Computer Facial Animation for Sign Language Visualization

Dean Barker

Thesis presented in partial fulfilment
of the requirements for the degree of
Master of Science
at the University of Stellenbosch

Pectora roborant cultus recti

Dr. L. van Zijl
April 2005

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Stellenbosch University http://scholar.sun.ac.za

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Stellenbosch University http://scholar.sun.ac.za

Abstract

Sign Language is a fully-fledged natural language possessing its own syntax and grammar; a fact which implies that the problem of machine translation from a spoken source language to Sign Language is at least as difficult as machine translation between two spoken languages. Sign Language, however, is communicated in a modality fundamentally different from all spoken languages. Machine translation to Sign Language is therefore burdened not only by a mapping from one syntax and grammar to another, but also, by a non-trivial transformation from one communicational modality to another.

With regards to the computer visualization of Sign Language; what is required is a three dimensional, temporally accurate, visualization of signs including both the manual and non-manual components which can be viewed from arbitrary perspectives making accurate understanding and imitation more feasible. Moreover, given that facial expressions and movements represent a fundamental basis for the majority of non-manual signs, any system concerned with the accurate visualization of Sign Language must rely heavily on a facial animation component capable of representing a well-defined set of emotional expressions as well as a set of arbitrary facial movements.

This thesis investigates the development of such a computer facial animation system. We address the problem of delivering coordinated, temporally constrained, facial animation sequences in an online environment using VRML. Furthermore, we investigate the animation, using a muscle model process, of arbitrary three-dimensional facial models consisting of multiple aligned NURBS surfaces of varying refinement.

Our results showed that this approach is capable of representing and manipulating high fidelity three-dimensional facial models in such a manner that localized distortions of the models result in the recognizable and realistic display of human facial expressions and that these facial expressions can be displayed in a coordinated, synchronous manner.

Stellenbosch University http://scholar.sun.ac.za

Opsomming

Gebaretaal is 'n volwaardige natuurlike taal wat oor sy eie sintaks en grammatika beskik. Hierdie feit impliseer dat die probleem rakende masjienvertaling vanuit 'n gesproke taal na Gebaretaal net so moeilik is as masjienvertaling tussen twee gesproke tale. Gebaretaal word egter in 'n modaliteit gekommunikeer wat in wese van alle gesproke tale verskil. Masjienvertaling in Gebaretaal word daarom nie net belas deur 'n afbeelding van een sintaks en grammatika op 'n ander nie, maar ook deur beduidende omvorming van een kommunikasiemodaliteit na 'n ander.

Wat die gerekenariseerde visualisering van Gebaretaal betref, vereis dit 'n driedimensionele, tyds-akkurate visualisering van gebare, insluitend komponente wat met en sonder die gebruik van die hande uitgevoer word, en wat vanuit arbitrêre perspektiewe beskou kan word ten einde die uitvoerbaarheid van akkurate begrip en nabootsing te verhoog. Aangesien gesigsuitdrukkings en -bewegings die fundamentele grondslag van die meeste gebare wat nie met die hand gemaak word nie, verteenwoordig, moet enige stelsel wat te make het met die akkurate visualisering van Gebaretaal boonop sterk steun op 'n gesigsanimasiekomponent wat daartoe in staat is om 'n goed gedefinieerde stel emosionele uitdrukkings sowel as 'n stel arbitrre gesigbewegings voor te stel.

Hierdie tesis ondersoek die ontwikkeling van so 'n gerekenariseerde gesigsanimasiestelsel. Die probleem rakende die lewering van gekordineerde, tydsbegrensde gesigsanimasiesekwensies in 'n intydse omgewing, wat gebruik maak van VRML, word aangeroer. Voorts word ondersoek ingestel na die animasie (hier word van 'n spiermodelproses gebruik gemaak) van arbitrre driedimensionele gesigsmodelle bestaande uit veelvoudige, opgestelde NURBS-oppervlakke waarvan die verfyning wissel.

Die resultate toon dat hierdie benadering daartoe in staat is om hoë kwaliteit driedimensionele gesigsmodelle só voor te stel en te manipuleer dat gelokaliseerde vervormings van die modelle die herkenbare en realistiese tentoonstelling van menslike gesigsuitdrukkings tot gevolg

Stellenbosch University http://scholar.sun.ac.za

het en dat hierdie gesigsuitdrukkings op 'n gekordineerde, sinchroniese wyse uitgebeeld kan word.

Stellenbosch University http://scholar.sun.ac.za

Acknowledgments

As with all efforts of this magnitude, many individuals need to be thanked for various contributions; in particular, Dr. Lynette van Zijl, my supervisor, for her mentorship, motivation and unrelenting patience throughout the tenure of my postgraduate study; my parents for their continual support (both emotional and financial) and Imelda Carstens, my girl-friend, who has shown me more understanding and love than any man deserves.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Stellenbosch University http://scholar.sun.ac.za

Contents

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

List of Figures

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

List of Tables

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 1

Introduction

> *Grand challenges like space exploration and weather prediction are expanding human frontiers, but the grandest challenge is the exploration of how we as human beings react to the world and interact with each other. Faces are accessible "windows" into the mechanisms which govern our emotional and social lives. The technological means are now in hand to develop automated systems for monitoring facial expressions and animating artificial models. Face technology of the sort we describe, which is now feasible and achievable within a relatively short time frame, could revolutionize fields as diverse as medicine, law, communications, and education.*
>
> – PAUL EKMAN & TERRENCE J. SEJNOWSKI [7]
>
> from the Executive Summary of the *Report To NSF of the Planning Workshop on Facial Expression*, 1992.

