# Automatic Facial Animation Analysis

Daniel Bardsley, University of Nottingham

## *Introduction*

A number of recent research projects have been focusing their attentions to the open-ended problem of generating realistic facial animations from 3D head models. Initially we discuss the motivation for researching automated facial animation methods. Secondly we investigate a variety of traditional and cutting edge approaches to the problem.

## *Aims*

The primary aim of automated facial animation is to allow us to take a 3D head model of a subject in a single expression (usually neutral) and to generate a range of expressions and facial poses without requiring the intervention of a traditional animator. Given the ability to automatically generate a full range of expressions and head poses for a specific model we are able to use the model as a "talking head" for enhanced communication, within training scenario software, for use in games or movies or for face / expression recognition applications. It is highly desirable to fully automate the animation process since this reduces the need for animators to define key frames whilst vastly reducing the amount of time required to produce realistic virtual characters. Considering video games for a moment, in-game characters must respond dynamically and realistically to actions presented by the player. In a game that allowed highly detailed face models the large number of potential expression permutations may be difficult and time consuming to animate by hand. Furthermore scenarios in virtual worlds where a player may wish their in-game avatar to visually represent their real life appearance would benefit greatly if the system was able to reconstruct a head model and then accurately animate expressions mapping a players emotions to their virtual avatar. Finally, assuming a suitable model and animation structure it would be possible to map animations from one head model to another with little or no work, thus allowing facial animations to be reusable across a large range of head models.

The problem of animating a human face is by no means simple. The face contains 268 voluntary muscles of three differing types. On top of the muscles a layer of skin, which can slide over the muscles, may bunch up as muscles are contracted, causing ridges to form in the skin. This type of surface deformation is not easily captured using simple animation techniques and is not intuitively easy to model mathematically. Finally a face animation must take into consideration the structure of the scull. Thankfully the scull only contains a single joint (the jaw) and can be treated as a rigid body. All other expression is the result of soft tissues movements.

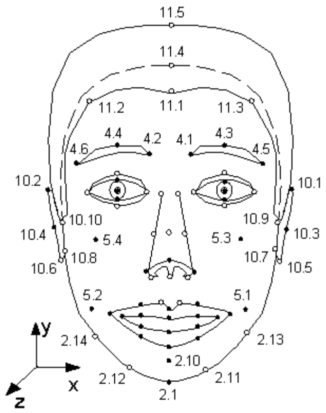
The traditional approach to facial animation involves the use of key frames to define each of the major emotions and facial expressions. It then becomes possible to use linear blending techniques to combine these key frames and to produce smooth transitions between each of the base states. Producing key frames for each model is a computationally inexpensive method of achieving facial animation, however, it requires a skilled animator to produce the base models. Furthermore, the key frames are not transferable between head models thus reducing their usefulness in real world applications. These drawbacks are what we are explicitly trying to avoid by investigating automated facial animation methods.

## *Muscle Models*

The first attempts to properly define human facial expression were carried out by Ekman and Friesan in 1978. They proposed the Facial Action Coding System (FACS). This describes facial movement in terms of the muscles involved. They define 46 action units pertaining to expression related muscles and a further 20 units for head movement and eye gaze. This method remains the basis for many head muscle models in use today. In the following sections we summarise the most commonly used facial animation models.

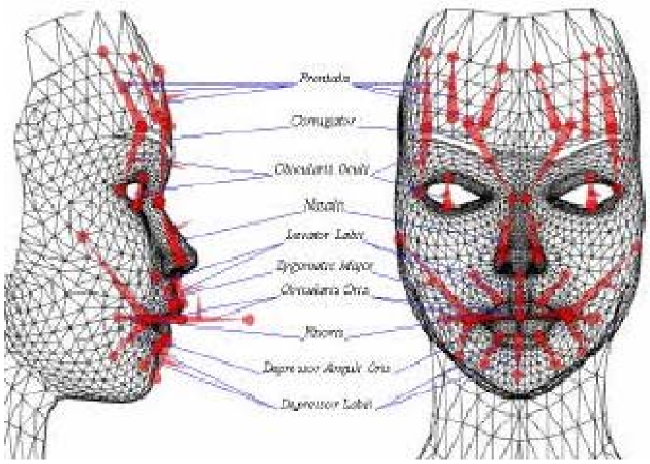
## MPEG-4 Standard

The MPEG-4 specification defines a standard for facial animation which has been implemented in a number of packages. The MPEG standard defines 68 facial action parameters (FAPs) covering movement of the face. Most of the FAPs define a rotation or translation of one or more feature points. In order to prepare a model for animation a number of Facial Definition Points (FDPs) must be selected on the model along with the region of influence on the model surface. Whilst the most common implementations of the MPEG-4 animation standard require manual selection of vertices for the FDPs it would be relatively trivial to map FDPs on a generic model to novel head models using a method such as ICP. An open source framework for working with head models and the MPEG-4 facial animation standard is the XFace toolkit (xface.itc.it).




## Muscle Vectors

Muscle vectors have an attachment point (to the bone) and an insertion point (to the skin). As the muscle is contracted nearby skin vertices are influenced more strongly along the direction vector of the muscle.



