# Sign Language Synthesis: Skeleton modelling for more realistic gestures.

Maxime Delorme
LIMSI/CNRS
B.P. 133, 91403
Orsay Cedex, France +33 6 10 57 39 71
Maxime.Delorme@limsi.fr

**ABSTRACT**

In this paper, I describe the summary of what I plan to do for my PhD thesis on sign language synthesis. The modelling of the skeleton of virtual characters would lead to improvements in the quality of animations. Thus the first step is to study how to build such a model, by analysing corpora data and integrating biological data. Then, we will implement the model and confront it to users to evaluate it.

**Categories and Subject Descriptors**

H.5.1 [**Information Systems**]: Information interfaces and presentation - *multimedia information systems.*

**General Terms**

Design, Experimentation, Human Factors.

**Keywords**

Sign language generation, Signing avatars, Statistical modelling

## 1. CONTEXT & MOTIVATIONS

Sign Languages (SL) are considered to be the most natural way to communicate for deaf people and therefore supposed to be the best way. Since deaf people are not always able to read plain text, we choose to display SLs on screen to give them access to every piece of information. We could use videos of people signing isolated signs to make complete utterances. However, such a method lacks flexibility because of image rights and absence of modularity. Moreover, SLs are natural languages and not just a transcription of a spoken language. They have a proper syntax and a proper lexicon and though are not simply a code like braille would be. We choose to generate animation for a dedicated humanoid called virtual signer (VS).

There are, at the moment, existing sign generation systems such as the European project eSIGN [4] or the Greek ILSP system [7]. These systems are generally relying on parametric models such as HamNoSys [12]. Parametric models describe the signs as tuples of values taken in finite sets. Studies [5] show that parametric models are insufficient to describe the whole majority of signs in French Sign Language (LSF). Based on a geometric and temporal approach [6], our team is working on a new generation systems covering the whole LSF lexicon. However, generation from geometric models often leads to perfect movement in the mathematical sense. And thus, the animation generated is very often mechanical and robotic.

Improving the quality of the animation will lead to a more understandable VS, and thus, to a better efficiency in communicating with deaf people.

My work as a PhD student is to improve the overall quality of the animation by making them more realistic. This is done by studying and modelling the skeleton of the VS in human-like ways. The study is oriented along two main axes:

- Firstly by studying the anatomy of the human skeleton and by applying biomechanical and kinesthesic constraints on the VS. For instance by preventing the wrist of the VS to be bent more than 90°.
- Secondly, by studying motion capture corpora and extracting statistical models. From these two separate studies, I will build a general model of the skeleton allowing a more realistic management of movements. The model will then be applied to the current animation system and be evaluated.

The next section deals with realistic modelling of the skeleton, the approaches covered in the litterature and the way I am doing my studies. The third section deals with the later parts of the PhD work: implementation and evaluation. The last section is a brief conclusion and what, in my sense, I will gain from the doctoral consortium.

## 2. REALISTIC MODELS OF THE SKELETON

The human skeleton as been deeply studied in the medical field. Lately, these studies have been extended to the domain of computer science for many purposes. The field of computer graphics have seen a particular interest in these researches with the increasing demands of realism in special effects and animation films. On one hand, the modelling of the skeleton including exclusively bones and joints has been extensively studied by many teams and have been successfully summarized up until 1993 by Badler and al. [1]. On the other hand researches on the biomechanical functioning of the skeleton have been made quasi-exclusively for medical purposes and not for computer animation, leading to very costly solutions. This is mainly because computer animation is actually ruled by data-driven approaches such as motion capture [9]. These approaches are cheaper in terms of computation time and provide good results as far as a professional animator checks the result of the process. We then need to find an intermediate between the very realistic models of biomechanical simulation and data centered approaches which lack of modularity.

The skeleton we consider is a simple sequence of joints (elbow, shoulder, wrist, etc.) connected by bones (upper arm, forearm, etc.). The first step of the study is the statistical analysis of a motion capture database [3]. Because of the large range of motions included in the database, we assume that the motion capture data studied is representative of the motions of a human. Then I extract, for each joint of the upper body, the limit values on each axis of rotation: for instance, to what extent the wrist can be bent. Aside from the limits, I also compute contribution of each joint in composed sequences. For instance if we want to bend forward the back of the VS about 80°, how much should we rotate each joint of the spine to make the posture realistic ? The upper-body of the skeleton has been modelled this way, from the spine to the wrists.

The second step of the study is to integrate various biological data to our model to complete the gaps such as the absence of wrists/hands information on the MoCap database. These data are comprised of biomechanical and clinical data [2], anthropometry data [11], biomechanical modelling [8] and anatomical data [10]. The integration of the two previous studies will lead me to build a single final model for the skeleton of the VS.

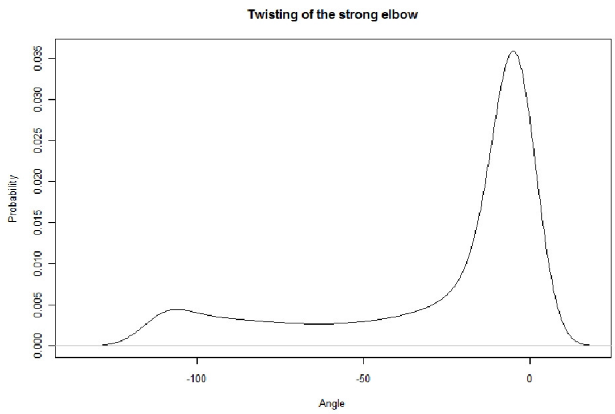


**Figure 1. Density of probability for the twisting of the strong (right) elbow.**

## 3. IMPLEMENTATION & EVALUATION

The model will next be integrated to the animation generation system to calibrate the skeleton. The system is currently taking sign descriptions as input. Each of these description gives us sets of constraints on the joints of the skeleton through time. For instance, the sign *WARDROBE* in LSF is made as shown on figure 2. We can see for each arm three key-postures (dots), and between them, two transitions (arrows). The key-postures are straight-forward constraints on the body ("place your wrist here"), le transitions are interpolation from a set of constraints to the next.

