Annotation of Non Manual Gestures: Eyebrow movement description

Emilie Chételat-Pelé, Annelies Braffort & Jean Véronis

LIMSI/CNRS, Orsay, France
Provence University, Aix-en-Provence, France
Emilie.chetelat@limsi.fr, annelies.braffort@limsi.fr, Jean@Véronis.fr

Abstract

Our study tackles Non Manual Gestures (NMGs) annotation within the context of Sign Language (SL) research and more particularly within the context of automatic generation of French Sign Language (LSF). Present descriptions need instantiation for the animation software. Thus, we propose a new annotation methodology, which allows us precise description of NMGs and which takes into account the dynamic aspect of LSF. On the video corpus, we position points on elements to be annotated, to obtain their coordinates. These coordinates are used to obtain precise position of all NMGs frame by frame. These data are used to evaluate the annotation by means of a synthetic face, for numerical analysis (by using curve), and, finally, to obtain numerical definition of each symbol of our annotation system based on arrows

1 Introduction

This paper deals with non manual gestures (NMGs) annotation involved in Sign Language (SL) within the context of automatic generation of SL. Many researches in SL emphasize the importance of NMGs at different language levels (lexical, syntactical, pragmatic…) and recognize that NMGs are essential for the message comprehension. However, the NMGs structure knowledge is limited. Our purpose is to refine the knowledge of NMGs structure and their roles. To acquire this knowledge, it is necessary to have precise NMG descriptions. These descriptions are obtained from the observation and annotation of a video corpus. Depending on the degree of precision we need, the first step is to conceptualize an annotation methodology. We suggest in this paper a methodology, which allows us a numerical annotation of NMGs for a precise description of NMGs structure. This study is based on French Sign Language (FSL) but can be used for another SL.

The next section presents the context of this study: the available descriptions and transcriptions of NMGs and the presentation of our purposes. In the third section, we suggest a new annotation methodology, which allows us to study the NMG movement dynamics.

2 Problematic

At present, descriptions of NMGs are symbolical. Transcription systems like HamNoSys (Prillwitz and Zienert, 1989), D'Sign (Jouison, 1995) or SignWriting (Sutton, Gleaves, 1995), describe the NMG posture with more or less iconical graphical forms (Figure 1: "eyebrows high" transcribes by different systems).

This type of description is not suitable for automatic generation systems because they do not contain numerical indication. Moreover these descriptions relate to a given instant and do not allow us to describe the movement intensity and dynamics. For example, for a description such as "Eyebrows high", we would like to know the movement intensity and the raising duration. Thus, these systems are not accurate enough to study the importance of these elements in the meaning transmission.

In this article, we suggest a new methodology applied to the eyebrow and eye movements. This allows us to study the NMG movements with the aim to provide precise descriptions of these movements. Describing NMGs precisely imply a rigorous annotation of the different NGM movements that can be observed on a video corpus. The methodology must provide the means to describe all the phenomena and the study of the NGM movement dynamics. The methodology has also to provide a formal definition of NMG structure.

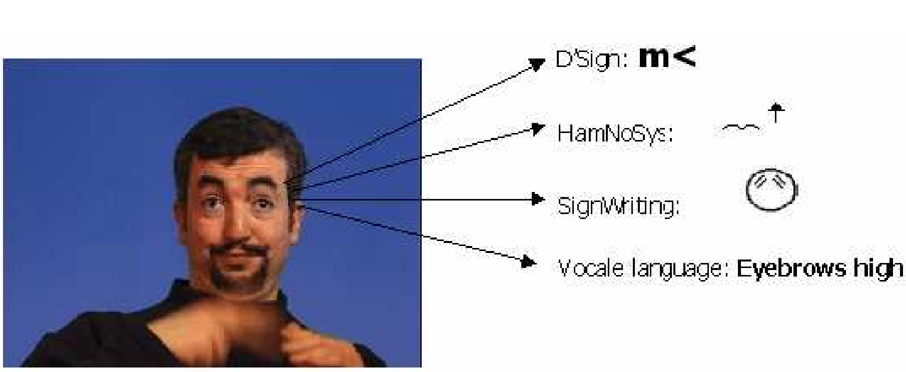

Figure 1: Many transcriptions of "eyebrows high".

