Overview of the Sign3D Project High-fidelity 3D recording, indexing and editing of French Sign Language content

François Lefebvre-Albaret, Sylvie Gibet, Ahmed Turki, Ludovic Hamon, Rémi Brun

Overview of the Sign3D Project
High-fidelity 3D recording, indexing and editing of French Sign Language content

Francois Lefebvre-Albaret
Websourd 99 route d'Espagne, Bat A
31100 Toulouse, France
francois.lefebvre-albaret@websourd.org

Sylvie Gibet
IRISA
University of Bretagne Sud
Campus de Tohannic, BP 573
56017, Vannes, France
sylvie.gibet@univ-uns.fr

Ahmed Turki
Mocaplab
5, cité Riverin
75010 Paris, France
ahmed.turki@mocaplab.com

Ludovic Hamon
IRISA
University of Bretagne Sud
Campus de Tohannic, BP 573
56017, Vannes, France
ludovic.hamon@univ-uns.fr

Rémi Brun
Mocaplab
5, cité Riverin
75010 Paris, France
remi.brun@mocaplab.com

ABSTRACT

The Sign3D project aims at creating a range of innovative tools to allow the recording and the processing of motion captured French Sign Language (LSF) content. The challenge is to design a complete workflow from the movement capture (including all upper body part articulations, facial expression and gaze direction) to their restitution using a 3D virtual signer. We present the main innovation challenges at each step of the project. As accessibility for Deaf people through Sign Language is one goal of this project, a project overview of the project in SL is accessible at the following address : http://sign3d.websourd.org/sltat

Keywords

Sign Language, Virtual Signer, Automatic Language Processing.

1. DATA DRIVEN VS PROCEDURAL ANIMATION

The first researches about virtual signers producing French Sign Language (LSF) used explicit animation commands in order to produce animations [5]. Even if this approach allows theoretically to produce any range of movements of any body part, the lack of parameterized linguistic models in LSF often leads to idealized movements or lack of signer expressivity (particularly regarding facial expression) [7]. In the *Sign3D* project, we decided to tackle this issue by directly generating sign language sentences from a concatenation of motion captured items. The capture leads to the restitution of any manual and non-manual parameters of the signs and its inflections [3]. The main challenge is then to find the best rules in order to choose the good sign variant depending on the sentence that will have to be produced.

2. CORPUS CREATION

Motion capture is still an expensive technique. Consequently, we decided to capture only about twenty sentences to start our project. As automatic signed information in public areas is one of the most promising applications of virtual signers [6], we decided to concentrate our efforts on messages such as "opening hours", "entrance fees" or "perturbation messages". The biggest challenge was to build the corpus in order to be able to compose other utterances by recombination and to have enough variability for each sign.

3. HIGH-FIDELITY MOTION CAPTURE

Each sentence has been recorded by an optical motion capture system combined with a head-mounted oculometer. Markers are placed on the whole signer's upper body, including her face and fingers which allows for a complete performance capture. One of the challenges of this process is to find a compromise between motion capture cost, measurement (space and time) accuracy, and spontaneity of the production (if the motion capture equipment is too invasive, the signer will not be able to sign in a natural way). After motion capture, the marker set is rigged onto a 3D virtual signer mesh in order to animate both its skeleton and its face.

4. ANNOTATIONS

During motion capture, a reference video (a frontal view of the signer) is also recorded. Then a deaf signer annotates

Figure 1: Motion capture session

this video with the Elan software [2]. Sentences are segmented into signs, labelled by a string conveying its meaning. Other meta-data can also be added to the segments about handshape, face expression, body posture, or any other feature that may be relevant to choose a good sign variant when creating new utterances.

5. COMPOSITION INTERFACE

The goal is then to combine the signs into new informative sentences that respect sign language organization rules. It is tempting to think of this composition problem only as the substitution of some words in a written sentence, which would lead to a kind of “signed French”. In order to avoid such a pitfall, most parts of the interfaces will be composed of only visual elements (icons, handshapes, sign pictures).

