Visualization of talking human head

Martin Simunek
xsimunek@fel.cvut.cz

Department of Computer Science and Engineering
Czech Technical University
Prague / Czech Republic

Abstract

This paper describes the system for visualization of human face and its animation. The system uses the representation of real facial muscles on a model of a face. The animation is based on the deformations of muscles. These deformations are described in the medical literature. We used the advantage of defining objects in VRML (Virtual Reality Modelling Language). The application has been written in C++ language with support of OpenGL library and platform independence has been achieved by use of QtLib library.

Keywords: visualization, face, facial animation, speech, VRML, OpenGL.

1. Introduction

Modern multimedia computer systems are able to present rich and realistic information to the users. Advanced technology of graphic hardware enables common use of the 3-dimensional graphics. There are possibilities to create systems that use virtual humans and animate them in real time. Virtual human can be represented either by the whole body or only by head or face. The virtual human faces are often able to depict expression, even though these systems usually depict human speech. Such virtual humans have a large field of use. They can be part of web pages, educational programs, language training, animated agents etc. Virtual humans are unusual way to present information to the handicapped people. For example deaf people can practice their ability to read lips of other person.

The creation of a virtual human face is complex task. We have to find a proper method for data representation of the model. These methods are described in chapter 2. The next step is a creation of the facial muscles and subdividing of the model. A possible way of implementation is described in chapter 3. Chapter 4 describes an analysis of translation phonemes to visemes.

The virtual human systems are usually interconnected with voice synthesis and they create homogenous applications.

2. Model representation

Proper data representation is necessary for an easy modification of existing models and for creating of the new models of virtual humans. Model of a face could be represented in the several ways. The

developers of virtual human systems often create their own data formats (see 2.1.) or they use existing formats such as VRML (see 2.2.)

2.1 Raw data

This type of representation usually uses one or more data files. The files usually contain a set of numbers. These numbers describe vertices, polygons (usually triangles), colours and control points or vectors of the model.

A few data files are used in the application of Expression toolkit [3]. These files contain coordinates of vertices, indexes of polygon vertices and coordinates of vectors.

2.2 VRML

VRML has all necessary features to describe any model and it can serve as a good language for model description. VRML has following advantages:

- it supports the definition of structured user data types, therefore we can create model with all properties we need
- we can inspect created model in 3D environment
- we can use the large scale of defined structures and data types

All data are included in one file and it is the main advantages of this model representation. The next advantage is the exact definition of data types and structures.

This type of model representation was used in our application.

3. Representation of muscles and dividing of model

One of main problems in virtual human systems is an appropriate subdividing of the model and definition and implementation of facial muscles.

3.1 Facial muscles

Main part of virtual human systems depicting human speech is implementation of muscles around the mouth. Following pictures contains all muscles, which are important for movement of lips and the lower jawbone.

