Methodological issues in automatic recognition of pointing signs in LSQ using a 3D motion capture system

Julie Rinfret
Anne-Marie Parisot
Karl Szymoniak
Suzanne Villeneuve
UQÀM
Thierry L. Chevalier
(C-Motion Inc.)

rinfret.julie@uqam.ca

Theoretical Issues in Sign Language Research 11, University College London • July 2013

INTRODUCTION

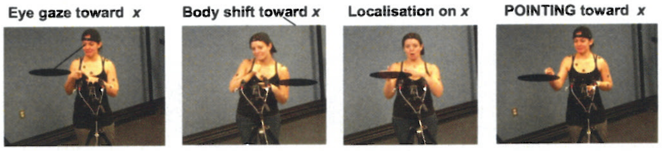
Prior works on the use of space in sign languages are mainly based on qualitative analysis. With respect to the articulatory possibilities in the production of sign discourse, there is an increasing interest among sign language researchers in understanding and quantitatively describing the use of space, especially for sign recognition and animation purposes.

Jantunen *et al.* (2012) point out that most motion capture (from now on <u>mocap</u>) studies on sign language have consisted of relatively small sets of isolated expressions (single signs (Wilbur, 1990), phrases (Tyrone et al., 2010)), and that the collection of biomechanical discourse-type data has been marginal (cf. Duarte and Gibet, 2010). Even though recording, processing and analyzing biomechanical data is time consuming, we believe that the understanding of the interaction of signs in space in discourse has to be analyzed using discourse data. This requires a robust markerset in order to automatically recognize and identify specific signs such as pointing signs.

CONTEXT : SPATIAL ASSOCIATION IN LANGUE DES SIGNES QUÉBÉCOISE (LSQ)

- **The Marqspat research project works on the linguistic description of four strategies that are used to identify and track referents within a discourse (semantic, morphosyntactic and pragmatic perspective).**
- **Instantiation of a noun in discourse can be accomplished through spatial association, using one of the four following strategies** (Parisot, 2003; Parisot and Rinfret, 2008; Rinfret, 2009):

Eye gaze toward *x* | **Body shift toward *x*** | **Localisation on *x*** | **POINTING toward *x***



- **The processing of the data is long and arduous (transcription, visual tracking and identification of each form).**
- **We use a 3D motion tracking system as a tool for quantifying:**
  - **The phonetic contours of the spatial trace left by spatial association;**
  - **The scope of each strategy of spatial association on one another;**
  - **The spatiotemporal patterns of spatial association.**
- **Using the 3D motion data, we aim to convert the manual identification of the strategies of spatial association to an automatic operation, in order to end up with corpora that are more representative of our data, in a lesser amount of time.**

Objective

Given that the pointing sign is, from the biomechanical perspective, easier to isolate than the other forms (body shift, localisation), the goal of the project is to mark out the linguistic and the biomechanical parameters which will allow the automation of the identification of this form.

PROBLEM: NEED FOR A MINIMAL MARKERSET

- The description of sign languages using a 3D mocap system is quite new. Early studies exploiting mocap data are Wilbur (1990) and Wilcox (1992). More recent works are, among others, Tyrone *et al.* (2010), Duarte and Gibet (2010), and Jantunen *et al.* (2012).
- Due to the great variability of hand shapes and types of contact in sign languages, the movements of the fingers involved are much less regular than those in handling activities (e.g. prehension task). Thus, the post processing (manual labelling of markers) becomes a really complex task, especially if the number of markers is important.
- There is a primary need of a robust enough but minimal markerset without affecting the reliability of the data.
- To our knowledge, there is no description in the literature on mocap studies in sign languages of a robust minimal markerset that allows the automatic recognition of signs in continuous discourse data.

