

# A PERFORMANCE BASED PARAMETRIC MODEL FOR FACIAL ANIMATION

*Anna Wojdel and Leon J. M. Rothkrantz*

Knowledge Based Systems Group,  
Faculty of Information Technology and Systems,  
Delft University of Technology,  
Zuidplantsoen 4  
2628 BZ Delft, The Netherlands  
A.Wojdel@cs.tudelft.nl, L.J.M.Rothkrantz@cs.tudelft.nl

## ABSTRACT

In this paper we present a parametric model for facial animation and a method for adapting it to a specific person. Every facial expression can be described as a contraction or relaxation of the facial muscles. P. Ekman and F.W. Friesen selected 44 Action Units corresponding to visual changes on the face, which cannot be decomposed into smaller ones and which combinations universally represent all facial expressions. Our model for facial animation is built on the basis of that Facial Action Coding System (FACS). The model adaptation is based on performance measurements of the subject's facial movements. Our model combines the advantages of parametric animation models such as wireframe model independency with the accuracy of face movements reproduction (cloning) obtained with performance driven models. The described model forms a part of the facial animation system that is currently under development at Delft University of Technology. A brief description of this system is also given in this paper.

## 1. INTRODUCTION

In the nineties, virtual reality became a very popular topic in computer science. Researchers around the world try to create virtual worlds which do not only look but also “behave” as realistically as possible. Especially the creation of a virtual human is a very interesting part in it. We need the virtual characters in entertainment industry (movies and computer games) as well as in more “serious” industries: virtual humans can be used for medical purposes (in speech distortions therapy or prediction of plastic surgery) or in virtual learning and teaching.

The first 3D model of the human face was proposed in 1972 by F. Parke [1]. Animation was based on morphing between two masks previously generated and stored in the library of masks. Since that time, a lot of different 3D facial models were created. Some of them are still based on a very simple key-frame modeling (as in the first Parke's model) or parametrization introduced two years later also by F. Parke [2] and expanded further e.g. by K. Waters [3]. In the nineties however, new models – based on the anatomy of the face, structure and functionality of facial muscles – became more and more popular [4]. This physically-based models are most used in applications where we need high precision and realism as e.g. in prediction of plastic surgery results [5], while simpler ones are still used in a real-time animation applications [6].

Another important field in facial animation area is designing systems, which will automatically generate appropriate facial animation according to a given text [7, 8]. Such systems are usually based on a set of rules, which describes dependencies between the prosody and facial expressions. Those rules are based on the work of many psychologists, who tried to describe universal links between facial expressions and the verbal content of the message [9, 10]. They are the same for each person – independent from his/her age, temper or customs. But, we know, that those factors – related to specific behavioral pattern – influence a way of talking and leading a conversation. Moreover, the same person can say the same sentence in completely different ways, depending on his/her feelings in given a moment of time or outside conditions.

Therefore it seems for us, that systems for automatic generating facial animation are too universal in order to generate realistic – person specific animation. For that reason, our project is aimed at developing a system for *semiautomatic* facial animation. Our goal is to create an environment with a synthetic 3D face able to show some expressions where the user supported by the system can generate person-specific facial animations. We want to give a user the possibility to use a facial expressions' script language. With this script language the user can create an animation of a given face according to a text just by putting emblems representing appropriate expressions in a chosen place of the text.

In the following sections we will give an overview of the system for facial animation, which we currently develop. Further we will concentrate on the part of the system which is responsible for the model of facial movements. We will present the design of the model as well as how this model can be adapted to a specific person.

## 2. SYSTEM OVERVIEW

In 1970 P. Ekman and F.W. Friesen developed the universal Facial Action Coding System (FACS) based on Action Units (AUs), where each facial expression can be described as a combination of those AUs [10]. Because it became a standard in facial expressions analysis, we decided to base facial movements on those Action Units. The whole idea of the system for generating facial animation is based on a “facial expressions script language”, where basic variables are Action Units. According to P. Ekman and W.F. Friesen, using AUs all facial expressions can be shown. They function as characters in a “normal” language.

