Synthesizing Facial Expressions for Sign Language Avatars

Yosra Bouzid, Oussama El Ghoul, Mohamed Jemni
Research Laboratory of Technologies of Information and Communication & Electrical Engineering
University of Tunis
yosrabouzid@hotmail.fr, oussama.elghoul@rnu.tn, mohamed.jemni@fst.rnu.tn

Abstract

Sign language is more than just moving the fingers or hands; it is a visual language in which non-manual gestures play a very important role. Recently, research has paid increasing attention to the development of signing avatars endowed with a set of facial expressions in order to perform the actual functioning of the sign language, and gain wider acceptance by deaf users. In this paper, we propose an effective method to generate facial expressions for signing avatars basing on the physics-based muscle model. The main focus of our work is to automate the task of the muscle mapping on the face model in the correct anatomical positions and the detection of the jaw part by using a small set of MPEG-4 Feature Points of the given mesh.

Keywords: facial expressions, signing avatars, MPEG-4, feature points

1. Introduction

Thanks to the advances in virtual reality and human modeling techniques, signing avatars have become increasingly common elements of user interfaces for a wide range of applications such as interactive e-learning environments and mobile phone services, with a view to improving the ability of hearing impaired people to access information and communicate with others. In order to ensure maximum comprehension and clarity to these signers, digital humanoids are required to perform not just broad hand movements, but also many subtle clues and features like face movements and expressions, which must be clearly seen in order to understand the meaning. Eyebrow height, mouth shape, and other facial gestures are linguistically required in sign language, and identical hand movements can have different meanings depending on the facial expressions performed during the sentence (Neidle et al., 2000).

Different approaches have been taken to animate a three dimensional synthetic human face, but most require a significant effort and time-consuming to adjust animation parameters. For example, the process of rigging requires many hours of manual work to set up the bone structure for an entire face. Even simple method like shapes blending needs the intervention of an artist to create large libraries of key shapes. On the other hand, the production of expressive and realistic animations involves a high computational complexity for simulating the physical property of the underlying facial structure which includes the skeletal, different muscles forms and the subcutaneous fatty tissues.

To deal with these problems, we present in this paper an effective method capable of deforming a 3D mesh of an arbitrary synthetic human face to generate emotional expressions without considerable amount of manual intervention and artistic skill. This approach relies on the physics-based muscle model proposed by Waters (1987) to emulate the contraction of the muscle onto the skin surface. Our contribution consists essentially of the automatic construction of mimetic muscles as well as the jaw mesh detection using only MPEG-4 feature points of the given mesh.

2. Background

This section presents a brief description of the most popular techniques and approaches which are generally utilized in 3D facial animation today.

2.1 Blend Shapes

The blend shape animation method, also known as morph target animation or shape interpolation is the most intuitive and commonly used technique in the field since it is quite straightforward and easy to accomplish. The basis for this method is that during the animation, the interpolated facial model is created from a specific set of key facial poses called blend shapes through interpolation over a normalized time interval. Typically, a blend shape model is the linear weighted sum of a number of topologically conforming shape primitives. Varying the weights of this linear combination allows the representation of facial motions with little computation. However, it is important to note here that the generation of a significant range of highly detailed expressions usually implies the creation of large libraries of blend shapes which can be very time-consuming. Moreover, if the topology of the model needs to be changed, all the shapes must be redone (Ping et al., 2013).

2.2 Facial Rigging

Rigging is the process of setting up a group of controls to operate a 3D model, analogous to the strings of a puppet. It plays a fundamental role in the animation process as it eases the manipulation and editing of expressions, but rigging can be very laborious and cumbersome for an artist. This difficulty arises from the lack of a standard definition of what a rig is and the multitude of approaches on how to set up a face (Orvalho et al., 2012).