SOME REMARKS ON PRIOR STUDIES

- **Few have used continuous signing (c.f. Duarte and Gibet, 2010; Jantunen *et al.*, 2012).**
- **The size of the markers, when mentioned, makes them intrusive (c.f. Cormier, 2002 (25 mm); Mauk, 2003; Jantunen *et al.*, 2012), even more if they are wired (c.f. Malaia *et al.*, 2008).**

From Malaia *et al.* (2008)

From Tyrone *et al.* (2010)

- **Some have included markers on the back and the pelvis, which is essential for the calculation of the actual position of the body (e.g. body shift, body rotation).**
- **No precise description of clothing (c.f. Tyrone *et al.*, 2010); loose clothing (c.f. Jantunen *et al.*, 2012).**

From Jantunen *et al.* (2008)

METHODS:

1) GETTING THE LSQ DATA

- The stimuli are made of short simple scenes which present vignettes of everyday life. In these scenes (which involve two or three characters), individuals interact without any language (no speech, no signs, no written words). This material is designed to elicit specific linguistic structures.
- For each scene, the experiment contained the same following sequence:
  - Look at a scene presented on a screen;
  - Look at pre-videotaped questions on a screen. All participants have been presented with the same questions and had to answer with a short statement (e.g. a man and a woman);
  - Produce the answer after each question;
  - Based on the scene, produce a short narrative.

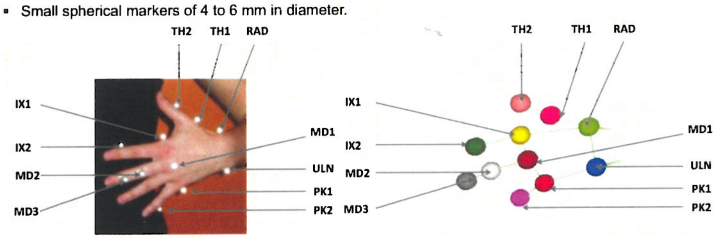
2) GETTING THE KINEMATIC DATA

Articulators	Markerset of the pilot study	New markerset
Head	3	4
Torso (front)	6	4
Torso (back)	0	2
Pelvis	0	4
Arm	4 x 2	5 x 2
Hand	5 x 2	11 x 2
TOTAL	27	46

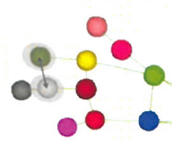
MS of the pilot study

New MS

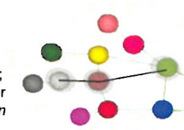
- The Eagle Camera system (Motion Analysis Corporation) is a passive marker system.
- Cameras strobe IR light that markers reflect.
- Cameras are independent from each other (can be pointed in any direction).
- Markers are not automatically labeled (need for post processing).
- Sampling rate of 60 Hz (to be temporally synchronized with our eye tracking system, faceLAB).
- Small spherical markers of 4 to 6 mm in diameter.



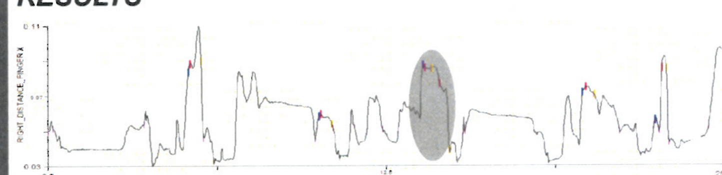
AUTOMATIC RECOGNITION OF POINTING SIGNS: BIOMECHANICAL CRITERIA



1) Calculation of the **linear distance** between markers on the index (IX2) and the middle finger (MD2);
2) Calculation of the **mean distance** between IX2 and MD2;
3) If the signal of the distance between IX2 and MD2 is greater than the mean distance, the event '*mean up*' is created (Visual 3D, C-Motion Inc.);
4) If the signal of the distance between IX2 and MD2 is less than the mean distance, the event '*mean down*' is created.
5) Calculation of the angle of flexion of the middle finger (MD2, MD1, RAD);
6) The beginning of the pointing sign is identified when the angle is greater than 80° (in the intervals defined by the events '*mean up*' and '*mean down*').



RESULTS



- The MS and the biomechanical criteria allow the automatic recognition of pointing signs in LSQ without false positives.
- **PROBLEM** : no automatic recognition of consecutive pointing signs
- **SOLUTION** : calculation of the median duration of pointing signs = isolation of consecutive pointing signs.

TO BE CONTINUED ...

- **Defining the biomechanical criteria for the recognition of**
  - **"loose" pointing signs;**
  - **non canonical pointing signs;**
- **The method does not work with all participants: need for a generic rule of recognition;**
- **Defining the biomechanical criteria for the automatic recognition of all the strategies of spatial association.**