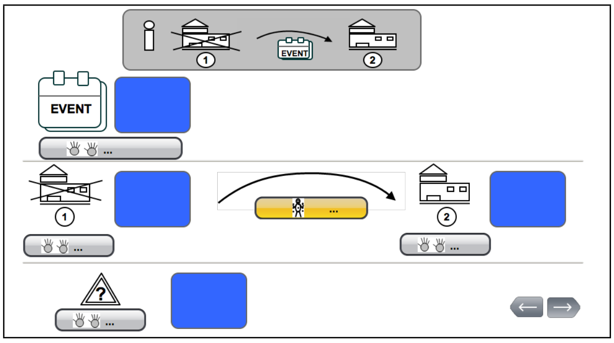

Figure 2: First version of the composition interface

6. DATABASE QUERIES

The corpus annotation allows a mapping between the meanings of the signs (and distinctive features) and their realization presented as motion captured data, using a method close to [1]. Once the new sentence is composed, the solution has to select the *good* sign sequence, which means retrieving in the data base the good variant of each sign. At this stage, it is important to point out that the goal is to find at least one way to express an information (*i.e.*: opening hours of the museum: 8am - 17pm[1]) but the way of signing it is not indicated in the interface. If the program tries to compose the new sentence only from one possible structure (*i.e.*: “The museum is opened from 8am to 17pm”), it is likely that the appropriate segments will not be inside our initial corpus. To bypass this issue, we try to find the matching segments in the database for several variants (*i.e.*:“The museum’s opening hours are 8am-17pm”, or even “The museum opens at 8am and closes at 17pm”). In other words, the goal is to compensate the small variation in the initial corpus by diversity in the syntactic structures that the system can handle.

[1]We make an analogy with a written language but the program naturally uses a language model dedicated to sign language.

7. FROM THE PLAY LIST TO THE ANIMATION

The previous step results in one (or several) list(s) of segments that have to be concatenated into a new SL sentence. If several lists are correct from a syntactical point of view, the selected one will have to be the one that optimizes the transitions between motion segments. Then, an animation engine will compute the combination of motion chunks. The assembling of signs occurs naturally between signs as temporal interpolations, and within signs as spatial blending when motion chunks are retrieved from different items from the database (*e.g.*: a sign D is made of the right hand of the sign A, the left hand of the sign B and the facial expression of the sign C). The overall principle is the same as the one described in [4].

8. RENDERING

The whole workflow ends with a rendering of the virtual signer signing the output sentence. The rendering must be high fidelity (motion-wise) but not necessarily photorealistic. The most important issue is to play back accurately each sign language parameter in order to be properly understood by deaf final users.

Figure 3: Rendering

9. FUTURE WORK

At the moment, our first corpus has been annotated and the database engine is able to retrieve signs by label or feature, and the corresponding motion capture chunk. The first interface has been designed and will shortly be connected to the database. As soon as the workflow from motion capture to synthesis is operational, we will submit the system to deaf users for evaluation, in order to find the most relevant criteria that enable the composition of novel SL sentences from motion captured data.

10. ACKNOWLEDGMENTS

The *SIGN3D* project is funded by the French Ministry of Industry (DGCIS: Direction Générale de la Compétitivité de l'Industrie et des Services, Program "Investissements d'avenir"). The interfaces have been designed by Guillemette Bosch who also annotated the whole video corpus.

11. REFERENCES

[1] C. Awad, N. Courty, and S. Gibet. In proc. of the 7th international workshop on content-based multimedia indexing (cbmi 2009), ieee cs (ed.). Chania, Greece, June 2009.

[2] O. Crasborn and H. Sloetjes. Enhanced elan functionality for sign language corpora. In *In: Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation.*, 2008.

[3] K. Duarte and S. Gibet. Heterogeneous data sources for signed language analysis and synthesis: The signcom project. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).

[4] S. Gibet, N. Courty, K. Duarte, and T. Le Naour. The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation. In *Transactions on Interactive Intelligent Systems*, volume 1. ACM, 2011.

[5] T. Lebourque and S. Gibet. High level specification and control of communication gestures: the gessyca system. In *Proc. of Computer Animation*, Genova, Switzerland, May 1999.

[6] J. Ségouat. Modélisation de la coarticulation en langue des signes franÃğaise pour la diffusion automatique d'information en gare ferroviaire a l'aide d'un signeur virtuel. In *PhD thesis, Université Paris Sud*, 2010.

[7] Websourd. Evaluation of the sign wiki. In *Dicta-Sign : Sign Language Recognition, Generation and Modelling with application in Deaf Communication, D8.2*, 2012.