

Real-time 3D gesture visualisation for the study of Sign Language

Roman Miletitch
GestuelScript (ESAD Amiens)
roman.miletitch@gmail.com

Claire Danet
GestuelScript (ESAD Amiens)
claire.danet@gmail.com

Raphaël de Courville
GestuelScript (ESAD Amiens)
raphael.de.courville@gmail.com

Patrick Doan
GestuelScript (ESAD Amiens)
pdoan.atelier@gmail.com

Morgane Rébulard
GestuelScript (ESAD Amiens)
morgane.rebulard@gmail.com

Dominique Boutet
UMR 7023 SFL, CNRS, Université
Paris 8 et Université EVE
dominique_jean.boutet@orange.fr

This work constitutes a contribution to the emergence of a common writing for French Sign Language in a graphical or even a typographical framework. In this article we present ThirdEye, an interactive visualisation tool designed for movement analysis. As a research tool, ThirdEye's main objective is to help study the importance of movement in the production of meaning within the context of sign language. In this paper, we will show why and how this device was conceived and how it can help to shed light on the relationships between sign language and writing.

Gesture visualisation. Sign Language. Motion capture. Writing. Graphical system. Iconicization.

1. INTRODUCTION

French Sign Language (FSL) is the first mean of communication of the French deaf community representing around 120.000 people in France. A law passed in 2005 recognised it as a full language, which was the starting point of broader recognition from the public and its use in public school system and administration. Signed languages (SL) are analogical, visual-gestural and multilinear (meaning that it allows for simultaneous transmission of several pieces of information) languages, thus distinct from vocal languages which are arbitrary, acoustic-vocal and monolinear. Up to now, due to this complexity, no satisfactory writing system was created and yet a Sign Language (SL) writing would offer deaf people the conditions for an unprecedented cultural enrichment.

One could think that French in its written form would fill that gap. However, in reality, it is very difficult for most deaf people to learn how to speak or even to read French and studies have shown how a harmonious development of conceptualisation relies on a SL-based education (Courtin 2002).

Not having the same roots and modality than their vocal counterparts, signed languages cannot be written with existing scripts. A writing system needs to be engineered specifically; one that would fit

their peculiarities: grammar, vocabulary and multilinearity among others. However, most endeavours in this direction ended up as codes for linguists rather than a real practical writing that could be used everyday by the deaf community.

The distinctive feature of our project is that it is based on a graphical approach to Sign Language, which aims at observing similarities between the gesture that signs and the gesture that writes. We believe that in this particular case, the language and its writing could have a lot more in common than vocal languages and their own scripts. Incidentally, this approach may help us to institute a relevant methodology for graphical research in general.

Our core hypothesis is as follows: the realization of the gestural sign casts traces in space that have a scriptural quality. Our study departs from the traditional segmentation of sign language in parameters (hand shape, movement, location, orientation — which direction the palms are facing during a sign — and facial expression) studied by Stokoe (1976) and Cuxac (2000) among others. To begin, we chose to focus on gesture as a whole, omitting all non-manual elements of the sign. To this end, we created ThirdEye, an interactive visualisation tool designed for movement analysis within the context of Sign Language. Its first purpose will be to assert how much of the meaning

is still present when you only keep the movement parameters in Sign Language.

In the next section we will show some of the various means of capturing movement and its use in linguistic prospects. Then we will describe ThirdEye, how it works, what it does as well as the user experience itself. Finally, we will present our test methodology, the language games that we experiment on with ThirdEye and the perspectives for this device.

2. LITERATURE REVIEW

2.1 Motion capture

Real-time Motion capture (mocap) has now become a commonplace technology. It is implemented in various ways in many consumer electronics (mostly mobile phones and videogames) and its wide availability as of-the-shelf hardware yields a lot of interest into both researchers and DIYers communities around the world (Chung Lee 2008). However, precise, affordable, real-time, non-bulky full body motion capture and hand shape detection, is not there yet. Let's review the technologies available today for gesture-related studies.