2.3 Parameterization

In parameterization, facial geometry is broken into parts where each part is exposed properly to its parameter sets or control points. This allows the animators to have control of the facial configurations (Ping et al., 2013). By combining different parameters, a large range of facial expressions can be produced. Facial Action Coding System (Ekman & Friesen, 1977) and the MPEG-4 Facial Animation standard are the most famous parametrizations that can be included in this category. The advantage of these approaches is that once control parameters are determined, they provide a detailed control over the face. But determining this is hard: complexity of creating an animation with these control parameters is related to the number of control parameters, as is the possible range of expressions (Ilie et al, 2012). Furthermore, the animation does not seem to respond to basic physical deformations of human faces since direct parameterizations make no attempt to represent the detailed anatomical structure. They model only the changes visible on the skin surface.

2.4 Physics-based muscle Modelling

Physics-based muscle methods have the potential to produce natural 3D animations by precisely simulating the real effects of the facial muscular tissues. They can generally fall into three different categories: mass spring systems, layered spring meshes and vector representations. Mass-spring systems (Platt & Badler, 1981) are designed to propagate muscle forces in an elastic spring mesh that models the elastic properties of the skin, while, layered spring meshes (Terzopoulos & K. Waters, 1990) extend the mass spring structure to three connected mesh layers to simulate the anatomical aspects of the face more faithfully. In vector representations, the actions of muscles upon the skin are modelled using motion fields in delineated regions of influence. The most successful muscle models were proposed by Waters (1987) and Kähler (2007) who proposes a muscle structure composed by quadric segments. Although the physics-based techniques are the most scientifically based, they are also among the most difficult to implement. The construction of the anatomical facial structure is an extremely tedious task which requires artistic skills and massive computation.

3. MPEG-4 Facial Animation

A widely used and validated parametrization for synthetic characters is the one defined inside the MPEG-4 specification, namely MPEG-4 Facial and Body Animation (FBA). Such a standard makes use of three main sets of parameters to specify a face model in its neutral state (Pandzic & Forchheimer, 2002).

The Facial Animation Parameters (FAPs) are used to direct control the face movements. They are based on the study of minimal perceptible actions (MPA) and are closely related to muscle actions, such as movements of lips, jaw, cheeks and eyebrows. They make up a complete set of basic facial actions that represent the most natural facial expressions. FAPs define 68 parameters. The first 2 are high level parameters representing visemes and facial expressions. Viseme is the visual counterpart of phonemes in speech while facial expressions consists of a set of 6 basic emotions for anger, joy, sadness, surprise, disgust and fear as prototypes. The rest of the low level FAPs deal with specific regions on the face, like right corner lip, bottom of chin, left corner of left eyebrow.

The Facial Definition Parameters (FDPs) are needed for the calibration of a synthetic face. These parameters are scalable; they can define the shape, texture or even the whole facial polygon mesh.

The Feature Points (FPs) are used to describe and define the shape of a standard face. There are a total of 84 feature points in a head model. They are subdivided in groups, mainly depending on the particular region of the face to which they belong. Each of them is labelled with a number identifying the particular group to which it belongs and with a progressive index identifying them within the group. A subset of these points can be affected by the facial animation parameters (FAPs) to control the animation.

