

Gesture Modelling for Linguistic Purposes

Guillaume J-L Olivrin

Meraka Institute*, City of Tshwane, SA
golivrin@meraka.org.za

Abstract. The study of sign languages attempts to create a coherent model that binds the expressive nature of signs conveyed in gestures to a linguistic framework. Gesture modelling offers an alternative that provides device independence, scalability and flexibility for the annotation and modelling of linguistic phenomena. This paper presents the requirements and initial experiments to build an input method editor for sign languages. The objective is to design interfaces backed up by computational methods that can infer and use linguistic guidance to model sign language gestures. This in turn produces a linguistically annotated corpus of gesture animations.

1 Introduction

Information in sign language is scarce on computational platforms. Digital video with emerging standards like MPEG annotation tools have not yet made computers a convenience for sign language users. Communities like Sutton Sign Puddles [4] continue to grow showing that there is a need for computational tools to acquire sign languages on computers. This paper introduces new entry mechanisms for sign language users and proposes to couple direct gesture modelling together with an annotation life cycle. Related works and requirements for computer-based sign language acquisition are first reviewed. Then designs and early experimentation on 3D humanoid interfaces are presented together with an approach for deriving linguistic annotation.

1.1 Related Work

Capturing sign language has traditionally been done through graphical forms, video and motion capture. For computer animation purposes, eSign [5] offers a notation - HamNoSys - accessible through interfaces and controls directly inspired by sign languages such as pointing to locations on a 2D character representation. A promising feature of the eSign editor is the introduction of Part of Speech as linguistic annotations for word categorization when adding a sign in the lexicon. Unfortunately it is not possible to layer and span this annotation scheme over multiple signs. To fully describe sign language phenomena

* In collaboration with the University of Stellenbosch [1, 2] and Thibologa Sign Language Institution [3]

where iconicity, nominalization and facial modalities are involved it is necessary to go beyond gloss and word categorization [6]. Video annotation tools such as SignStream and ELAN can annotate the time line and point to some visual elements. Beyond recent video corpus [7] there is a quest for modelling sign languages virtually based on signing space [8] and classifier predicates [9]. Providing spatial and temporal annotations is necessary to study phenomena like agreement and co-reference. A framework like MPEG-7 can then provide the basis for spatio-temporal reasoning and computational linguistics. How can we provide appropriate gesture modelling and useful annotation strategies to support the study of sign languages?

1.2 Requirements

From a linguistic point of view, the need for annotations depends on the research question. A "linguistically driven approach" is characterized by the introduction of specific expert knowledge during the language acquisition phase. Lenseigne also proposed that sign prediction be based on syntactic models boost the recognition process [7]. In order to study "word-order" in South African Sign Language, a simple video annotation approach might be sufficient. However since word order is intimately connected with the spatial agreement rules, then time-linear video tracks are not practical to represent such syntactic phenomena. Iconicity [7] and classifiers [9] have rather to be studied on a 3D model of signing space.

Ethnographic considerations are concerned with how data is acquired (c.f. [10]). This must be done faithfully by sign language experts. Or should it? Users of sign languages often need to capture and document their language for their own (e.g. for teaching purposes [3]) . Hence the use of computerized input method editors (IME). IME should reduce the effects of code-switching as well as the influence of informants and language experts on data. It should be possible for all Deaf signers to be involved in the data collection and the documentation of their language regardless of their level of literacy.

Device independence and scalability are two requirements inspired from the field of gesture recognition. T. Huang [11] noted that "the analysis of gestures increasingly relies on computer vision techniques as an unencumbered alternative to glove- and stylus-based approaches". On scalability and level of details, studies that use faster sampling and multiple modalities like video and gloves and infrared as in [12] already makes it easier to study particular details of the language. Virtual avatars provide a better level of abstraction and level of detail for observing all aspects of gestures at all scales. To achieve device independence in gesture recognition, Yan [13] also prescribes mapping the gestural observations to higher level predicates that are often inspired by linguistics notations. Modeling character animations cognitively adopts similar notions of higher level gestural predicates ([14–16]). Scalability in perception and in the level of details (local and global changes) is necessary to disambiguate signs and their meanings. Precise observations at small scale as well as at large scale are needed simultaneously globally and concurrently. This is where a parametric model of

signed content is well-suited for linguistic studies. Building a working scheme of predicates and parameterized gestures is one way to model sign language. Various generative models can then be offered and, depending on the context, a generative model can be invoked to compose sign language. Direct modelling is difficult so the ability to constrain (physics based animation, collision detection, anthropo-morphological constraints) is necessary to guide direct modeling operations and also to make sign credible with reference to a signer's profile. Anthropo-morphology i.e. the effects of physiognomy and the variations of signing styles (use of signing space, energy, speed and stress) between sign language users characterize one's manners of signing. Finally there are other important requirements for descriptive annotation tasks prescribed in [8]. Can gesture modelling facilitate the acquisition of sign language? Can sign language linguistics guide the modelling of gestures?