When we have AUs as characters we can define words of our

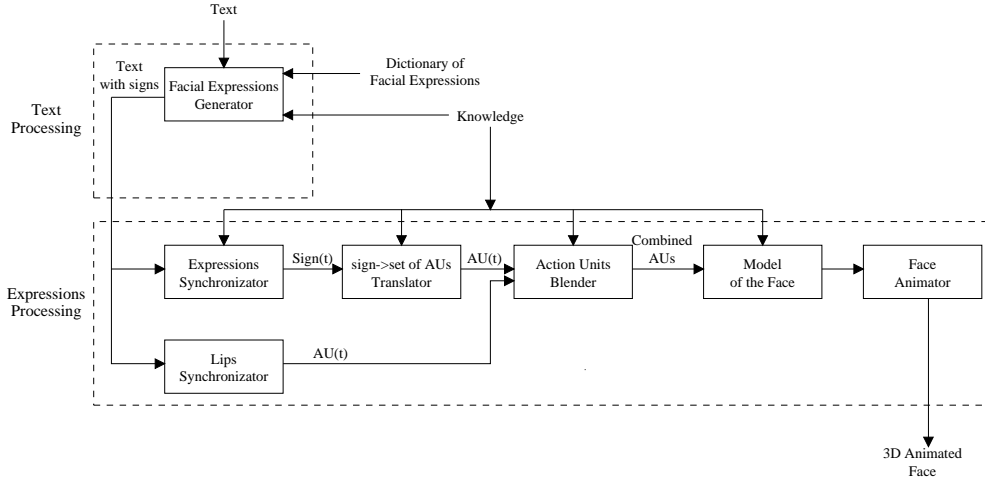


Figure 1: Design of the system for generating facial expressions

script language: facial expressions. Facial expressions, the same as words in “normal” language have their own syntax and semantics. Syntax is just a set of AUs which are involved showing a given facial expression, while the semantics of a facial expression is the description of a meaning of the facial expression in a verbal way (e.g. “ironical smile - mostly used while...”). All of the defined facial expressions will be collected in a nonverbal dictionary of facial expressions. Each position in this nonverbal dictionary will contain an emblem of a given facial expression, a description of the syntax (which AUs are activated) and semantics (what this expression means) as well an example of a synthetic face showing this facial expression. We have already developed the first version of the nonverbal dictionary of facial expressions. The current version of the dictionary contains about 200 facial expressions together with corresponding emblems and descriptions of their syntax.

But of course, the set of separate words doesn’t create a language. There has to be also a definition of a grammar. Grammar of script language should describe how to compose facial expressions together. That means, that grammar should “control” which facial expressions (e.g. with opposite meaning) cannot occur in the same time or close to each other; which facial expressions usually occur next to each other or which facial expressions usually occur in specific place of the sentence. Contrary to syntax and semantics, the grammar won’t be exposed to the user. The system will use it internally to support generating appropriate facial animation.

Summarizing; our design of the system for facial animation is as follows: it has a modular structure, where each module is dedicated to a given task and each module uses his own knowledge about dependencies between facial expressions for its level. The schematic design of the system is presented in Figure 1. More about design of the system can be found in [11].

### 3. MODEL DESIGN FOR A SINGLE ACTION UNIT

Each AU can be described in verbal terms in the way that it is observed on the face. For example one can describe the area of influence of the AU, how this influence changes within the defined area and finally what is the direction of changes. In order to implement any given AU in our system we need to transform this description

into some mathematical terms. In order to fulfill this requirements we propose the following components of the AU model:

$\varphi(\mathbf{v}) : \mathfrak{R}^3 \rightarrow \mathfrak{R}^+ -$  density function, which describes in fact both the area of AU occurrence and “density” of the movement inside this area,

$\Psi(\mathbf{v}) : \mathfrak{R}^3 \rightarrow \mathfrak{R}^3 -$  the direction of the movement triggered by an AU,

$\tau \in [0, 1] -$  the value of the activation intensity of a given AU.

Using the above components, for each point  $\mathbf{v}$  on the face surface its displacement under influence of a given AU can be calculated from the following formula:

$$\Delta \mathbf{v} = \Psi(\mathbf{v})\varphi(\mathbf{v})\tau \quad (1)$$

where:  $\Delta \mathbf{v}$  is a vector of displacement,  $\mathbf{v}$  contains the coordinates of a vertex,  $\Psi(\dots)$  is a direction function which defines the direction of the movement,  $\varphi(\mathbf{v})$  is the density function which defines how much a given vertex will be moved and  $\tau$  the value of the activation intensity of the given AU.