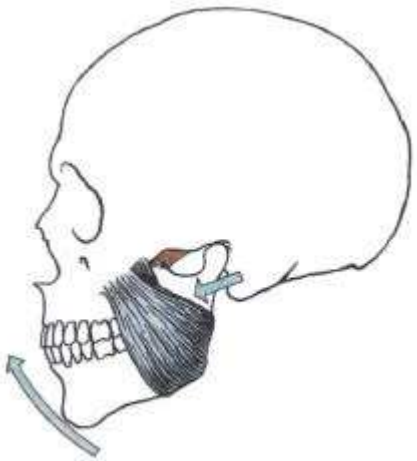

Figure 1: Musculus masseter and directions of the lower jawbone movement

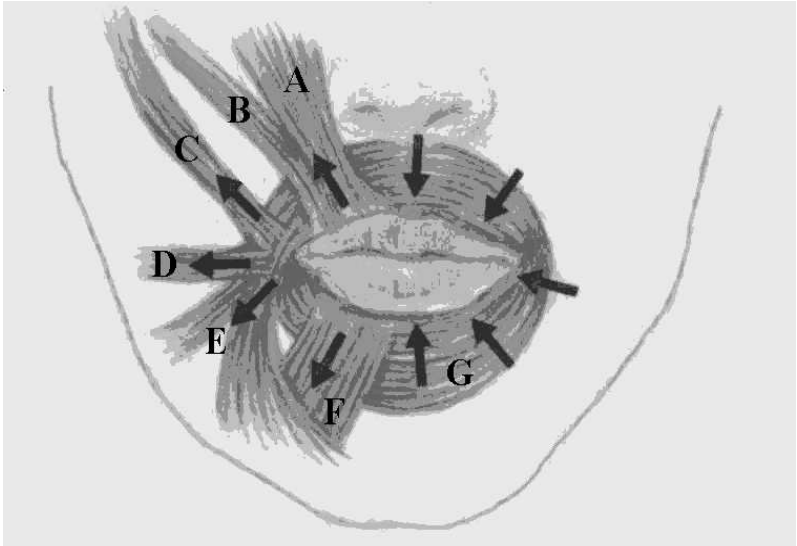

Figure 2: Facial muscles around the mouth and directions of muscle contraction (A – *m. levator labii superioris*, B – *m. zygomaticus minor,* C - *m. zygomaticus major*, D - *m. risorius*, E - *m. depressor anguli oris*, F - *m. labii inferioris*, G - *m. orbicularis oris*)

All depicted muscles must be implemented for the realistic movement of mouth. The lips are formed by *musculus orbicularis oris*. The function of this muscle is narrowing of lips.

Musculus levator labii superioris, *musculus zygomaticus minor*, *musculus zygomaticus major*, *musculus levator angulis ori* and *musculus risorius* are used for the movement of the upper lip. *Musculus levator labii superioris* and *musculus zygomaticus minor* have similar function and they

can be implemented as one muscle. Their function is lifting of the middle part of the upper lip. We can also merge the function of *musculus zygomaticus major* and *musculus levator angulis ori*. They lift the corner of mouth and they move it aside. *Musculus risorius* moves the corner of mouth aside.

Musculus depressor anguli oris, musculus depressor labii inferioris are used for the movement of the lower lip. *Musculus depressor anguli oris* moves the corner of mouth downwards and aside. *Musculus depressor labii inferioris* moves the middle part of the lower lip downwards.

Musculus masseter moves the lower jawbone upwards and downwards. More information about muscles can be found in medical literature [1, 2].

3.2 Control points

It is suitable to implement all facial muscles as vectors. Two points of the vector determine places, where the muscle is attached on the lip and on the skull. The first point is mobile and we call them control points. The second point is immovable. The movement of the muscles is implemented as extending or reducing of a distance between points of the vector. This reduction or extension performed by movement of control point. We have to determine limits of vector length. They must be defined by anatomy of a human face [1, 2].

We used 12 control points for the movement of the lips. They are depicted on following picture.

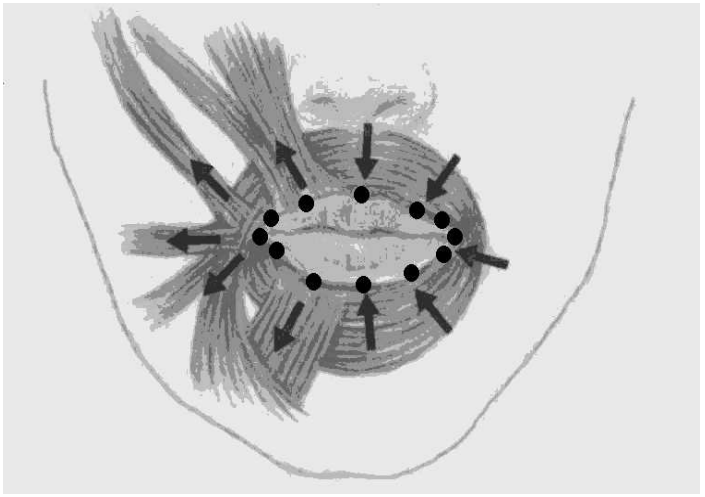

Figure 3: Control points

MPEG-4 specification [4] contains another way to define control points.

3.3 Subdividing of the model

We realised during the analysis of model, that the movement of control point deforms only a part of the polygonal mesh. For that reason it will be advantageous to divide model to parts. This dividing minimizes amount of operations performed simultaneously. For that reason we need to regenerate only areas with changes. It is suitable to use defined areas of a human head. Following picture depicts these areas.