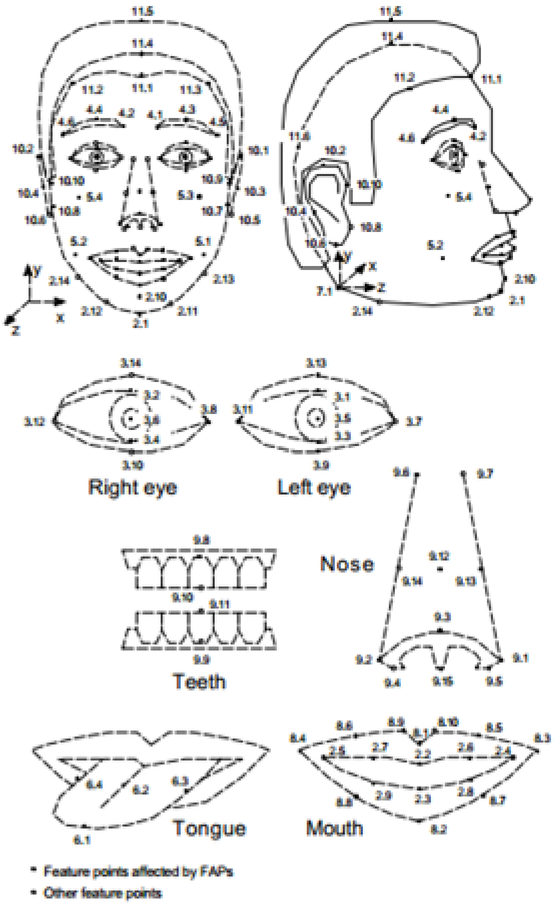

Figure 1: The 84 Feature Points (FPs) defined on a neutral face (Pandzic & Forchheimer, 2002)

4. Approach

Our approach aims to create plausible facial animation with a variety of facial expressions by automatically contracting a group of individual muscles and rotating the jaw mesh. To elaborate the mimic musculature, we have relied on Water's muscle model in which two types of muscles are defined: linear muscles that pull and sphincter muscles that squeeze. The mapping of such musculature to the face model is achieved by identifying the key nodes of each muscle with the appropriate set of MPEG-4 features points. The selection of FPs is done according to the anatomical properties that characterize the given muscle. In fact, the use of MPEG-4 features points as key nodes of each muscle will certainly reduce the amount of work that must be done manually by animators.

4.1 Muscle Modeling

To emulate the behaviour of muscles upon skin, Waters presents one of the most popular and complete parametric muscle models that are based on the human facial anatomy. This model is computationally cheap and easy to implement, it includes two types of muscles, linear and sphincter, independent of the bone structure. Each of these muscles can be defined by two key nodes, an area of influence which presents a skin portion affected by the contraction, and a deformation formula for all influenced vertices.

The linear muscle is modeled as a vector from a bony attachment point that remains static, to an insertion point which is embedded in the soft tissue of the skin. Its influence area is represented by a cone shape (Fig.2). For example the Zygomaticus Major which acts to draw the angle of the mouth up and back to smile or laugh, is a linear muscle.

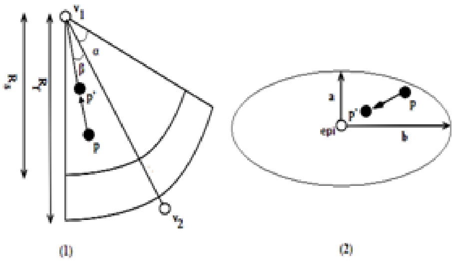

Figure 2: Linear and sphincter muscles

When a linear muscle contracts, all points in its influence area are displaced towards its point of attachment. The displacement of a point p affected by the muscular action is given by the Equation 1 (k is a fixed constant representing the elasticity of the skin).

$$p' = p + a\, r\, k\, \overrightarrow{pv_1} \quad (1)$$

$$a = \frac{(\cos\beta - \cos\alpha)}{(1 - \cos\alpha)}$$

$$r = \begin{cases} \cos\left(\frac{1 - \|v_1 p\|}{R_s} \cdot \frac{\pi}{2}\right), & \|v_1 p\| < R_s \\ \cos\left(\frac{\|v_1 p\| - R_s}{R_f - R_s} \cdot \frac{\pi}{2}\right), & else \end{cases}$$

For sphincter muscles, we can identify only two instances in a human face: the Orbicularis Oculi muscle around each eye and the Orbicularis Oris which circles the mouth. This kind of muscle attaches to the skin both at the origin and at the insertion. Its influence area has an elliptical shape defined by a virtual center and two semi-axes (Fig. 2). When a sphincter muscle contracts, the points in its influence area are displaced towards the center of the spheroid. The displacement of a point p affected by the action of muscle is given by the Equation 2 (k is a fixed constant representing the elasticity of the skin).

$$p' = p + r\, k\, \overrightarrow{po} \quad (2)$$

$$r = \cos\left(\left(1 - \frac{\sqrt{p_x^2 b^2 + p_y^2 a^2}}{ab}\right) \cdot \frac{\pi}{2}\right)$$

It is important to note that Waters combined the muscle actions sequentially by applying the displacements caused by them on a vertex one by one. However, such a process can produce undesirable effects especially when a mesh vertex is under the influence of multiple muscle actions: the vertex will be shifted outside the influence area of adjoining muscle vectors. To avoid the undesired effects, we have used the Wang approach (Wang, 1993) which summarizes the displacements and then applies it to the vertex.