Most of the AUs occurring on a human face can be described using formula (1) even though it assumes linear dependency between AU intensity and effective displacement. This doesn’t hold however for AUs that incorporate long movements on a large areas, where nonlinearity becomes evident. Such AUs are e.g. head movements (AU51 – AU56). In this case a more generic formula can be used:

$$\Delta \mathbf{v} = \Psi(\mathbf{v}, \varphi(\mathbf{v}), \tau) \quad (2)$$

The above proposed formalization of the AU description can be used in semi-automated implementation of the AU for a given person. The implementation proceeds in three steps which will be described below.

The first step is to make 3D measurements of the real human face with a given AU 100% activated. Some selected points on the face and their movements have to be measured. It is advisable that the measured points relate somehow to the used wireframe, but it is not absolutely necessary. The only important thing is that the measurements describe accurately changes on the face when applying the given AU.

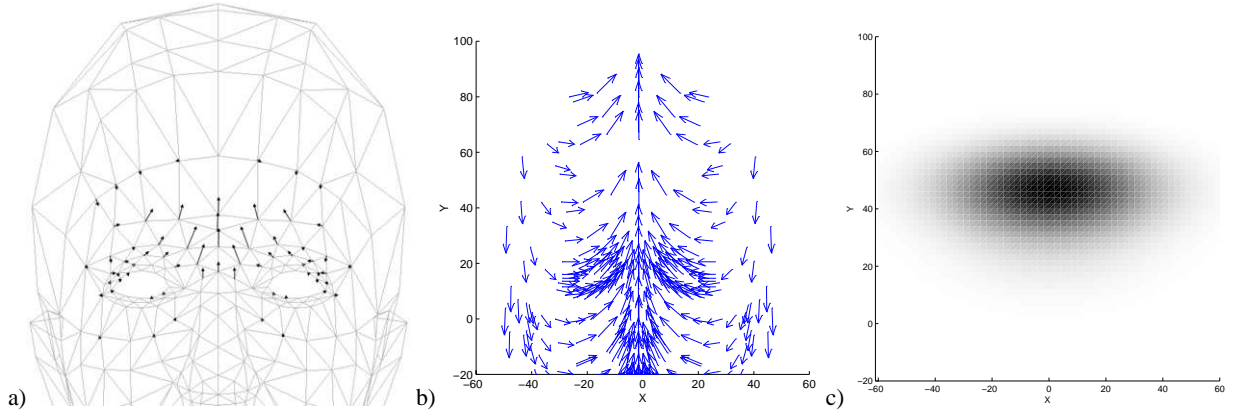


Figure 2: Model deformation for AU1. (a) manual adaptation of the model, (b)  $\Psi$  function, (c)  $\varphi$  function

In the second step, a hypothesis on the generic form of functions  $\Psi$  and  $\varphi$  must be stated. This hypothesis should be based on the character of changes inflicted on the face (as measured in the previous step). While  $\Psi$  and  $\varphi$  heavily depend on the AU itself, their generic forms can be defined once for all subjects. For example, independently on the person modelled, the  $\Psi$  function for AU51 (head turn left) will always be a 3D rotation and  $\varphi$  will be a smoothed step function with values changing from 1 to 0 between the chin and bottom of the neck. In this way, once the form of those functions is defined for a given AU, it can be reused (with different parameters) for modelling different persons with different wireframe models.

Eventually the parameters of both functions characterizing a given AU must be adjusted so that the resulting displacement will optimally fit the measured data. In the generic case (2) the number of free parameters that have to be optimized could grow considerably with the complexity of the functions. The form (1) however was designed in such a way that each of the functions can be optimized independently. In this way we can first optimize the parameters of the  $\varphi$  function so that it fits the lengths of the measured displacements. In next step the parameters of the  $\Psi$  function can be optimized so that it fits the directions of the displacements. This approach provides a significant improvement in both speed and accuracy of optimization.