2.1.1. Commercially available mocap solutions

Videogame motion controllers have become an object of interest for researchers (Vaughan-Nichols 2009) as they provide a cheap of-the-shelf solution for motion capture.

Optical human motion capture is traditionally used in the cinema and game design industries. It usually requires pose estimation algorithms to resolve ambiguities and occultations (Moeslund, & Granum 2001) and combine the signals from several cameras. This makes it an inadequate solution for real-time applications.

Inertial measurement units (IMUs) combining accelerometers, gyros and magnetometers seem to provide a viable alternative, especially with the recent addition of efficient sensor fusion algorithms (Madgwick, Harrison & Vaidyanathan 2011) though they are not yet as widely available as their console accessories counterparts.

2.1.2. Mocap applied to linguistics

In the field of sign language studies, the question of motion analysis is a crucial one, whether it is for segmentation (Lefebvre-Albaret 2010) or for automated translation (Starner 1995).

The work of Thad Starner (1995) at MIT and projects done at the Center for Accessible Technology in Sign show that it is possible to get

accurate recognition of the signs by a computer within a limited vocabulary of signs (Zafrulla 2010).

Furthermore, François Lefebvre-Albaret (2010) showed that the dynamic of the movement can be a good indicator for the automated segmentation of the signs.

These studies show the efficiency of the movement parameter as a differential indicator in natural language processing (human-to-computer communication). It still has to be shown if this holds true in human-to-human communication.

2.2 Writing gestures

To this day, a usable writing system for signed languages still doesn't exist as such. The only examples of writing available to us are therefore tied to vocal languages. The later exhibit a great variety, as Clarisse Herrenschildt (2007) showed, demonstrating that relations between language and writing depend, among other factors, on the type of script used: logogrammatic, consonant or alphabetic. In logogrammatic scripts, understanding is compulsory as there is no connection between the written form and phonetics, reading and comprehension are closely intertwined in the script. Consonant scripts on the other hand require from the reader knowledge of the language in order to complement the information given by the text with unwritten vowels. Finally, the readers of a complete alphabetical script can decipher a text even if they don't know the language or understand the meaning. So, depending on the type of script (expression form — sound — or content form — concept — (Hjelmslev & Leonard 1996) comprehension and reading may switch places; comprehension necessarily comes first with logograms which can be read regardless of the language used, whereas reading an alphabetical script — even clumsily (as the phonology of each language orients the actual pronunciation of the graphemes) — can be done without understanding the content.

Types of relationships between writing and its object are either an arbitrary convention between alphabet and sound, or aspire to an analogy between the logogram and the concept. However, whether it goes from sound to concept (alphabet) or from concept to sound (logogram), reading is always confronted to the arbitrarily stemming from the difference in modality between vocal languages and the script. Thereby, we see a wide gap between the vocal language that relies on a vocal-auditory modality and its writing totally dependant on a gestural-visual canal, regardless of the vocal language and the script.

Signed languages (SL) are iconic languages; the shapes of the signs maintain an analogical relationship with their referent. Moreover, unlike vocal languages, signed languages and writing share the same visual gestural modality. Freed from the arbitrariness gap, there might be a chain of analogy going from the referent to the language then to the writing. This point is important as it would make for a major advantage over vocal languages. In existing systems for SL annotation, we observe different sign/script relationships: from conventional encoding of the grapheme-like part of each sign as in Hamnosys (Prillwitz, Siegmund *et al.* 1989) to the schematic representation of the signers body in motion (SignWriting). Our take on the problem is to assume that there could be a reduplication of the analogy, not between the sign and its referent this time, but between the sign and its graphical representation. The seed of a writing for signed languages would be enclosed within its own oral form: the shape cast by the hands of the signer in the space around them as they sign. An unprecedented fact in the history of writing as — aside from signed languages, oral never shows a continuity of canal with its written form.