4.2 Muscle Construction

Our facial musculature comprises essentially 31 muscles including three sphincter muscles that are used to represent the orbicularis oris and orbicularis oculi, and 11 pairs of linear muscles that are placed symmetrically through the face to accomplish the major face movements. The remaining linear muscles, namely Cheek Sup, Cheek Center and Cheek Inf, are located on each cheek and don't exist in a real human face. They have been added to our model to simulate specialized expressions in sign languages like cheek movements. The complete facial muscle structure is shown in Fig.3. The face model is represented as a single layered mesh with no skeleton, it is expected to consist of triangular or quadratic polygons, a high-poly or low-poly mesh.

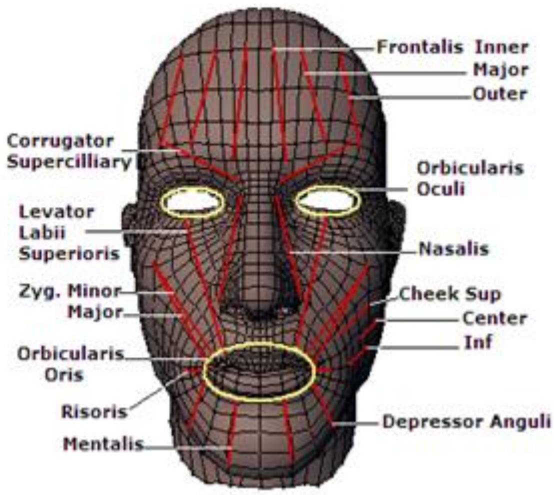

Figure 3: Facial musculature

The construction of the proposed musculature involves two basic steps. At first, the anatomical positions of the muscle control points will be defined with the suitable MPEG-4 Feature Points. Second, the set of vertices that belong to each influence area will be detected.

For a linear muscle, three points are needed to define its location on the input face mesh: an attachment point AP, an insertion point IP and a reference point RP which was not used in Waters model, we have added it in our method to facilitate the determination of the mesh part affected by the muscle action. The obtained properties of muscle vectors are given in Table 1. For instance, the Left Nasalis which depresses the cartilaginous part of the nose is characterized by:

- an AP which coincides with the feature point 9.7 located in the left upper edge of the nose bone
- an IP which coincides with the feature point 9.1 located in the left nostril border.
- an RP which coincides with the feature point 9.3 located in the nose tip .

Muscle name	**Attachment point**	**Insertion point**	**Reference point**
Frontalis Inner	$11.1+\frac{1}{6}(11.2-11.1)$	4.2	$4.2+\frac{1}{2}(4.4-4.2)$
	$11.1+\frac{1}{6}(11.3-11.1)$	4.1	$4.1+\frac{1}{2}(4.3-4.1)$
Frontalis Major	$11.1+\frac{2}{3}(11.2-11.1)$	$4.2+\frac{1}{2}(4.6-4.2)$	4.6
	$11.1+\frac{2}{3}(11.3-11.1)$	$4.1+\frac{1}{2}(4.5-4.1)$	4.5
Frontalis Outer	11.2	4.6	4.4
	11.3	4.5	4.3
C.S	$4.2+\frac{1}{2}(4.2-4.4)$	4.4	$4.2+\frac{1}{2}(4.2-3.8)$
	$4.1+\frac{1}{2}(4.1-4.3)$	4.3	$4.1+\frac{1}{2}(4.1-3.11)$
Nasalis	9.6	9.2	9.3
	9.7	9.1	9.3
Levator Labii Superioris	3.10	2.7	$8.9+\frac{1}{3}(8.6-8.9)$
	3.9	2.6	$8.10+\frac{1}{3}(8.5-8.10)$
Z. Minor	9.2	2.7	8.4
	9.1	2.6	8.3
Z. Major	5.4	2.9	9.15
	5.3	2.8	9.15
Risoris	5.2	2.5	$2.5+\frac{1}{3}(2.5-9.2)$
	5.1	2.4	$2.4+\frac{1}{3}(2.4-9.1)$
Depressor Anguli Oris	$8.4+(8.4-8.6)$	8.6	8.8
	$8.3+(8.3-8.5)$	8.5	8.7
Mentalis	$2.1+\frac{1}{2}(2.1-2.12)$	2.9	8.2
	$2.1+\frac{1}{2}(2.1-2.11)$	2.8	8.2
Cheek Inf	$K+\frac{1}{2}(K-2.11)$	$K=5.1+(5.1-8.2)$	5.1
Cheek Center	$K+\frac{1}{2}(K-2.13)$	$K=5.1+(5.1-8.2)$	5.1
Cheek Sup	$K+\frac{1}{2}(K-2.11)$	$K=5.1+(5.1-8.3)$	5.3