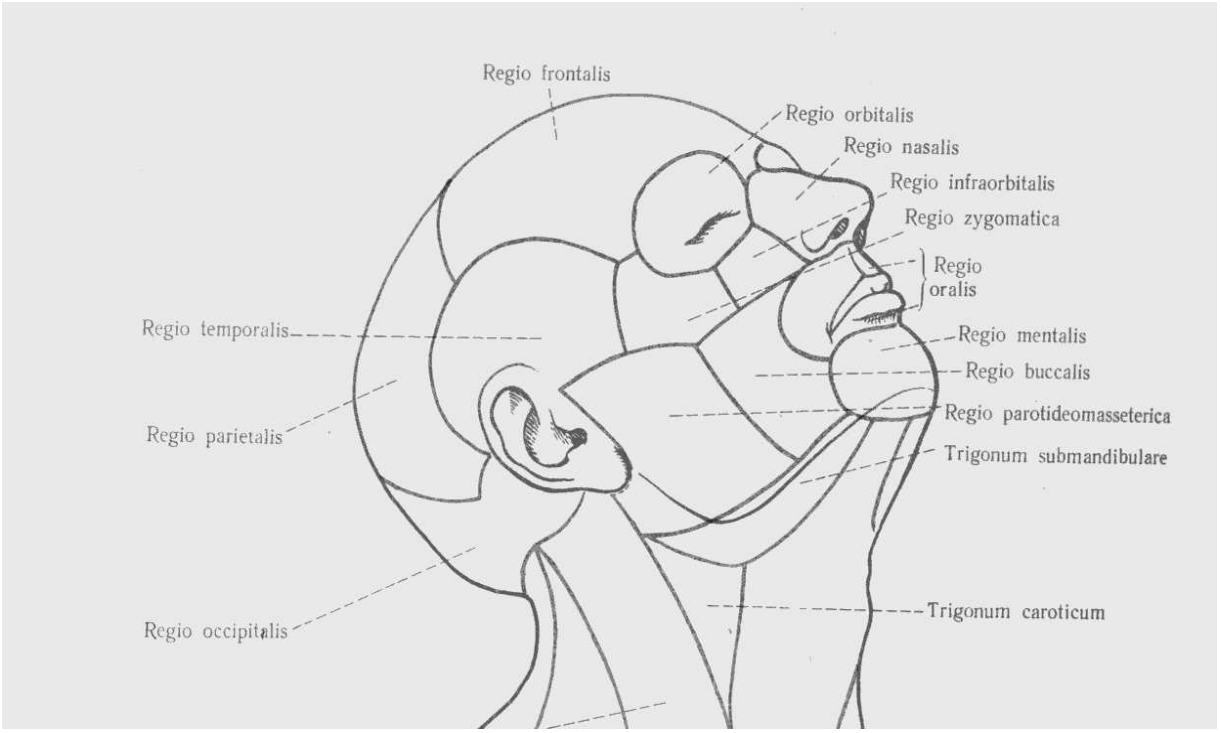

Figure 4: Areas of a human head

4. Phoneme and viseme analysis

4.1 Phonemes and visemes

Applications performing facial animation work with special terms - phoneme and viseme. The phoneme is the sound representation of the letter. Each phoneme can have a few variations. These variations depend on phonemes surrounding current phoneme. The viseme is a visual image of phoneme. It can be interpreted as an appearance of a face and lips during pronunciation of the phoneme.

The relationship between phonemes and visemes do not vary in most languages, but the relationship between letters and phonemes differ in various languages. The typical example is the pronunciation of vowels in Czech and in English language. An English letter ”e” is in some occasions pronounced as a Czech letter ”i” and in some occasions also as a letter ”e” etc.

4.2 Groups of visemes

The visemes could be divided into different groups, because some phonemes have the same viseme. A difference of pronunciation of such phonemes is caused by position of a tongue and mainly by function of vocal cords. Dividing phonemes into groups is a step to make viseme implementation easier. It is sufficient to create one viseme for the whole group of phonemes. For example one such group is formed by phonemes for ”b”, ”p”, ”m” etc. Some phonemes and visemes cannot be assigned to any group. Vowels are the typical example of such phonemes. Every vowel has unique phoneme and unique viseme, therefore every vowel needs to have a specific representation.