2 Sign Language Modelling

First we look at building new sign language acquisition modules for a signing avatar gesture editor as shown in Fig.1 and developed at Stellenbosch University [1, 2]. The second aspect of the research is to computationally derive linguistic annotations from such modelled animations.

2.1 Interfaces

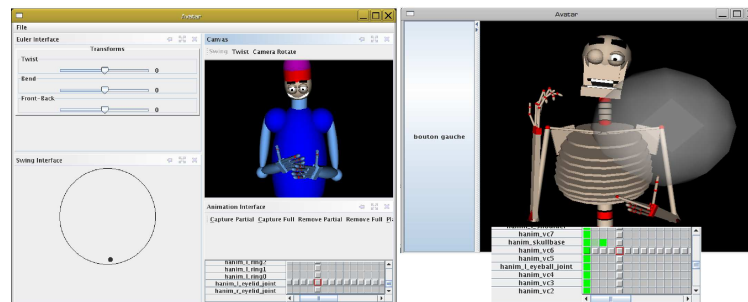


Fig. 1. Interface modules for capturing sign language [2].

The research first focuses on humanoid interfaces to provide better controls and features for modelling signs. Linguistic phenomena like agreements and co-referencing can be integrated in the interface to guide the acquisition of sign language. Various user interfaces strategies are studied in this research to provide better controls over sign language composition : chord keyboards, trackball, mouse gesture recognition, 2-hands interfaces, tactile screen, head and eyes tracking, capture of facial expression or traversable expression maps. Each interface

provides new input strategies over a 3D avatar model. Eventually these interfaces can be combined in a multimodal fashion for even richer composition strategies. There are many constraints associated to modelling gestures in a 3D fashion like in Poser or in VSign. If the controls were simpler, interaction with a human model should be a natural act closest to the ideas of "Universal design" and contact sign language. Following the capture of experimental data using a direct gesture modelling software, VSign [17], our sign language modelling team made suggestions to facilitate their modelling experience : physics-based animation to help model precise contacts and positioning through proprioceptive sensors and actuators ; the ability to control more than one joint at a time via inverse kinematics and multiple direct manipulations, e.g. using tactile screens ; guides and constraints in signing space and signer's profile (snap-grids, anthropo-morphology).

It is believed that the semantic MPEG-7 annotation scheme that describe sign language phenomena will support gesture animations. Current motion tweening techniques makes model act rather robotically. Reasoning on agreements, transitivity, directionality, stress and sign amplitude can be integrated in the composition process.

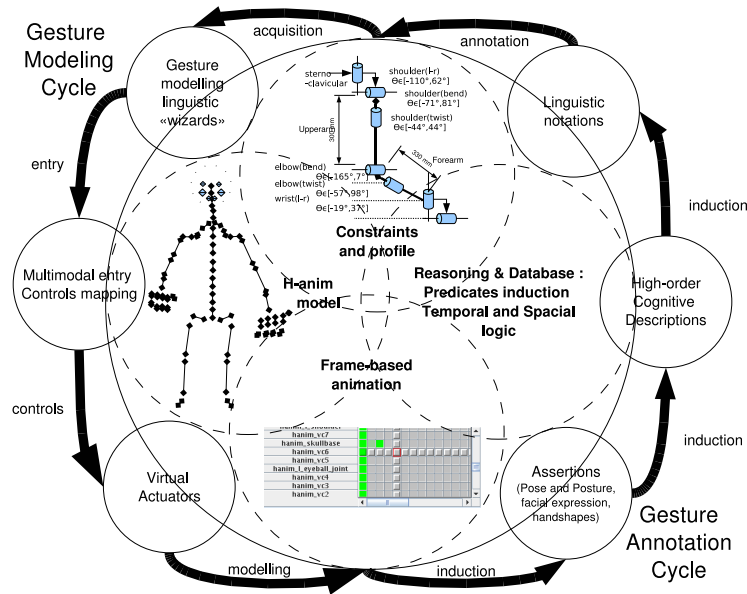


Fig. 2. Gesture modelling and annotation life cycles.