Table 1: Linear Muscle Properties (Right and Left)

Similarly, each sphincter muscle is defined by three key points: the epicenter of the spheroid EP, a semi-major axis SJ and a semi-minor axis SN. The obtained properties of sphincter muscles are shown in Table 2. For instance, the Orbicularis Oculi Left which closes the eyelids of the left eye is characterized by:

- an EP which coincides with the midpoint of the line segment formed by the two points 3.7 and 3.11
- SJ is equal to the length of the line segment formed by the two points 3.7 and EP
- SN is equal to the length of the line segment formed by the two points 3.9 and EP.

Muscle name	**Semi-major axis**	**Semi-minor axis**	**Epicenter**
Orbicularis Oculi	3.12	3.10	$\frac{1}{2}(3.8+3.12)$
	3.7	3.9	$\frac{1}{2}(3.7+3.11)$
Orbicularis Oris	8.3	8.2	$\frac{1}{2}(8.3+8.4)$

Table 2: Sphincter Muscle Properties

Once the positions of muscle control points are computed and mapped on the facial mesh, we can determine then the set of vertices which will be influenced by the muscle contraction. For a linear muscle, all influenced vertices should match the following conditions (see Fig. 2):

$$\|\overrightarrow{pv_1}\| > 0,\ \|\overrightarrow{pv_1}\| \le \boldsymbol{Rf},$$
$$\boldsymbol{\cos\beta} \ \ge\ \boldsymbol{\cos\alpha}$$

Similarly, all influenced vertices of a circular muscle should be within its spheroid (see Fig. 2).

$$\left(\frac{p_x - epi_x}{a}\right)^2 + \left(\frac{p_y - epi_y}{b}\right)^2 + \left(\frac{p_z - epi_z}{\sqrt{a^2 - b^2}}\right)^2 < 1$$

Fig. 4 shows the obtained result after applying the proposed algorithm to the Nasalis muscle. The face on the left illustrates the set of vertices having a distance from AP less than the length of muscle fiber. The second face illustrates the set of vertices that will be displaced when muscle contracts.

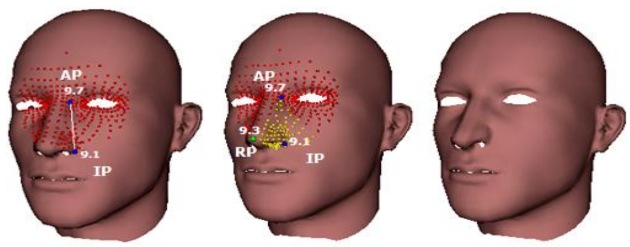

Figure 4: The set of vertices that belong to the influence area of Nasalis Left is colored in yellow

4.3 Jaw Articulation

As we mentioned above, the face model does not have a skull, so the jaw is not a particular mesh, it will be rather detected automatically from the initial mesh. To do so, we have used some features points for approximating the vertices of the chin and lower lip that are affected by the jaw rotation.