3. DEVICE & GRAPHICAL METHOD

The ThirdEye motion capture system was designed to be modular. It can be broken down in three parts: the device itself, the capture part of the code, and the rendering. The device itself can change, as can the rendering, or even just the output. This makes ThirdEye a very versatile tool that can be used with any kind of motion capture system.



Figure 1: A signer equipped with the markers

3.1 Device

In ThirdEye, the motion capture is done using two luminous spherical markers. As we pointed out in section 2.1, most motion capture systems use multiple cameras or combine marker based

tracking with other sensors, which make them quite onerous. One of our goals with ThirdEye was to reach out the deaf community, whether individuals or schools. This means a low cost production and a relying on a DIY (Do It Yourself) approach.

In practical terms, ThirdEye uses a single point of view and a bespoke algorithm to track the markers in 3D space. We are currently using a Sony PS3 Eye camera, which delivers 60 frames per second at a 640x480 resolution. This is enough for most sign language gestures. To use the system, the signer clips a marker on each hand and faces the camera. The image on the screen then mirrors his hand movements in an abstract 3D environment. A foot-switch triggers the writing.

3.1.1. The algorithm

The algorithm in ThirdEye is separated into two threads, one for the capture and the writing, the second for the rendering and the graphical user interface.

The capture is based on a Monte Carlo algorithm and a tracking algorithm. We use two markers: one blue and one green. We set the camera to a low exposure in order to get the best contrast. We try to locate each marker. The program searches the image, looking at random pixels until either it finds one that has the appropriate colour, or it reaches a threshold (set at 10000 iterations in our code). We know that the spheres will appear as circles in the image. Then, based on the approximation of this projection, we draw lines up, down, right and left, until we are not anymore in the colour range we defined (which implies we are out of the marker). We have then two segments (from up to down, and from left to right). We take the x-value of the centre of the left to right segment, and the y-value of the centre of the up to down segment. As we said, the marker is seen as a circle, hence we can approximate the centre by a dot defined by those x and y values.

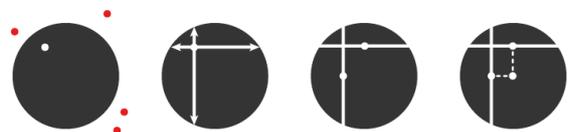


Figure 2: Detection of a marker and its centre

On the following frame, we do the same, except that we don't search the whole image again. We bypass the Monte Carlo subroutine and first check the previous centre. Since our camera works at 60 fps, the centre of the circle in a frame n is most of the time still within the circle in frame $n+1$.

To get the z-value, we try to systematically fill the majority of the pixels of the disk. The filling algorithm works by, first scanning the surface from top to bottom, extending lines left and right of the vertical segment we already have, we scan the surface line by line by searching matching pixels left and right. This usually gives a rather good filling of the surface, but we still miss a little part from the top and bottom of the disk. Hence in this subset of "left to right" lines, we take the top one (lowest y-value), and the bottom one (highest y-value) and from each pixels of the upper one, we extend lines up, and do the same downwards for the bottom one until we reach out of the disk. During all that process, we count the total number of pixels. This number is correlated to the surface of the disk, and since this surface is linked to the diameter of the disk, we can deduce from it the size of the spherical marker, and hence the distance between the camera and the marker. One interesting thing to note is that in this process, most of the processing is done on markers' pixels, hence reducing the cost in time of image processing since we never actually go through every pixel in the image.



Figure 3: Filling algorithm

Furthermore, our tracking algorithm works fast enough to deliver 60 positions per second and still leave us time to work on the data before hand. One of our main perspectives is the use of that time in order to filter data before use.

For the rendering, we use SFML for all GUI elements and OpenGL for 3D rendering. First, we display the position of the marker as two pointers on the screen. Their shape is a simple 3D object in OpenGL. When the signer activates the writing, we link all these shapes in consecutive frames, creating a ribbon. An outline of the body is displayed as a backdrop in order to help the signer get reference points in space. Another feedback is given by the colour of the pointers that changes when they are close to a conventional coronal plane in 3D space, giving an feeling of the z-axis.