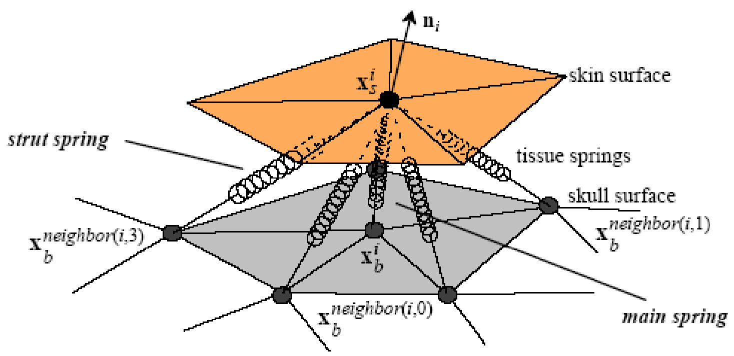
The disadvantages of the muscle vector method are as follows:

- Treats skin as a 2D surface with no concept of curvature.
- Artefacts may appear when vertices are within the influence of two muscles.

## Mass-Spring Model

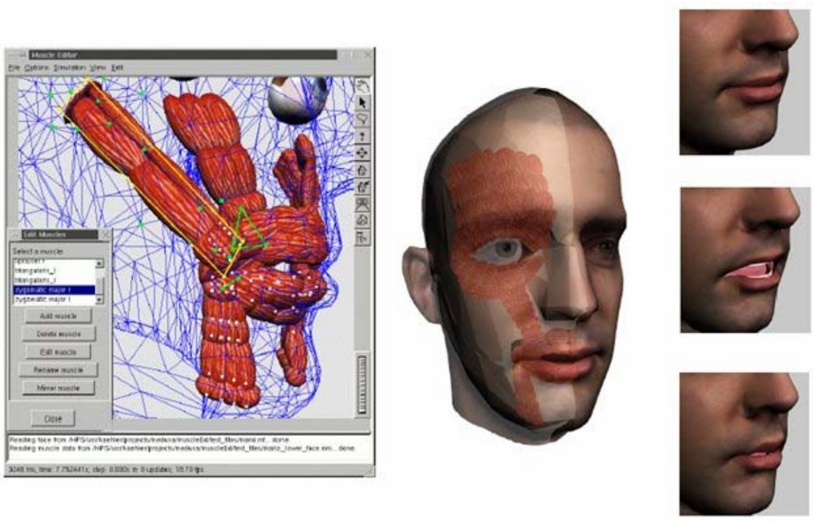
Using this method the skin, muscle and bones are modelled as a number of point masses connected by springs. A number of spring models for facial animation exist however, one of the most common is the Terzopoulos and Waters [1-2] method where the entire face is modelled as a three-layer mass-spring system. Features of the Terzopoulos and Waters method include:

- Simplifies implementation: everything is handled by a single system.
- Capable of real time expression generation.
- Provides some wrinkle effects.
- Unrealistic model of muscles and bone.
- Cannot be controlled via muscle activations.




## Kähler, et al.

In the Kähler [3] model the muscles are represented as ellipsoids, with long or curved muscles divided into piecewise linear segments. As a muscle is contracted the diameter is scaled as the length changes to implement surface bulging in a volume-preserving manner. Kähler also uses a mesh of preservation springs between the skin and the bone to prevent inter-layer penetrations.



A number of papers regarding Kähler's model along with some interesting videos can be found at the Facial Animation and Modelling Webpage (Computer Graphics Group, MPI Informatik, Saarbrücken, Germany): www.mpi-inf.mpg.de/resources/FAM/. This method provides some of the most accurate, fast and expressive facial animation to date.

### Finite-Element Models

In this case the system is broken down into a regular discreet representation (eg. tetrahedrons). This method is more accurate and more stable than the various mass-spring models, however, it is much more computationally expensive. It can take many minutes for a single frame of animation to be calculated.

## *Summary*

Finite-element models of facial animations perhaps provide the most accurate results, however, the computation time for even simple animations is too extensive to consider using where interactive results may be required. The muscle vector model is simple to implement and efficient enough to consider using for interactive work, however, the accuracy and realism of the models produced is probably not good enough for use where detailed animations are required. Furthermore anomalies in the skin surface may occur where vertices are between two muscles. Both mass spring models and Kähler's model both produce high accuracy animations in real or near-real time. In addition to being accurate and fast, these methods ability to model folds in the skin seriously adds to the realism available in the final renderings. Finally, Kähler's method is suitable for transferring muscles models between different 3D heads. This allows a single animation to be defined and then applied to any number of models. This is essential if we wish to be able to automatically apply animations to novel 3D data as it is presented to the system.

The MPEG-4 model has the advantage in ease of implementation as open source libraries are readily available for editing and publishing MPEG-4 compatible animations. Suitably accurate animations can be created given the availability of annotated high resolution 3D models. Parts of the MPEG-4 standard specifies the ability to use key frames to generate some emotions, however, it should be possible to synthesise these key frames from a generic model. The MPEG-4 standard does not appear to contain methods for mapping from high level emotional instructions to FDP movements however it should be possible to generate such mappings with a little work. With the availability of open source libraries and the potential to fully automate the FDP mapping process the MPEG-4 standard is probably a good initial starting point for generating facial animation, however, it appears that Kähler's research is more sophisticated, accurate and already possesses the ability to map muscle movements between head models.

## *References*

[1] D. Terzopoulos, Waters, K., *Physically-Based Facial Modelling*, Analysis, and Animation, The Journal of Visualization and Computer Animation, 1990

[2] K. Waters, A *muscle model for animating three-dimensional facial expressions*, SIGGRAPH'87

[3] K. Kahler, J. Haber, H.-P. Seidel, Geometry-*based muscle modeling for facial animation*, Proceedings Graphics Interface 2001