

# Modeling synchrony and co-occurrence for nonmanual signals in American Sign Language

Jerry Schnepf, Rosalee Wolfe, John McDonald  
DePaul University, Chicago, USA  
{jschnepf, wolfe, jmcDonald}@cs.depaul.edu

Any automatic system for translating spoken to sign language must be capable of portraying all aspects of the target language including facial nonmanual signals (NMS). This is difficult because facial nonmanuals can co-occur, each potentially influencing facial features in common (Neidle, Nash, Michael, & Metaxas, 2009) (Sandler & Lillo-Martin, 2006) (Wilbur, 2000) (Wilbur, 2009).

This matter compounds when affect is incorporated into sign synthesis. For example, both the WH-question marker and angry affect lower the brows and both the YES-NO question marker and happy affect raise the brows. If a signer asks a WH-question in a happy manner, there are two competing influences on the brows: one is a downward motion, and the other is upward (Reilly & Bellugi, 1996) (Weast, 2008).

Previous computational models hampered avatar technologies because they did not fully support synchrony and co-occurrence. Researchers were limited to choosing a single facial process instead of portraying all of the co-occurring influences (Platt & Badler, 1981), (VCom3D, 2012).

Our computational model overcomes this limitation by using linguistic notation as a metaphor. This notation can indicate co-occurring facial nonmanual signals that naturally maps to a software user interface. Because tiers in the interface are linguistically, not anatomically based, it is possible to specify nonmanual processes that influence facial features in common. The interface facilitates quick synthesis and editing of ASL via a signing avatar. It is abstract enough that animations can be drafted quickly, while still providing access to fine tuning details for a greater level of control.

We conducted tests to evaluate the clarity and acceptability of the synthesized ASL produced with this technique. Deaf participants viewed animations displaying combinations of competing nonmanual signals and affect. Results indicate that even when multiple processes simultaneously influence a facial feature, each individual process is perceived.

The computational model is potentially useful for language study. It provides researchers with a highly detailed level of control and a consistent repeatability unattainable by humans. Avatars can perform to exact specifications of timing, alignment and location, and never suffer fatigue. This allows researchers to synthesize utterances that conform exactly to a theoretical model, which can then be evaluated.

As a feasibility study, we created animations that displayed nonmanual and classifiers involving size. We portrayed the phrase “a {large, medium, small} cup of coffee”. For each classifier, participants viewed one version in which the nonmanual signal was consistent with classifier, one version where the nonmanual signal conflicted with classifier, and one version with no nonmanual signal.

Although animations without nonmanual signals were comprehensible, the data indicate that the nonmanual lends context to the classifier. The presence of a nonmanual increases perceived clarity. Notably, this occurred even in most cases when the nonmanual conflicted with the size classifier. This is consistent with a common participant remark, “You always need to have a nonmanual”.

Future work includes automatic synthesis of other language processes.

## References

- Neidle, C., Nash, J., Michael, N., & Metaxas, D. (2009). A method for recognition of grammatically significant head movements and facial expressions, developed through use of a linguistically annotated video corpus. *Proceedings of the Language and Logic Workshop: Formal Approaches to Sign Languages*. Bordeaux, France. Retrieved from [http://www.bu.edu/asllrp/papers/Neidle\\_ESLLI\\_2009.pdf](http://www.bu.edu/asllrp/papers/Neidle_ESLLI_2009.pdf)
- Platt, S., & Badler, N. (1981). Animating Facial Expressions. *Proceedings of the 8th annual conference on computer graphics and interactive techniques (SIGGRAPH 81)* (pp. 245-252). Dallas, Texas: ACM.
- Reilly, J., & Bellugi, U. (1996). Competition on the face: affect and language in ASL motherese. *Journal of Child Language*, 23, 219-239.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge: Cambridge University Press.
- VCom3D. (2012, 09 27). *Sign Smith 3.0 user manual*. Retrieved April 19, 2012, from [www.vcom3d.com: http://www.vcom3d.com/docs/Sign\\_Smith\\_Studio3.0\\_UsersGuide.pdf](http://www.vcom3d.com/docs/Sign_Smith_Studio3.0_UsersGuide.pdf)
- Weast, T. (2008). *Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows*. The University of Texas at Arlington: ProQuest Dissertations and Theses. Retrieved 03 11, 2010, from <http://search.proquest.com/docview/304837448?accountid=10477>
- Wilbur, R. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In K. Emmorey, & H. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 215-241). Mahwah, NJ: Lawrence Erlbaum.
- Wilbur, R. (2009). Effects of Varying Rate of Signing on ASL Manual Signs and Nonmanual Markers. *Language and Speech*, 52(2/3), 245-285.