASL shares some vocabulary with English, it is not a direct translation of English words and sentence structure. It presents many of the same challenges of any language translation process, but adds the complexity of changing modality from aural/oral to visual/gestural [Alko99a].Â However appealing the animations might appear to a hearing person, we think it is imperative that our approach be tested with people who are likely to use the translator.