4.3.1. Jaw Detection

To define the chin vertices, we have taken the following steps: first, we project the facial mesh on the plane P passing through the midpoint of the segment joining the FPs 10.8 and 10.7, with the normal vector defined by this midpoint and the vertex 10.8 in order to get a profile view. Second, the projections of 2.14, 10.8 and 8.3 respectively p1, p2 and p3 are marked on the projected mesh. The vertices whose projects are inside the angular sector $(p1,\widehat{p2},p3)$ are considered to be in the chin influence (Fig. 5).

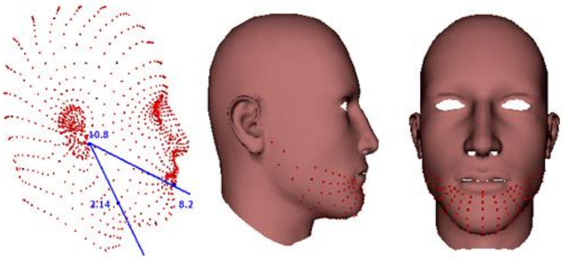

Figure 5: Detecting chin vertices

The process is about the same for the lower lip but by using other feature points. It is important to note here that the vertices located on the inner contour of the lower lip should also be taken into account since the projection is incapable of detecting them. The extraction of these vertices is done as follows: the algorithm will browse all the edges of the mesh to find those that belong only to one surface. The selected edges depict the contours of the facial model such as the openings of the eyes, nose as well as the space between the two lips. Using this set of edges and some MPEG-4 FPs, we can distinguish the inner contour of the lower lip. All we have to do is finding the closest edge to the point 2.3, and then its neighboring segments (Fig. 6). For each new segment found, we perform the same process until finding edges closest to the points 8.3 and 8.4 which define the corners of the lips.

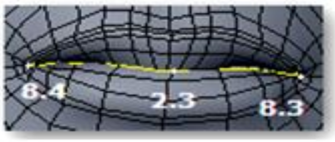

Figure 6: The inner lower lip contour

4.3.2. Jaw Rotation

To rotate the jaw, the vertices of the chin and the lower lip are rotated around a line passing through the feature point 10.8 and parallel to the X axis (Fig. 7). The final positions are calculated by the following equation where φ represents the degrees of rotation and (x1, y1, z1) the coordinates of 10.8.

$$\begin{pmatrix} x \\ y' + y1 \\ z' + z1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} * \begin{pmatrix} x \\ y - y1 \\ z - z1 \end{pmatrix} \quad (3)$$

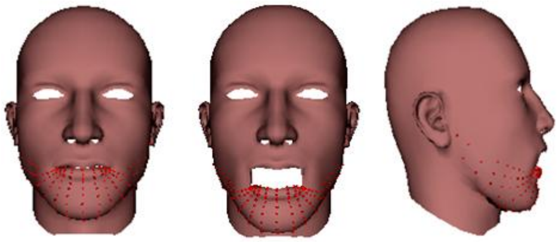

Figure 7: Jaw rotation

5. Results & Evaluation

The production of emotions is the result of a contraction or relaxation of one or more facial muscles. Fig. 8 shows some examples of basic facial expressions (happiness, anger, sadness, surprise, disgust and fear) on different face models, while Fig. 9 illustrates the contraction of Cheek Sup, Cheek Center and Cheek Inf which are used to emulate the tongue motion on the right cheek.

Figure 8: The simulation of some basic facial expressions on different face models

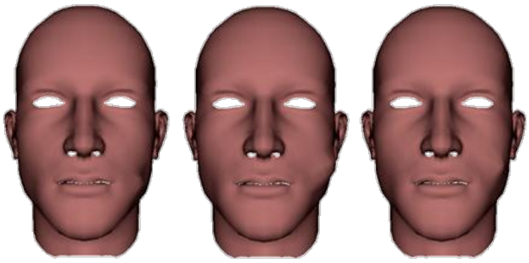

Figure 9: The contraction of cheek Sup, Cheek Center and Cheek Inf

In order to validate our approach, we have tested the performance of facial muscles in different face models. The goal of this evaluation is to check the choice accuracy of features points in the definition of muscle key nodes as well as the fitting of those features with anatomically correct positions on the face. Fig. 9 shows the recognition rate of each muscle motion is calculated with 50 models.

