Inferring biomechanical kinematics from linguistic data: A case study for role shift

Rosalee Wolfe, John C. McDonald, Robyn Moncrief, Souad Baowidan, Marie Stumbo
DePaul University
Chicago IL, USA
E-mail: {wolfe,jmcdonald}@cs.depaul.edu, { rkelley5, sbaowida, mstumbo} @mail.depaul.edu

Over the past two decades, researchers have made great strides in developing avatars for use in Deaf education (Efthimiou & Fotinea, 2007), automatic translation (Elliott, Glauert, Kennaway, & Marshall, 2000), (Filhol, 2012) interpreter training (Jamrozik, Davidson, McDonald, & Wolfe, 2010), validation of transcription (Hanke, 2010), and improving accessibility to transportation and government services (Segouat, 2010) (Ebling, 2013) (Cox, et al., 2002). Creating lifelike, convincing motion continues to be one of the key goals of signed language synthesis research. Avatars that sign with smooth, natural movements are easier to understand and more acceptable than those that move in an unnatural or robotic manner.

Motivation

Current research efforts in sign synthesis either use libraries of motion captured signs (Awad, Courty, Duarte, Le Naour, & Gibet, 2010) or libraries of sparse key-frame animations transcribed by artists (Delorme, Filhol, & Braffort, 2009). Entries from libraries are then procedurally combined to produce longer signed utterances. An excellent review of the current literature on sign synthesis can be found in (Courty & Gibet, 2010).

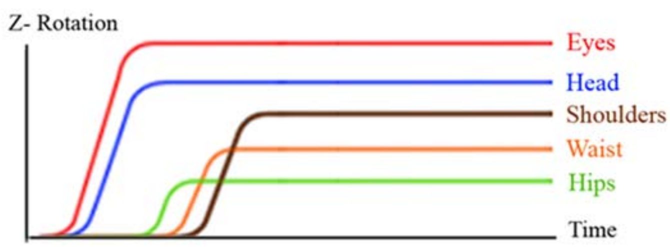
Sign synthesis based on motion capture produces outstanding natural motion. The myriad tiny and subtle details in the data create smooth, naturally flowing movement in an avatar. However, it is difficult to maintain the same naturalness in the transitions when modifying the data to accommodate new utterances. The high temporal density of captured detail that creates the beautiful movement also requires substantial resources to modify.

Applying linguistic rules to modify animation is easier with sparse sets of keys that correlate well to the structure of linguistic models. Unfortunately the ease of modification is offset by a lack of realism in the animation. The linguistic parameters contain no information about the subtle body movements which are not considered to be linguistically significant, but are nonetheless required for natural motion.

The ideal system would combine the best aspects of both approaches. It would support ease of key modification while still producing natural, lifelike motion. This presentation details a step towards a new method that automatically layers biomechanical, sublinguistic movement under the motion dictated by linguistic data. The approach is designed to improve the quality of avatar motion without requiring researchers to acquire more data.

Science

This presentation will discuss the theory of the new approach in the context of generating role shifts. In a role shift, a signer uses a body turn to assume the role of a protagonist in a constructed dialog (Lillo-Martin, 2012). From the linguistic information, an animation system can compute a global orientation that dictates the avatar's pose when assuming a role. Previous work (McDonald, et al., 2013) used range of motion data to distribute the global orientation down the spinal column as local rotations, but the timing of the transitions proved problematic. In a turn, the transition begins with the eyes, followed by the neck, hips, spine, and shoulders. The eyes and head complete their rotation before the remaining linkage begins movement.



A nontrivial problem occurred with the previous work because the eyes and head were descendants of the hips in the transformational hierarchy. Whenever the hips rotated, the eyes and head rotated in concert. This induced an additional rotation on the head that was not the intent of the animator.