3 Methodology

This part presents an application of our methodology on eyebrows and eyes movements. For annotation, we used LS-Colin corpus (Braffort et al, 2001; Segouat, Braffort, Martin, 2006). The video quality and the close-up shot are particularly precious for our study. Moreover, we used Anvil software because this software offers the possibility to annotate with personal icons and colors, which is of great help for a visual perception of phenomena.

Moreover, Anvil allows us to directly annotate on the video frames by means of points. Their coordinates can then be exported for further treatments.

The first section (3.1) presents how the video was annotated based on the FACS system (Facial Action Coding System). In a second section (3.2), we explain in detail the annotation data processing. Then, the last three sections (3.3, 3.4, and 3.5) present three data uses that permit to analyze and evaluate the annotation.

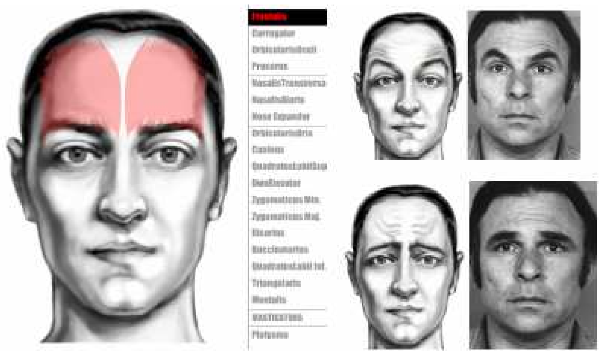

Figure 2: Frontal muscle frontal and the associated eyebrows AUs: outer extremity rise (AU1) and inner extremity (AU2). Pictures extracted from the Artnatomy[1] website (Contreras Flores, 2005) and the FACS manual[2].

The corrugator supercillii muscle, the orbicularis oculi muscle and the procerus muscle allow lateral movement of the eyebrows, which is inducing a variation of the distance between the eyebrows (Figure 3).

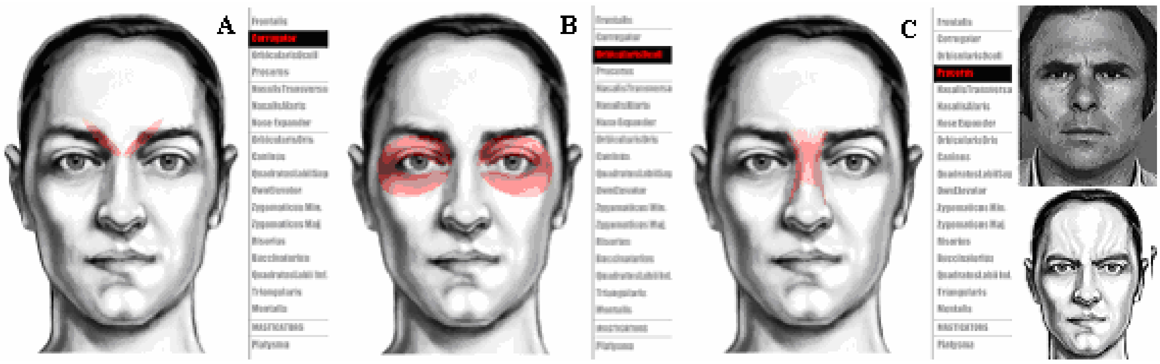

Figure 3: The corrugator supercillii muscle (picture A), the orbicularis oculi muscle (picture B) and the procerus muscle (picture C), responsibles of the AU4. Pictures extract of Artnatomy (Contreras Flores, 2005) and FACS manual.