#### 4. EXAMPLE AU IMPLEMENTATION

In this section we will show, how the previously described procedure was used in order to obtain an implementation of AU1 (inner brow raiser). In order to obtain the necessary measurements, we asked a subject to show this AU and took pictures of a neutral and an AU1-showing face. We used 36 control markers on one side of the subject's face and we took simultaneously pictures of the frontal and lateral view of the face. In this way it was relatively easy to manually measure facial movements. Moreover, as control points we used also positions of such facial features as eye-contour and eye-brows. The results of those measurements are depicted on Fig. 2a.

To model a  $\Psi$  function we used the following formula:

$$\begin{aligned} \Psi(\mathbf{v}) &= [\cos(\alpha) \cos(\beta), \sin(\beta), \sin(\alpha) \cos(\beta)] \\ \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= \mathbf{A}\mathbf{v} + \mathbf{c} \end{aligned} \quad (3)$$

where  $\mathbf{v}$  is a point in 3D space,  $\mathbf{A}$  is a  $2 \times 3$  matrix and  $\mathbf{c}$  is a 2D vector. In this way the image of the mapping  $\Psi$  is a set of unit-length vectors with a linearly changing angle. The actual values of the parameters  $\mathbf{A}$  and  $\mathbf{c}$  were optimized to the measurements using Matlab toolkit. We used Nelder-Mead method for nonlinear unconstrained minimization, and minimized the following function:

$$E_{\Psi} = \sum_{i=0}^n |(\Delta \mathbf{v}_i - |\Delta \mathbf{v}_i| \Psi(\mathbf{v}_i))| \quad (4)$$

where  $\mathbf{v}_i$  is  $i$ -th measured point and  $\Delta \mathbf{v}_i$  is measured movement of this point. In this way, the error in the direction of the vector was weighted by the extent of its movement. The resulting  $\Psi$  function is depicted on Fig. 2b.

As a density function  $\varphi$  we used a Gaussian shape:

$$\varphi(\mathbf{v}) = \eta e^{-\frac{(\mathbf{v}-\mathbf{m})^T \mathbf{B} (\mathbf{v}-\mathbf{m})}{2}} \quad (5)$$

where  $\eta$  is a real number,  $\mathbf{m}$  is a 3D vector and  $\mathbf{B}$  is a  $3 \times 3$  matrix. Again those 3 parameters were optimized to fit the measured data. This time, only the length of the movement was optimized, so the goal function was:

$$E_{\varphi} = \sum_{i=0}^n (|\Delta \mathbf{v}_i| - \varphi(\mathbf{v}_i))^2 \quad (6)$$

The resulting density of the AU1 is depicted in Fig. 2c.

Finally, both functions were applied to the wireframe model of the subject's face according to the formula (1). On Figure 3c we present the resulting rendering of the deformed wireframe. For comparison, Figure 3a contains the original subject's picture and Figure 3b a rendered model deformed according to exact markers' measurements.

#### 5. VALIDATION OF THE MODEL

In order to validate our model, we assumed that the difference between positions of the vertices in the wireframe for the neutral face and the wireframe manually adapted to a specific AU are adequate to the 3D measurements taken on the real face. The method for taking those measurements is not defined in our model. Therefore we do not validate here the accuracy of taken measurements.

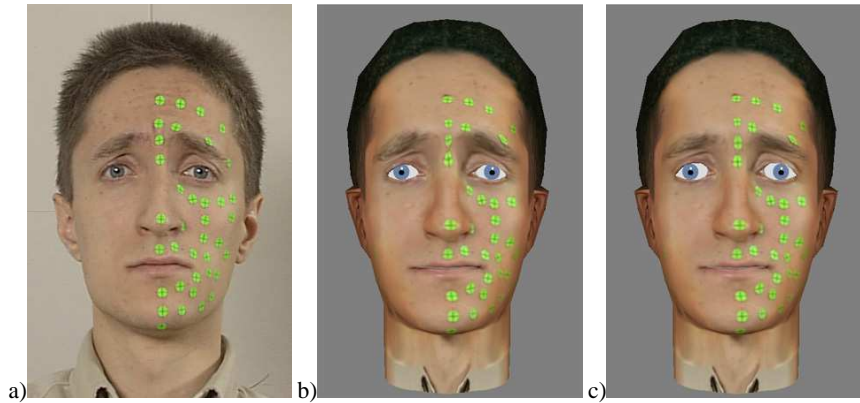


Figure 3: Activation of AU1. (a) original photo, (b) manually adapted model, (c) parametrized deformation of the model

The goal of this evaluation is to validate accuracy of choice of the generic forms of the direction and density functions ( $\Psi$  and  $\varphi$  respectively) as well as the fitting of those functions.