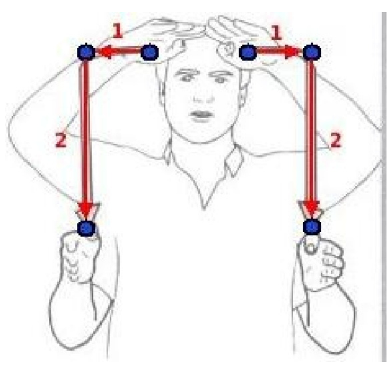


**Figure 2. Description of the sign *WARDROBE* in LSF.**

Currently, the system takes these constraints and applies basic CSP/optimization methods to solve them (Inverse Kinematics). Since constraint solving problems are underspecified problems (i.e. from a constraint set we can find an infinity of solution), the solutions given by the methods are generally very far from a natural posture.

The statistical model I build will be used as a selection system for generated postures of the skeleton. The system will generate, for each constraints set, a number of solutions. Each of these solutions will be assigned a score depending on how likely they are to be found in the

MoCap database. The final posture will be the one with the highest score, assuming that it will be the most natural. On the other hand, every unconstrained degree of freedom of the hands will be managed by the biomechanical models. Since the hands are the most constrained parts of the body in sign language synthesis we think that we can greatly improve realism by modelling inter-fingers reactions.

Once the model will be completely integrated to the animation generation system, I intend to evaluate and validate it by confronting the signs to users. The users should be both deaf and hearing persons, as well as signers or not. The first step will be to show animations build with and without the model, and to ask for a measure of "naturalness" in the generated motion. Then I will show to deaf signing people results of the generation with and without the model and ask for a measure of understandability.

Another clue on evaluation I would like to study is the possibility to feed generated animations to a sign language recognition system and see whether the signs are recognized as LSF signs. However, we the recognition systems to be developed and also evaluated before doing so.

## 4. CONCLUSION & PROSPECTS

The work I intend to do during my PhD will aim to build a model of the skeleton for VS. This model will be sufficiently realistic to palliate the robotic motion problems but also quick to compute to allow quasi-real time generation of animation.

The overall statistical study has been completed and is currently being implemented to be tested. The biomechanical parts of the model are still being studied but we expect finishing this study before the end of the implementation of the statistical scoring system.

The implementation phase will be followed by a calibration process that will more than probably take us some time. We don't know yet how we are going to proceed to calibrate the various parameters (such as the number of posture to generate for the scoring system).

In the end of my PhD, I want to release a set of tools for easily manipulating and animating the skeleton of the signing avatar. Hopefully, this work will be the first step into more and more realistic tools for the signing avatars, but also for general purpose animation.

## 5. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n°231135.

### 6. REFERENCES

[1] Badler, N.I., Phillips, C.B., Webber, B.L.: Simulating Humans, Computer Graphics Animation and Control. Oxford University Press (1993).

[2] Bogduk, N., Mercer, S.: Biomechanics of the cervical spine. I: Normal kinematics. *Clinical Biomechanics* vol.15, 633-- 648 (2000).

[3] Carnegie Mellon University Graphics Lab Motion Capture Database : http://mocap.cs.cmu.edu/

[4] Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I., Safar, E.: Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society (UAIS),* Special issue 6/4, 375--392. Springer (2008)Conger., S., and Loch, K.D. (eds.). Ethics and computer use. *Commun. ACM 38*, 12 (entire issue).

[5] Filhol, M., Braffort, A.: A sequential approach to lexical sign description, In: Workshop on sign languages, LREC 2006, Genoa, Italy (2006).

[6] Filhol, M.: Zebedee: a lexical description model for Sign language synthesis. LIMSI Report (2009).

[7] Fotinea, S.-E., Efthimiou, E., Caridakis, G., Karpouzis, K.: A knowledge-based sign synthesis architecture. Universal Access in the Information Society (UAIS), Special issue 6/4, 405--418. Springer (2008).

[8] Lemay, M.A., Crago, P.E.:A dynamic model for simulating movements of the elbow, forearm, and wrist. *Journal of Biomechanics*, vol 29, no 10, 1319--1330 (1996).

[9] Monzani, J-S. Baerlocher, P., Boulic, R., Thalmann, D.: Using an Intermediate Skeleton and Inverse Kinematics for Motion Retargeting. *Eurographics 2000*, vol 19, Issue 3 (2000).

[10] Musculoskeletal Radiology, Muscle Atlas, University of Washington, Department Of Radiology, http://www.rad.washington.edu/academics/academic- sections/msk/muscle-atlas

[11] NASA, Man-Systems Integration Standards, vol 1, section 3, *Anthropometry and Biomechanics*, http://msis.jsc.nasa.gov/sections/section03.htm

[12] Prillwitz, S., Leven, L., Zienert, H., Hanke, T., Henning, J., HamNoSys Version 2.0: Hamburg Notation System for Sign Languages - An introductory guide. In: International Studies on Sign Language and the Communication of the Deaf, vol 5., Hamburg (1989).