2.2 Gesture Annotations and a modelling life cycle

Modelled gestures cannot yet be exported with the sign editor in Fig. 1. Experiments have been conducted with data captured using VSign[17]. The idea is to associate the annotation directly on the avatar model : a knowledge representation that semantically links parts-of-gestures to linguistic notations. Series of derivations and inductions form the cognitive transformation process which transfers the joint values inherited from the animated humanoid model into predicates. Further assertions are made by reasoning on these basic predicates: possible handshapes, directions, movement, facial expressions (eye squint, pinched lips, puffed cheeks). More reasoning follows with the induction of higher-order predicates like symmetries, repetitions and type of contacts occurring in gestures. All these observations can be used to annotate the model and make predictions and decisions on which linguistic constructs are being modelled. These constructs can also be selected during the sign modelling phase using language wizards. The annotation of gestures remains flexible and unobtrusive since it is also part of the modelling process and the induction cycle does a lot of guesswork. Because we are planning on an integrated gesture modelling cycle where annotation takes place alongside the acquisition of the language, and where linguistic guides are called while modelling a sign, the transformational process consists in applying successive rules and unifications to the initial humanoid animation model joint and pose values and rewrite the model with parameterized predicates. Basic predicates will often correspond to classic elements found in notation systems such as handshapes, direction of movement, facial expression elements (eye squint, pinched lips, puffed cheeks). MPEG-7 is considered for the output format of the sign language composer together with XML models [18] and the linguistic annotation found in [19]. Huenerfauth [20] has published a formalism to represent coordination for sign language animations that can also be implemented within the MPEG-7 schematics and as a standalone output format.

3 Conclusion

Overall, an original approach to building a sign language linguistic corpus has been presented. This linguistically driven approach suggests new gesture modelling strategies which combine interfaces with annotation inference cycles. Future work include the predication and application of derivation rules on H-anim animations (Fig 2), the development of annotations export module for the sign editor (Fig 1) and the design of new interfaces and linguistic "wizards" for acquiring South African Sign Language.

References

1. Zijl, L.V., Fourie, J.: Design and development of a generic signing avatar. In: Proceedings of Graphics and Visualization in Engineering, Florida, USA (2007)
2. Potgieter, D.: Signing avatar: Sign editor. <http://www.cs.sun.ac.za/~dpotgieter> (2006)

3. Mohlamme, I.: Thibologa sign language institution. <http://www.thibologa.co.za> (2007)
4. Sutton, V.: SignWriting: Sign Language Transcription. The Center For Sutton Movement Writing, Inc., California, USA. (September 2006)
5. Thomas, H., Popescu, H., Ehrhardt, U., Davies, B., Thomas, N., Sheard, M., Glauert, J., Elliott, R., Tryggvason, J., Schmaling, C., Wells, M., Zwitserlood, I.: Animating Sign Language : The eSIGN Approach. IST program EU FP5 : Essential sign language information on government networks, EU. (September 2003)
6. Cuxac, C.: La langue des signes française - les voies de l'iconicité. Ophrys (2000)
7. Lenseigne, B.: Intégration de connaissances linguistiques dans un système de vision, application à l'étude de la langue des Signes. PhD thesis, Toulouse, FR (2004)
8. Filhol, M., Braffort, A.: A sequential approach to lexical sign description. In: LREC 2006, Orsay, FR (2006)
9. Huenerfauth, M.: Spatial representation of classifier predicates for machine translation into american sign language. In: LREC 2004, Lisbon, PT (2004)
10. Crasborn, O.: A guide for starting sign linguists. Technical report, Sign Language Linguistics Society, Amsterdam, NL (2005)
11. Huang, T., Pavlovic, V.: Hand gesture modeling, analysis, and synthesis (1995)
12. Rezzoug, N., Gibet, S., Courty, N., Heloir, A., Kamp, J.F., Multon, F., Pelachaud, C., Gorce, P.: Hugex: Humanoides doués de gestes expressifs de communication. IEEE/SMC International Conference on Systems, Man and Cybernetics (2006)
13. Yan, D.: Device independence and extensibility in gesture recognition. IEEE Virtual Reality (2003)
14. Ma, M.: Virtual human animation in natural language visualisation. (2006)
15. Funge, J.: Cognitive modeling: Knowledge, reasoning and planning for intelligent characters. AI games (1999)
16. Huang, Z.: Step: A scripting language for embodied agents. In: Lifelike Animated Agents. (2002)
17. Pyfers, L.: VSign Builder Manual. Pragma, Hoensbroek, NL. (2002)
18. del Puy-Carretero, M., Oyarzun, D., Ortiz, A., Aizpurua, I., Posada, J.: Virtual characters facial and body animation through the edition and interpretation of mark-up languages. In: Computers & Graphics. Number 29 in Computers & Graphics, San Sebastian, Spain, VICOMTech Research Center (2005) 189–194
19. Fernandez-Carbajales, J.M., Grupo, F.M.: Description-driven generation of 3d humanoids within authoring 744 v. In: IEE Multimedia Week 2005, Spain, Universitaria de Cantoblanco (December 2005)
20. Huenerfauth, M.: Representing coordination and non-coordination in american sign language animations. In Francis, T., ed.: Behaviour & Information Technology. Volume 25 of 4. (July 2006,) 285–295