Towards Computerized Synthesis of Nonmanual Signals in American Sign Language

Jerry Schnepp
DePaul University
School of Computer Science
Chicago, IL
jschnepp@cs.depaul.edu

Abstract

American Sign Language (ASL) is the language of the Deaf in North America. It is not simply a signed representation of English, but rather, a rich natural language with a unique structure, vocabulary, and grammar.

In 1965, William C. Stokoe published the first dictionary of American Sign Language. He was the first researcher to offer evidence that ASL was an independent language, separate from English. His dictionary describes signs only through a combination of hand configuration, location, and movement (Valli and Lucas, 2000).

During the 1970's, researchers began to examine the role of facial expressions used in conjunction with hand signals. Kegl and Wilbur found that *nonmanual signals* (NMS) have specific grammatical meaning, functioning as modifiers and separating certain clause structures (1976). Bridges and Metzger documented 49 NMS. They identified multiple ways in which NMS can function (1996).

• **Phonemic** - Facial expressions comprise an important component of ASL's NMS that are present in the phonemic features of the language (Dively, 2001; Reilly, 2002). These are not ancillary expressions that accompany syntax as in spoken languages, (Takeuchi and Nagao, 1993), but are essential elements of the language. For example, the only phonemic difference between the ASL signs LATER and NOT-YET is the presence of a TH- NMS.

• **Syntactic** – In a syntactic role, a facial expression is required to ask either a Yes/No or WH-question. Consider the three glosses WOMAN KISS-FIST FOOTBALL. With a neutral expression, these mean "The woman really likes football." With raised eyebrows and a head tilt, the meaning becomes, "Does the woman really like football?" With furrowed brows, the meaning becomes, "Which woman really likes football?"

• **Prosody and Turn Taking** - Similar to intonation and temporal pauses in spoken languages, facial expressions in ASL are used to convey prosody and turn taking. A signer will typically pause between sentences, dropping eye gaze and tilting the head downward. Likewise, the signer will hold gaze at the end of a sentence when expecting a response.

• **Emotional Affect** - There are close parallels in the roles of facial expression in ASL and spoken languages in conveying emotional affect. It is apparent when a signer is happy, angry, or puzzled simply by observing the facial expressions co-occurring with signing.

Although NMS are an essential part of the language, no previous research effort has addressed the problem of acquiring or storing NMS for use in displaying ASL through computer synthesized animation. Any digital representation must take them into account. It is simply

impossible to reproduce certain signs without incorporating NMS. Further, they are essential to forming questions.

By creating animations using 3D coordinates, independent data for each part of the body is collected and stored. The animations can be broken down to their essential elements and synthesized with other animations to create new signs.

It is advantageous to be able to add NMS to manual signs and use animations interchangeably. Moreover, it is important to isolate and synthesize specific facial elements (eyebrows, cheeks, mouth, gaze).

Current technologies promise rapid and precise representation of NMS. One of them is motion capture. However, the current technology is inadequate for this work. A NMS contains intricate and subtle nuances. The motion capture devices currently available have been designed for the video game and animated film industry and are best suited for recording gross motor movement. There are parts of the face that current motion capture devices cannot record, including the teeth and tongue. While they can capture the coordinates of specified targets, they do not provide the accuracy required for depicting NMS. Further, the amount of effort necessary to clean up the data is prohibitive.

Another alternative, lip syncing software, does not provide control of some facial elements, such as the tongue. In addition, lip syncing software does not allow for some of the key poses necessary to represent NMS not found in speech.

The challenge that lies ahead is a comprehensive automatic synthesis of facial animations that depict NMS. Initial manual attempts have been very time consuming. Especially important will be the development of an abstraction of NMS for reuse.

Future work involves developing the high level representation of NMS which will allow for discretization of facial elements and parameterization of each NMS. Also needed is the capacity to composite several NMS together, specifying different elements from each one, and specifying different times. An example would be when changing a frown to a smile, while holding the brows up.

Works Cited

[1] Bridges and Metzger. *Deaf Tend Your: Non-Manual Signals in ASL*. Silver Spring, MD: Calliope Press. 1996.

[2] Dively, V., 2001. *Signs without hands: nonhanded signs in American sign language*. In: Dively, V. et al (Eds): Signed languages : Discoveries from international research, International Conference on Theoretical Issues in Sign Language Research 6, Washington, DC, Gallaudet Univ. Press, pp. 62-73.

[3] Kegl, Judy and Ronnie B Wilbur.: "When does structure stop and style begin? Syntax, morphology, and phonology vs. stylistic variation in ASL" In: Mufwene, Salikoko S. et al. (eds): Papers from the *12.Regional Meeting, Chicago Linguistic Society*, April 23-25, 1976. Chicago, Ill. : Chicago Linguistic Society (1976) - pp. 376-396.

[4] Poor, G., Project Director, with P. Graybill, and D. Wilkins, *ASL Video Dictionary and Inflection Guide on CD-ROM*. National Technical Institute for the Deaf, NY, 2002.

[5] Stokoe, William, Dorothy Casterline, and Carl Croneberg . *A Dictionary of American Sign Language on Linguistic Principles*. Washington, DC: Gallaudet College Press, 1965, 1976 (2nd ed.).

[6] Takeuchi, A. and Nagao, N, 1993. Communicative *facial displays as a new conversational modality*. Proceedings of the conference on Human factors in computing systems 2003, Amsterdam, the Netherlands, 187-193.

[7] Valli and Lucas. *Linguistics of American Sign Language*. Washington, D.C.: Gallaudet University Press, 2000.