1-10-2011

Avatar to Depict Sign Language: Building from Reusable Hand Animation

Rosalee Wolfe
DePaul University

John McDonald
DePaul University

Jerry C. Schnepp
Bowling Green State University, schnepp@bgsu.edu

An Avatar to Depict Sign Language: Building from Reusable Hand Animation

Rosalee Wolfe, John McDonald, Jerry Schnepp
School of Computing
DePaul University
Chicago, IL 60604
{wolfe,jmcdonald,jschnepp}@cs.depaul.edu

An essential component of an automatic speech-to-sign translator is a method of representing and depicting the target language. Unlike many spoken languages, sign languages do not have a universally-accepted written form, and depicting it convincingly is an open question. The goal of our work is to develop a representation capable of synthesizing novel utterances of American Sign Language (ASL) as realistic and believable 3D animations.

Any representation must be flexible enough to modify vocabulary algorithmically so that it conforms to the grammatical structure of ASL. Additionally, realism is necessary because a viewer should not be distracted by differences between animations signed by an avatar and video of utterances produced by a human signer. Animations signed by an avatar should be as easy to understand as video of a human signer.

A survey of the state of the art for modeling, rigging and animating a 3D character reveals a well-established body of knowledge used in films and computer games for over twenty years. Likewise, linguistic scholars have made great strides in characterizing sign language since Stokoe first discovered that the visual/gestural system of communication he observed at Gallaudet was a true language in its own right.

Using conventional techniques to represent phonemes have yielded promising results, but user studies show that the resulting animations fall short of the goal of realism. User feedback has shown that the standards for physical accuracy in depicting ASL are far higher than that required for film or games. The motion for a movie, once created, never changes, and for a game, is limited to a small number of choices. In contrast, subtle changes in sign production result in vastly different perceptions, requiring an improved approach to animation reusability.

To address the modeling and rigging issues, we have developed a more accurate behavior for the human thumb, created internal handshape movement, and improved the positioning of the arms via "JK" mathematics. This, coupled with a user interface that facilitates positioning via linguistic terms, allowed us to synthesize grammatically correct sentences involving agreement verbs and to retarget signs to different avatar geometries. To address the challenge of nonmanual signals, we created dynamically interacting sets of texture maps to portray the forehead and introduced a rubber band rig for a more natural mouth. The internal representation accommodates multiple linguistic processes that co-occur on the face. The current system stores this information in a parameterized, retrievable format for customization in future sentence generation.

Current challenges include the representation and application of phonetic information and kinematics, which are necessary for convincing animation, but are distinct from phonemic information. An additional challenge is the representation and application of timing to support adverbs that manifest as "quality of motion."