5. System implementation

5.1 Data import

System uses VRML model representation. The data of model are described in a prototype based on the IndexedFaceSet node. This node contains all necessary information about polygonal mesh. The prototype contains a few arrays. Two similar arrays determine the membership of each vertex or face to the groups of vertices or faces (described in section 3.3). The order of each element in the array corresponds to the order of the element (vertex or face) in IndexedFaceSet node. Next array determines control points of the model. This array contains indices to the array of vertices.

We used 130 vertices and 220 faces for description of our model.

5.2 Data structures

We used mostly two-dimensional arrays for data representation. Each row in array contains information for one vertex (or face, colour etc.). A real data are only in the arrays containing vertices and colours. A data in other arrays are indices to the another arrays. This structure provides easy access to the every information. The structure of arrays contains some duplicate information, but this information is not redundant. An array of vertices contains information about membership of the vertex in the group of vertices and an array of membership of the vertices contains information about the vertices in the group of vertices. An array of faces uses the same duplicity of information. It allows very quick access to all vertices (or faces) of some group and it also allows to learn the group containing the vertex (or face).

5.3 Subdividing of the model

This subdividing can use some of advantages of the OpenGL structures. The model contains quite large amount of graphic elements, that need to be created once and they must not be recomputed in every moment, when the picture refreshes. It is advantageous to use the OpenGL Display-list. The Display list is a structure created in memory only once (at the time of list creation). Using of Display-list decrease number of operations. The Display-list has the one major disadvantage – contents of the list cannot be modified. Only possibility to change Display-list is to move, rotate and scale list, therefore it is absolutely useless for parts of the model, that change their appearance in different ways than the moving, rotating and scaling.

The Display-list can be used for face areas without the facial muscles. It includes areas of temples, some segments of a forehead, the segments around the lower jawbone and the part of the nose.

It is necessary to dynamically regenerate the shape changing parts of the model. The areas around mimic muscles are parts of the model, that change their shape. In each step of a model refresh positions of vertices and appropriate faces are recomputed and they are redrawn.

We divide the model to 6 parts – one static part and five dynamical parts. The dynamical parts are: *regio oralis*, *regio buccalis*, *regio mentalis*, *regio paratideomasseterica* and *trigonum submandibulare* (see figure 4). They are regenerated in every step, when they change.

5.4 Control points

The deformation of polygon mesh is based on the concurrent deformation controlled by control points of the model. They are defined in input VRML file.

We use all control points defined in the chapter 3.2. In addition to them we use a special pair of points. They define joints of the lower jawbone and axis of its rotation. The area of lower jawbone (see 5.3) is rotated around this axis. The definition of movement limits of jawbone is based on information in literature [1].

Control points are implemented as vertices of polygonal mesh of the model and they deform neighbouring faces.

5.5 Text to viseme conversion

We use 12 groups of visemes and rules for Czech pronunciation. Chapter 4.2 describes how to divide phonemes and visemes into parts. The implementation of visemes (single visemes and groups of visemes) consists of two phases.

The first phase is creation of the new shape and position of lips. It consists of translation of one or more control points. These translations represent contractions of facial muscles.

The second phase consists of the rotation and translation of lower jawbone area. The results of both phases are displayed simultaneously.

6. Conclusions

The main tasks of creation of virtual human system were introduced in this paper. We described various ways of data representation of the model including its subdividing. The movement control issued from analysis of a human face anatomy. Our method of facial animation was based on movement of control points, that deformed polygonal mesh of the model..

7. Future work

This project has been implemented by one person during one semester. There are still a lot of things, that can be done for improving this application. Further improvements will include

- implementation of the function of remaining facial muscles [1,2]
- interconnection with the voice synthesis application
- implementation of the protocol for communication between instances of this application

All these improvements are the theme of my diploma thesis.

8. References

[1] Prof. MUDr. DrSc. Cihak, R.: Anatomie 1, pages 361-380, Prague 1987. [in Czech]

[2] Sinelnikov, R.D.: Atlas anatomie cloveka 1, pages 284-297, Moscow 1978. [in Czech]

[3] The Expression Toolkit, expression.sourceforge.net

[4] MPEG-4: Generic coding of audio-visual objects, ISO/IEC FDIS 14496-2