3.2 Setup

The signer is equipped with the markers and stands in front of the screen. The markers are tracked by the system (see 3.1) and their trajectory appears as 3D strokes on the screen as soon as the signer presses the foot-switch. The written strokes are a mirror projection of the signer's and

an extension of the body in the virtual space. The whole experiment is designed to look and feel as much as possible as using a writing tool. It merges signing and writing to make their similarities appear.

In our experimentations we used different setups. The first one is without the foot-switch, we just give the signer the markers, and they play with the trace in order to discover the system and the 3D space with the help of some simple tasks such as connecting two dots in 3D. The second one is the whole setup described in the previous paragraph, which can be used while another subject watches the screen, unaware of the movement of the signer and comments on it. The last one is a chat setup where two signers are both equipped with markers and a foot-switch, while each sees on his screen the strokes produced by the other. Numerous language games can be done with this setup, and some are described in section 4.2.

3.3 The experiment

In the comments we got from the first deaf beta testers, we understood that their first impression is surprise. The visual feedback on their language production is something unusual that makes them rethink the way they sign. They also note that some signs produce a trace that is counter-intuitive. We still have to investigate this question but we believe that the movement may not have the same value in every sign.

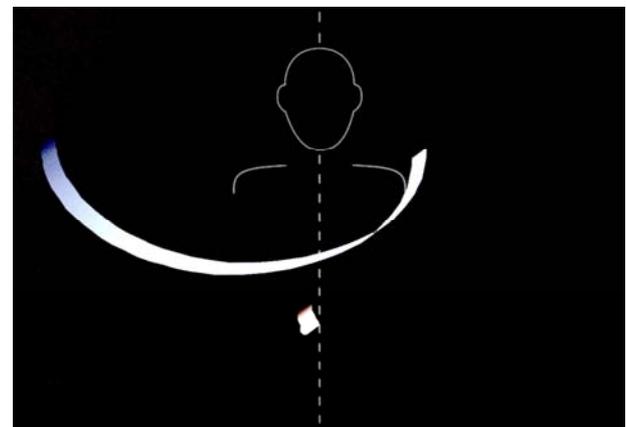


Figure 4: Trace of a sign: [succeed]

4. TEST METHODOLOGY

4.1 Our hypothesis

We observed that the trajectories followed by the hands into space are different for each sign. Our hypothesis is that these trajectories share a link with the meaning of the sign itself, the nature of which we will have to find if our hypothesis is proven. The way we proceed is we visualize the

trajectories of the hands in real-time so that the signers gain instant feedback on the shapes they unknowingly produce as they sign.

4.2 The language games

We imagined several experiments designed as games (except the first one). Each of them is a communication scenario and they all aim at understanding how our very abstract and simplified rendering of the movement can act as a communication medium and to what extent.



Figure 5: A signer using ThirdEye

The scenarios go from the most simple to the most complex communication situation. The information to be transmitted is getting more and more abstract and we give fewer clues.

Scenario 1: Two signers are equipped with markers and they must realise the same sign. Signer A writes a trace that will stay on screen and serve as a guide for the signer B. Then A and B can switch places. There is no goal in this scenario but the two signers get to experiment with the device and it gives them the opportunity of sharing their impressions about the graphical representation of the gestures.

Scenario 2: The two signers are shown 4 images (objects or simple concepts). Signer A secretly chooses an image and tries to explain it to signer B through traces on the screen only.

Scenario 3: Signer A is given a theme (e.g. animals, job titles, family...) and has to communicate it to signer B through traces on the screen only.

Scenario 4: We give the two signers a list of subjects, verbs and objects and a simple sentence structure [subject] [verb] [object]. Signer A creates

a sentence and tries to make signer B guess its components through traces on the screen only.