An error was calculated as a distance between positions of the vertices in the manually deformed wireframe and the wireframe obtained after application of our model. The average displacement error for a single vertex is 0.7074 units which is equivalent to 1.36 mm on the real face. For comparison, the average movement on the face is 1.8353 (3.54 mm). For different AUs it varies between 0.3049 units (0.59 mm) and 1.5427 units (2.97 mm). It seems, that the average error depends on the size of the area of AU occurrence, but not on the number of vertices, which were displaced. If the area of occurrence is large, the average error is remarkably higher. In 75% of the implemented AUs, the average error is lower than 1.45 mm and only 25% of those AUs produce the average error higher than that.

We can also compare the mean displacement error to the maximal facial movement for a given AU; which is the most important in what we see as a result. This ratio is for the most of implemented AUs about 14% with the minimal and maximal values respectively 6% and 17%.

Another interesting conclusion can be done when we compare the maximal displacement error and the maximal movement for each AU. We can observe, that for AUs with relatively small maximal movement (less than 9 mm) the ratio of the maximal error to the maximal movement is higher (about 44%) than for the rest of AUs (about 32%). It indicates, that our method provides better results for AUs with bigger facial movements and worse for subtle facial changes.

## 6. CONCLUSIONS

In this paper we described a generic facial animation model that can be used for a person specific facial animation. Proposed approach incorporates fitting of the generic facial model on measured facial changes of a specific person. It is independent of the measurement technique and wireframe model that are used. Therefore it can be widely used in virtual person cloning applications. In order to use our model, we need measurements of a subject's facial movements and we have to model his/her face with a wireframe. These aspects were so far done manually, so automatization of those processes remain for future development. An other thing that wasn't covered in this paper is the blending of AUs. But it is a

separate module in our system and it has to be developed independently.

## 7. REFERENCES

- [1] F. I. Parke, "Computer generated animation of faces," in *Proceedings of the ACM National Conference*, pp. 451–457, 1972.
- [2] F. I. Parke, "Parametrized models for facial animation," *IEEE Computer Graphics*, vol. 2, no. 9, pp. 61–68, 1982.
- [3] K. Waters, "A muscle model for animating three-dimensional facial expressions," *Computer Graphics (SIGGRAPH'87)*, vol. 21, pp. 17–24, July 1987.
- [4] Y. Lee, D. Terzopoulos, and K. Waters, "Realistic modeling for facial animation," in *Computer Graphics Proceedings, Annual Conference Series*, pp. 55–61, 1995.
- [5] R. M. Koch, M. H. Gross, and A. A. Bosshard, "Emotion editing using finite elements," Technical Report 281, ETH Zürich, Institute of Scientific Computing, January 1998.
- [6] N. Magnenat-Thalmann, P. Kalra, and I. S. Pandzic, "Direct face-to-face communication between real and virtual humans," *International Journal of Information Technology*, vol. 1, no. 2, pp. 145–157, 1995.
- [7] C. Pelachaud, N. I. Badler, and M. Steedman, "Generating facial expressions for speech," *Cognitive Science*, vol. 20, no. 1, pp. 1–46, 1996.
- [8] C. Pelachaud and S. Prevost, "Sight and sound: Generating facial expressions and spoken intonation from context," in *Proceedings of the second ESCA/AAAI/IEEE Workshop on Speech Synthesis*, (New Paltz, NY, USA), September 1994.
- [9] E. Hajicova and P. Sgall, "The ordering principle," *Journal of Pragmatics*, pp. 435–454, 1987.
- [10] P. Ekman and W. F. Friesen, *Unmasking the Face*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, Inc., 1975.
- [11] A. Wojdel and L. J. M. Rothkrantz, "Intelligent system for semiautomatic facial animation." to be published in *Proceedings of Euromedia'2000*, May 2000.