3.1 Annotation on the videos

For the eyebrows movement description we use the FACS system, which has been designed for the description of emotion mimics. FACS is a description system of facial expression, which is based on facial muscles (Ekman and Friesen, 1978). Actually, Ekman and Friesen use these muscles as a base for the definition of all face movements. FACS measurement units are the Action Units (AUs), which represent the muscular activity that produces momentary changes in facial appearance.

For the eyebrows, Ekman and Friesen distinguish four muscles allowing three actions: rise of eyebrow inner (AU1), rise of eyebrow outer (AU2) and eyebrow lowering (AU4).

The frontal muscle (Figure 2) is responsible of the rise of the eyebrow inner and outer extremities.

The figure 4 shows three Aus combination: AUs 1 with 4 (inner rise and eyebrow lowering), AUs 1 with 2 (inner and outer rises), and AUs 1, 2 and 4 (inner and outer rises and eyebrow lowering).

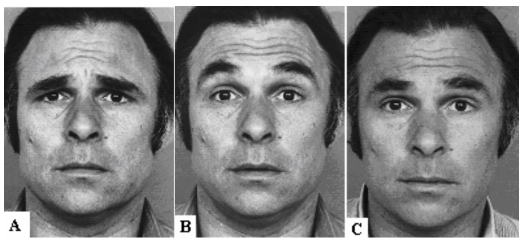

Figure 4 : Three AUs combinations. Picture A: AU1 + AU4 ; Picture B: AU1 + AU2 ; Picture C: AU1 + AU2 + AU4

1 www.artnatomia.net
2 http://www.face-and-emotion.com/dataface/facs/manual/TitlePage.html

Theses pictures show that the size of the eyebrow can change according to the AUs and their combinations. Moreover, the middle of the eyebrow rises with a bigger amplitude than its outer extremity, implying a more important perception of movement in this area.

FACS is a formal coding system is useful for facial expression description. However, it does not allow a description of dynamics (temporal analysis…). Then we only use FACS as a base, from which we have elaborate our own methodology.

For the eyebrows, FACS distinguishes two points (inner and outer extremities), which can move on horizontal and lateral axes. We retain these points for the video annotation. But because of its greater movement amplitude, we also consider the middle of the eyebrow and annotate it.

Moreover, to limit the annotation imprecision involved by the eyebrow thickness we double the extremity points (inner and outer) for each eyebrow and triple the middle point, the most difficult to accurately position.
Finally, to determinate the eyebrows movements independently of the head movement, we consider reference positions: the two extremities of each eye.
Thus, we position 18 points on each frame of the video (25 frames by second). The figures 5 and 6 show the location of each point.

After having annotated the whole video, we export the 2d coordinates x and y of each point. Calculations on theses coordinates give us precise data of the eyebrows movements.

3.2 Calculation on the 18 point coordinates

For the data processing, we used Scilab[3] software, free software for scientific calculation, which allows us, within a script, the automation of calculations. The input is coordinates of each point. These data are used for calculations to compute the position of each point independently of the head movement, frame by frame:

1. First, we calculate the average coordinates of the extremity and middle of each eyebrow for each frame (2-3, 7-8, 11-12, 16-18 for the extremities, and 4-5-6, 13-14-15 for the middles).
2. The news coordinates are used to calculate the distance (D) between these 3 points of each eyebrow to the extremity points of the eyes (for example, the distance between the point 1(x1,y1) and the average point 2-3 (x2,y2)): $D = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$.
3. We calculate the variation (V) of the position at the frame (n) by means of the Distance (D): $V(n) = D(n) - D(n-1)$. This variation can be positive (for a rise) or negative (for a lowering).
4. Then, the variation (V) allows us to calculate the position (P) of each element, independently of the head movements, for each frame of the video: $P(n) = V(n) + P(n-1)$.

These final data are used for the annotation evaluation and analysis.