Scenario 5: We give signer A an image (e.g. the depiction of a historical event like the first moon landing). He has to tell the story to signer B through traces on the screen. We then interview signer B to see what amount of information they could recover from what they saw.

Each of these scenarios can be transposed in the chat setup if the material is available.

5. CONCLUSION

In this article, we presented a system that provides an easy way to visualise hand trajectories in real-time using a marker-based motion tracking system and a single camera, and its rendering. Our first experiments allowed us to gather feedbacks about its use in a sign language context, would that be about a potential writing, the linguistic aspect of movement or for educational uses. The comments we collected gave us new ideas about graphical means of showing the gesture, which we will compare and test in upcoming experiments with a larger audience.

For now, the limitations of the system are mostly occultation issues and the fact that it relies on optimal lighting conditions (requiring calibration). Also, the markers are attached to the edge of the hand so there is an offset in the measurement that can't be corrected with using only our system. The extra bulk on the hand can also "get in the way" during certain signs. All these issues could be addressed using inertial measurement units with robust sensor fusion (Madgwick, 2011) that could make IMUs a viable and rather affordable alternative to optical motion capture and give the extra information of orientation. We will look into this solution as the material becomes more widely available.

We are contemplating other applications than mere experimentation with our device. One is a set of tools to learn how to sign (ThirdEye could compare a sign in a database to the one currently produced on screen, advice a signer about its speed, position etc.) Another application would be a sign language input device that could replace the keyboard for sign language. And last would be to create a corpus of signs to be used by the community, for instance to help avatar rendering of sign language

As of our research, we plan on designing an experiment that will reveal the nature of the link between trajectory and meaning of the sign. One of our main objectives would then be to characterise this link, and to check if it's the same for every sign

(e.g. ample movement VS tiny movement). We will also check if there are categories of signs for which the trajectory is more meaningful than others. All of this will be a step forward in the understanding of the movement parameter of sign language, its link to writing and progress toward the creation of a written form of sign language.

6. REFERENCES

Chung Lee J., 2008. Hacking the Nintendo Wii Remote, *Ieee Pervasive Computing* 7 (3) p. 39-45.

Herrenschmidt C., *Les Trois écritures : Langue, Nombre, Code*, Bibliothèque Des Sciences Humaines, ISSN 0768-0570, 1 vol ([Paris]: Gallimard, 2007) pp.109 and after.

Hjelmslev L. & Léonard A.-M., *Prolégomènes à Une Théorie Du Langage : Suivi De ; La Structure Fondamentale Du Langage*, trad Una Canger & Annick Wewer, Arguments (Ed. De Minuit, Collection), ISSN 0530-7228 (Paris: Ed. de Minuit, 1996).

Lefebvre-Albaret F., 2010. *Segmentation de la Langue des Signes Française par une approche basée sur la phonologie*, PhD Thesis.

Madgwick, S.O.H., Harrison, A.J.L. & Vaidyanathan, R., 2011. Estimation of IMU and

MARG orientation using a gradient descent algorithm. *2011 IEEE International Conference on Rehabilitation Robotics. IEEE*, pp. 1-7.

Moeslund, T.B. & Granum, E., 2001. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3), pp.231-268.

Prillwitz, Siegmund et al.: *HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide*. (International Studies on Sign Language and Communication of the Deaf; 5) Hamburg : Signum 1989 - 46 p.

SignWriting. <http://www.signwriting.org>. [Accessed March 10, 2012]

Starner, T., Weaver, J. & Pentland, A., 1995. Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video. *Media*, (466), pp.1-7.

Vaughan-Nichols, S.J., 2009. Game-Console Makers Battle over Motion-Sensitive Controllers. *Computer*, 42(8), pp.13-15.

Zafrulla Z. et al., 2010. *American Sign Language Recognition using Kinect Skeleton features*. <http://youtu.be/qFH5rSzmgFE> (retrieved 25 March 2012).