A traditional approach for holding objects in a given orientation is to add *lookat* constraints that apply a global rotation and ignore the transformation hierarchy. With lookat constraints, the timing and orientation of the eyes and head were preserved throughout the onset and duration of the role shift. Unfortunately, it proved difficult to blend between the global rotation in the lookat constraints and the local rotation used when the avatar is in the narrator role. The visual result was a visible head bobble at the end of the transition.

The transformational hierarchy also disrupted the staggered timing in the shoulders, which are supposed to remain stationary while the hips are starting their rotation. But because they are descendants of the hips, they began their rotation synchronously with the hips, defeating the attempt to stagger the timing.

Features

The new system uses no lookat constraints, and all joints remain in the transformation hierarchy. In a preliminary step, the system computes the transition of each joint as a global orientation. It then computes compensatory motion to implement timing as animation keys cast in local coordinates.

The present implementation is applied to a simple figure with controls to change the global orientation of the torso and the speed of the transition. We invite participants to a hands-on evaluation of the system at the conclusion of the presentation or any time during the course of the workshop.

Bibliography

Awad, C., Courty, N., Duarte, K., Le Naour, T., & Gibet, S. (2010). A combined semantic and motion capture database for real-time sign language synthesis. *Intelligent Virtual Agents,*, 432-438.

Courty, N., & Gibet, S. (2010). Why is the creation of a virtual signer challenging computer animation? *Motion in Games*, 290-300.

Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002). TESSA: a system to aid communication with deaf people. *Proceedings of the fifth international ACM conference on assistive technologies (ASSETS 02)* (pp. 205-212). Edinburgh, UK: ACM.

Delorme, M., Filhol, M., & Braffort, A. (2009). Animation generation process for Sign Language synthesis. *International Conference on Advances in Computer-Human Interaction (ACHI '09)* (pp. 386-390). Cancun, Mexico: IEEE.

Ebling, S. (2013). Evaluating a Swiss German Sign Language Avataramong the Deaf Community. *Third International Symposium on Sign Language Translation and Avatar Technology (SLTAT).* Chicago, IL.

Efthimiou, E., & Fotinea, S.-E. (2007). An envrionment for deaf accessibility to education content. *International Conference on ICT & Accessibility*, (pp. GSRT, M3. 3, id 35). Hammamet, Tunisia.

Elliott, R., Glauert, J. R., Kennaway, J. R., & Marshall, I. (2000). The development of language processing support for the ViSiCAST project. *Proceedings of the fourth international ACM conference on Assistive technologies (ASSETS 2000)* (pp. 101-108). Arlington, VA: ACM.

Filhol, M. (2012). Combining two synchronisation methods in a linguistic model to describe Sign Language. *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, 192-203.

Hanke, T. (2010, June 14-16). *An overview of the HamNoSys phonetic transcription system.* Retrieved December 28, 2011, from Sign Linguistics Corpora Website: http://www.ru.nl/publish/pages/570576/slcn3_2010_hanke.pdf

Jamrozik, D. G., Davidson, M. J., McDonald, J., & Wolfe, R. (2010). Teaching Students to Decipher Fingerspelling through Context: A New Pedagogical Approach. *Proceedings of the 17th National Convention Conference of Interpreter Trainers*, (pp. 35-47). San Antonio, TX.

Liddell, S. (2003). *Grammar, Gesture, and Meaning in American Sign Language.* Cambridge, UK: Cambridge University Press.

Lillo-Martin, D. (2012). Utterance reports and constructed action. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign Language: An International Handbook HSK 37* (pp. 365-387).

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., & Berke, L. (2013). Toward Lifelike Animations of American Sign Language: Achieving Natural Motion from the Movement-Hold Model. *Third International Symposium on Sign Language Translation and Avatar Technology (SLTAT 2013).* Chicago, IL.

Segouat, J. (2010). *Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel.* Doctoral Dissertation, Université Paris Sud, Orsay, France. Retrieved February 21, 2013, from http://hal.upmc.fr/docs/00/60/21/17/PDF/these-segouat2010.pdf