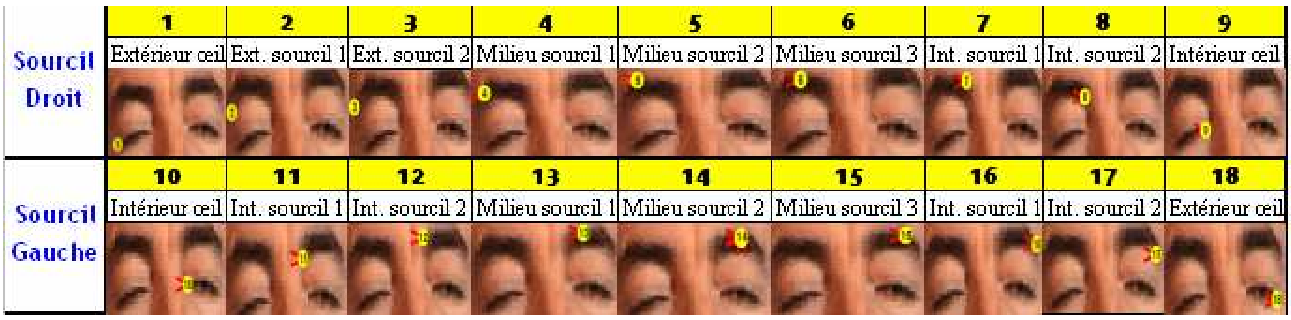

Figure 5: Site of 18 points

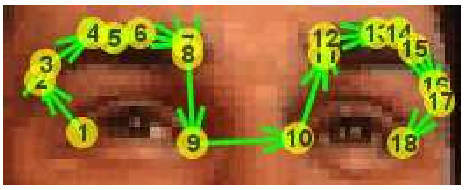

Figure 6: Corpus video extract with points.

3.3 Intermediate evaluation

These numerical data allow us to automatically generate the eyebrows animation on a synthetic face. For the generation, we used the Xface[4] software (Balci, 2006). Xface is a 3D talking head and was built for vocal production; not for SL production.

This automatic generation allows us to have a first qualitative evaluation of our annotation. We can compare the video and the Xface production simultaneously and evaluate if all phenomena are presents. Thus, we can adjust the annotation (for example put one more point) if necessary.

[3] http://www.scilab.org/
[4] http://xface.itc.it/

In figures 7 and 8, the left picture is extracted from our corpus (LS-Colin corpus, Braffort et al, 2001; Segouat, Braffort, Martin, 2006), and the right one from some Xface productions generated from our annotation.

Figure 7: Standard position

Figure 8: higher eyelid and distance between eyebrows lowered

Moreover, playing the Xface production and the video at the same time allows us to evaluate the synthetic face. We have yet identified the limits of the Xface face model and we can propose ameliorations for the synthetic faces used for automatic generation of LSF. For example, we observe that Xface do not have wrinkle and does not provide enough amplitude for the movements of eyebrow and eyelid. These limits induce perception problems for deaf users because it is very difficult to determinate the of eyebrow position. Thus, we can establish a list of necessary elements for synthetic face to produce realistic LSF.

This first use of the data allows us qualitative evaluation of the methodology. Data are then used for NMGs analysis.

3.4 Structural analysis of NMGs

Numerical data allows us to analyze the movement structure. For example, the curve presented figure 9 informs us of the amplitude of the eyebrow inner point and allows us a classification of the rises.

This curve shows three rise amplitudes for the eyebrow inner point: one small rise (1 unit for this person), one medium rise (2 units) and one high rise (3 units). These rises can be defined related to the small rise: a medium rise is two times higher than a small. A high rise is three times higher than a small rise. The precise numerical value of the rises amplitude can vary but the number of rise classes and their proportions are always the same. Then, a very high rise (7 units on the curve) is adopted in several steps: several rises of different degrees successively.

As show this example, the curves allow us to analyze the structure of the NMGs movements.