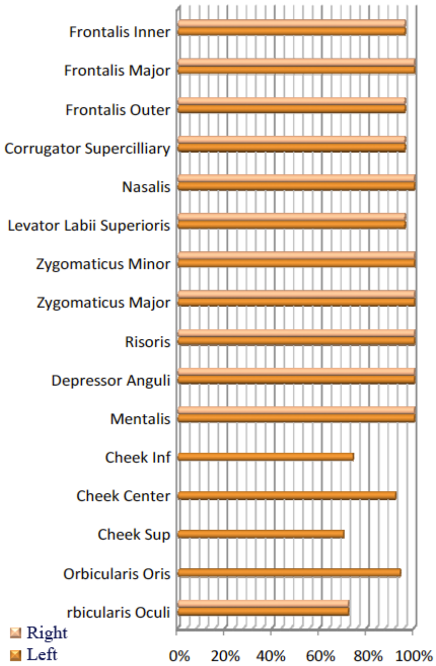

Figure 10: The recognition rates of muscle motion on 50 face models

We can notice that the contraction effects of the vast majority of muscles have been simulated. For the Frontalis Major, Nasalis, Zygomaticus Major, Zygomaticus Minor and Risoris, our approach achieves 100% recognition rate, whereas, the Cheek Inf, Cheek Sup and Orbicularis oculi achieve recognition rates ranging from 70% to 75%. This is can be explained by the fact that muscular activities do not give the desired animations for some meshes.

On the other hand, in order to study the sensitivity of muscle performance on high-poly and low-poly meshes, we have used three sets of face models that have the same appearance, the same polygonal resolution, but with different number of vertices: less than 1000 vertices, between 1000 and 4000 vertices and over than 4000 vertices. The obtained result is drawn in Fig.11. It is clear that most muscles are insensitive to the changes in the number of vertices, with the exception of Orbicularis Oculi. This is can be explained by the fact that in low-poly meshes, the eyelids and eyebrows have common polygons.

It should be noted that the proposed method depicts one module of the WebSign project (ElGhoul & Jemni, 2008) (Othman, ElGhoul & Jemni, 2011) which renders sign language animations in real time using a virtual avatar, from a writing text or a SignWriting notation (Bouzid & Jemni, 2013). The control of the muscles and jaw articulator is done via a scripting language called SML (Sign Modeling Language). Examples of face movements rendered by our virtual avatar are illustrated in Fig. 12.