3.5 Formalization evaluation

These numerical data also allow us a validation and a numerical instantiation of the formal description based on arrows that we had presented in a previous paper (Chételat-Pelé, Braffort, Véronis, 2007). This system is based on four properties:

- Movement description (instead of posture description): For example: "eyelid lowering" instead of "low eyelid".
- Movement decomposition: For example, the diagonal movement of shoulders is described with horizontal movement and vertical movement separately;
- Element decomposition: For example, we separate higher eyelid and lower eyelid;
- The use of a set of symbols rather than words (Figure 10). One symbol can describe many phenomena (for example with use of colors for the movement intensity, figure 11).

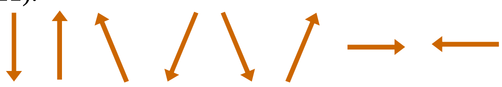

Figure 10: Set of symbols used.

Figure 11: Different degrees of intensity

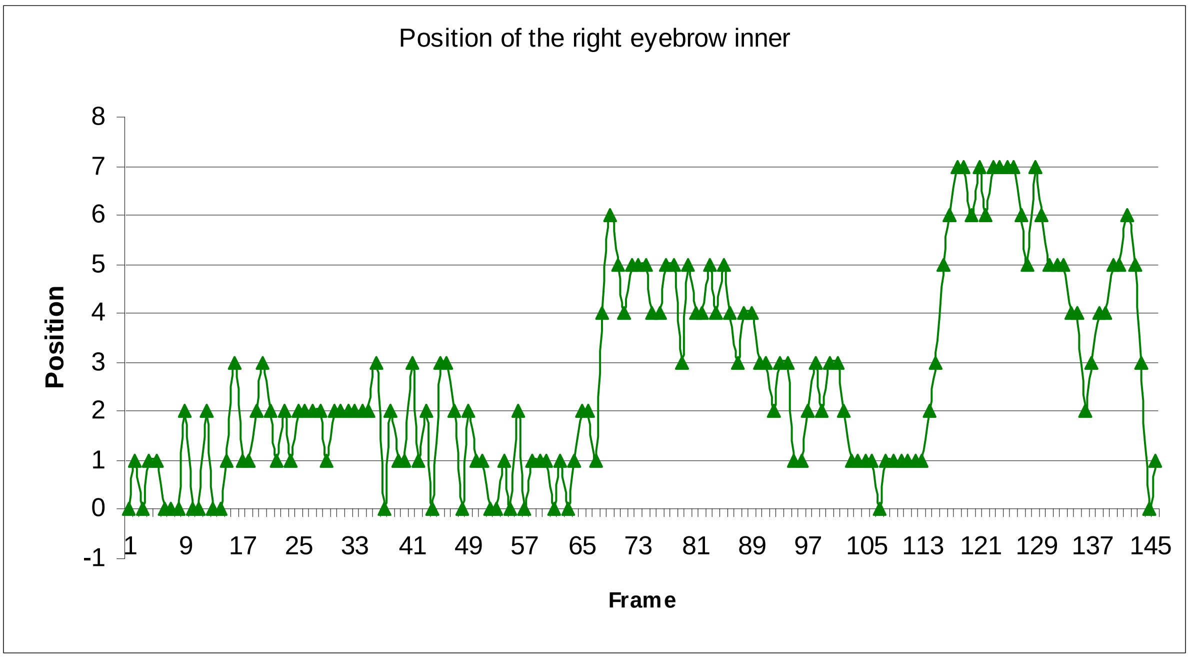

Figure 9: Position of the right eyebrow inner.