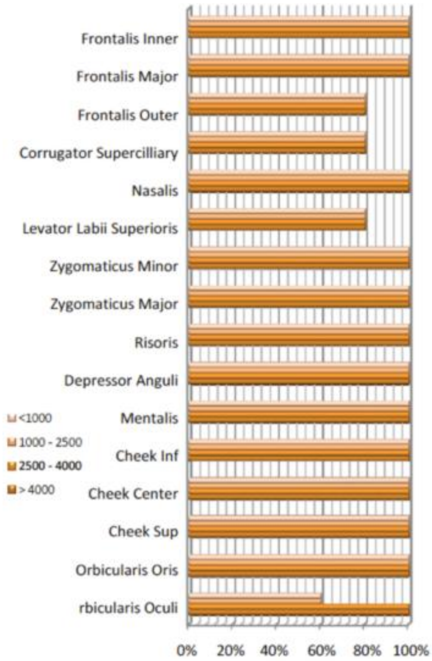

Fig. 11 The recognition rates of muscle motion on low and high poly meshes

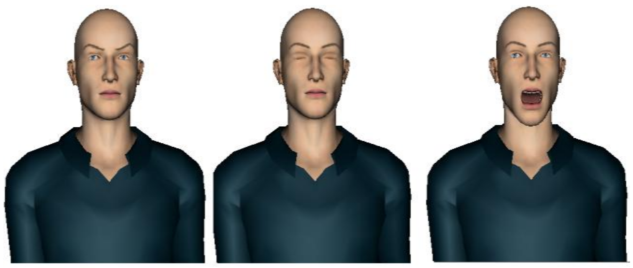

Fig. 12 Examples of face movements

6. Conclusion

We have presented an automatic facial animation method based on Waters vector model and some MPEG-4 feature points. The experimentation shows that more than 90% of tested facial actions can be animated without any human intervention. However, for some low-poly meshes we need to adjust the influenced area of sphincter muscles. To this end, we aim in our future work to modify the sphincter muscle model by ameliorating the algorithm used to detect its influence area.

7. References

Bouzid, Y., ElGhoul, O., Jemni, M. (2013). Synthesizing Facial Expressions for Signing Avatars using MPEG-4 Feature Points. In Proceedings of the 4th International Conference on Information and Communication Technology & Accessibility, Tunisia

Bouzid, Y., Jemni, M. (2013). An avatar based approach for automatically interpreting a sign language notation. In Proceeding of the 13th IEEE International Conference on Advanced Learning Technologies, China
Deng, Z., Noh, J. (2008). Computer facial animation: a survey. In Data-Driven 3D Facial Animation, London: Springer-Verlag, ch. 1, pp. 1-28.
Ekman, P., Friesen, W. (1977). Manual for the Facial Action Coding System. Consulting Psychologist, Palo Alto, California
Elghoul, O., Jemni, M. (2009). Automatic generation of sign language's facial expression. In Proceeding of ICTA 2009, pp143-146, Tunisia
Fratarcangeli, M. (2005). Physically based synthesis of animatable face models. Workshop on Virtual Reality Interaction and Physical Simulation
Ilie, M.D., Negrescu, C., Stanomir, D. (2012). Energy minimization tool for generating composite facial expressions in 3D facial animations, International Journal of Computer Theory and Engineering
Kähler, K. (2007). 3D Facial animation-Recreating human heads with virtual skin, bones, and muscles. VDM Verlag
Lee, Y., Terzopoulos, D., Waters, K. (1993). Constructing physics-based facial models of individuals. In Proceedings of Graphic Interface 1993, pp. 1-8
Liu, C. An analysis of the current and future state of 3d facial animation techniques and systems. (2009). M. S. thesis, School of Interactive Arts and Tech., Simon Fraser Univ., Burnaby, BC, Canada
Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee, R. (2000). The syntax of American Sign Language: functional categories & hierarchical structure. Cambridge: MIT Press.
Orvalho, V., Bastos, P., Parke, F., Oliveira, B. (2012). A Facial Rigging Survey. In EUROGRAPHICS 2012
Othmen, A., Elghoul, O., Jemni, M. (2012). An automatic approach for facial feature points extraction from 3D head. In GRAPP/IVAPP 2012, pages 369-372, Italy
Parke, F. I. (1972). Computer generated animation of faces. In Proceedings of the ACM Annual Conference, vol. 1, pp. 451-457
Parke, F. I., Waters, K. (2008).Computer Facial Animation, Wellesley, MA: AK Peters Ltd, pp. 85-362.
Pandzic, I. S., Forchheimer, R. (2002). MPEG-4 facial animation: the standard, implementation and applications, New York
Ping, H. Y., Abdullah, L. N., Sulaiman, P. S., Abdul Halin, A. (2013). Computer Facial Animation: A Review. International Journal of Computer Theory and Engineering.
Platt, S. M., Badler, N. I. (1981). Animating facial expressions. In Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques, pp 245-252. ACM New York, USA
Terzopoulos, D., Waters, K. (1990). Physically-based facial modeling, analysis, and animation. Journal of Visualization and Computer Animation, vol. 1, no. 4, pp. 73-80
Wang, C. L. (1993). A hierarchical spline based facial animation system with simulated muscles. Master's thesis, University of Calgary
Waters, K. (1987). A muscle model for animation three-dimensional facial expression. In Proceedings SIGGRAPH Computer Graphics.