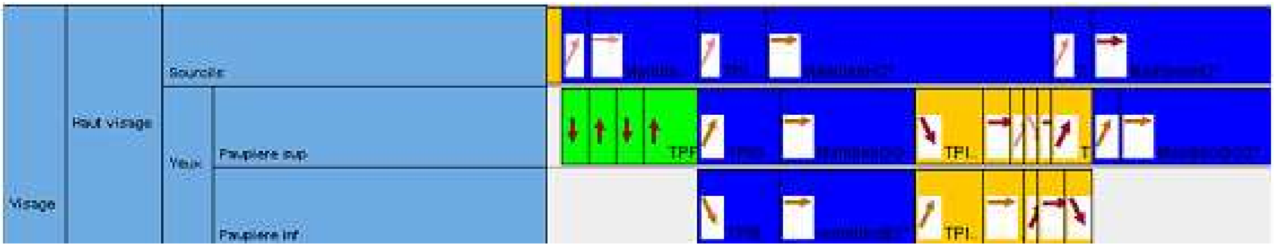

Figure 12: Annotation extract.

This description is simple and the use of the colors allows us to identify quickly the present phenomena (Figure 12). Our methodology allows us to define numerical values for each symbol. Moreover, we can automatically produce the annotation by means of the numerical data and validate our system. The numerical data have confirmed that there are three degrees of eyebrow movement (Figure 9). Applied on the whole arrow system we can determinate the pertinence of each symbol.

4 Conclusion

This study takes place within the context of automatic generation of SL and aims at enhancing of the NMGs structure knowledge to ameliorate the animation capacities of automatic generation system.

We have presented, in this paper, a system allowing a accurate numerical description of some NMGs. This system is based on the annotation of each video frame. Moreover, it allows us to obtain precise positions of the eyebrows, independently of the head movements.

The annotation will be extended on other video to validate our first observations. Moreover, the synthetic face evaluation will be extended to identify the properties that the faces have to respect to produce precise and understanding LSF.

Références

Balci K. (2004). MPEG-4 based open source toolkit for 3d facial animation. In AVI04, working Conference on Advanced Visual Interfaces. Gallipoli, Italie, 25-28 Mai 2004.

Braffort A., Choisier A., Collet C., Cuxac C., Dalle P. Fusellier I., Gherbi R., Jausions G., Jirou G., Lejeune F., Lenseigne B., Monteillard N., Risler A., Sallandre M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ?. In Journées Recherches sur la langue des signes. UTM, Le Mirail, Toulouse.

Chételat-Pelé E., Braffort A.,Véronis J. (2007). Mise en place d'une méthodologie pour l'annotation des gestes non manuels. In TALS 2007. Traitement Automatique des Langues des Signes 2007 : atelier de Traitement Automatique des Langues Naturelles 2007. Toulouse.

Chételat-Pelé E., Braffort A.,Véronis J. (2008). Signs Language Corpus Annotation: Toward a New Methodology. In LREC 2008. Conference on Language resources and Evaluation. Marrakech. (A paraître).

Victoria Flores V. (2005). Artnatomy/Artnatomia. url (www.artnatomia.net). Spain.

Ekman P., Friesen W. V. (1978). Facial Action Coding System (FACS). Manuel Palo Alto : Consulting Psychologists Press.

Jouison P. (1995). Ecrits sur la Langue des Signes française. Garcia, B. (éd). Paris : L'Harmattan, Paris

Kipp M. (2004). Gesture Generation by Imitation - From Human Behavior to Computer Character Animation" Boca Raton Florida: Dissertation.com

Prillwitz S., Zienert H. (1989). Hamburg Notation System for Sign Language: Development of a sign writing with computer application. Allemagne : S. Prillwitz & T. Vollhaber (Eds.): Current trends in European Sign Language Research Signum.

Segouat J., Braffort A., Martin E. (2006). Sign language corpus analysis: synchronisation of linguistic annotation and numerical data. LREC 2006. Fifth International Conference on Language Resources and Evaluation Genoa Italy : 2006

Sutton V., Gleaves R. (1995). SignWriter – The world's first sign language processor. Deaf Action Comittee for SignWriting, La Jolla, CA.