1.1 Motivation

In 2002, at the University of Stellenbosch, a project was initiated to develop the South African Sign Language Machine Translation System (SASL-MT System); a machine translation system from English to South African Sign Language (SASL). The need for such a system is high [42] given that in the absence of Sign Language interpreters, communication for the Deaf community in places such as clinics, hospitals, post offices and police stations becomes difficult. Unfortunately, human interpreters are prohibitively expensive and there will no doubt always be a shortage of appropriately skilled people. A computer translation system would enable mass translation for more situations in an affordable manner.

Stellenbosch University http://scholar.sun.ac.za

Sign Language is a fully-fledged natural language [53] possessing its own syntax and grammar; a fact which implies that the problem of machine translation from a spoken source language to Sign Language is at least as difficult as machine translation between two spoken languages. Sign Language, however, is communicated in a modality fundamentally different from all spoken languages. Machine translation to Sign Language is therefore burdened not only by a mapping from one syntax and grammar to another, but also, by a non-trivial transformation from one communicational modality to another.

Sign Language cannot be spoken and there exists no satisfactory method of writing Sign Language. The most popular textual representation of Sign Language is known as *glossy* notation in which signs are represented in their natural order by upper-case words taken from their nearest spoken counterparts. This approach is severely limited by its failure to represent the actual physical sign. As an attempted solution to this problem, iconic or pictorial representations have often been used [9]. These are however not ideal as they fail to capture spatial and temporal elements associated with the generation of signs [53]. What is required is a three dimensional, temporally accurate, visualization of signs including both the manual and non-manual components which can be viewed from arbitrary perspectives making accurate understanding and imitation more feasible.

1.2 The Significance of Facial Expressions in Sign Language

There exists a common misconception that Sign Language is communicated exclusively via motions of the hands and arms (*manual signs*); this could not be further from the truth. Whilst manual signs form the primary lexical entities of Sign Language, it has been shown [2] that so called *non-manual signs* (NMS) play significant lexical, morphological and syntactic functions. The significance of NMS is such that any system hoping to accurately synthesize Sign Language must be able to synthesize both the manual and non-manual components of signs.

From an engineering perspective, human communication can be viewed as an exchange of sequences of meaningful tokens or signs. The medium upon which the sign is carried is often referred to as the *sign vehicle*; hence it can then be said that human communication occurs over multiple synchronized channels i.e. via multiple sign vehicles. Specifically, BAKER and PADDEN [2] identified five sign vehicles used in Sign Language communication:

1. hands and arms;

Stellenbosch University http://scholar.sun.ac.za

2. head nods, shakes and tilts;
3. facial expressions;
4. eye gaze and blinks;
5. body posture, torso twisting or body shifting.

(2), (3), (4) and (5) constitute the aforementioned NMS. Functionally, NMS can be defined as "*those facial expressions, head movements, eye gazes and body shifts that are used to express lexical, morphological and syntactic meanings in natural Sign Languages, or that are functioning as discourse markers regulating conversation* [16]." It has been stated [16] that American Sign Language (ASL) is an incomplete language without non-manual signs. This will become apparent once the particular roles of NMS in ASL, specifically (2), (3) and (4) above, are made clear in the remainder of this section.

Unfortunately, most of the studies done into the significance of NMS have been in the context of ASL and almost no linguistic research on SASL has been published. For this reason the SASL-MT project has decided to standardize on ASL wherever specific linguistic information on SASL is not available; this includes the question of the significance of NMS in SASL.

1.2.1 Lexical NMS

Lexical NMS are those found within a single sign or produced alone without any associated sign. One particular role of lexical NMS is to distinguish what are known as *minimal pairs*. These are two signs which are identical except for the NMS which accompany them. For example, the ASL signs for NOT-YET and LATE are identical, but NOT-YET is accompanied by a protruding tongue and a head shake. Also, the ASL signs for UNEXPECTEDLY and WRONG are identical, but UNEXPECTEDLY is accompanied by raised eyebrows, widened eyes and a lowered jaw. Similarly in SASL the signs for WEIGHT and MAYBE represent a minimal pair. In these cases it is obvious that the NMS are essential for the correct understanding of the sign.

The lexical significance of NMS are further evidenced by certain NMS which have no accompanying manual sign [2]. These NMS are distinct lexical items in the ASL vocabulary. Examples are MENSTRUAL-PERIOD which is represented by puffed cheeks; OH-WHY-NOT? represented by a head tilt to the side and puffed and then released cheeks; OH-DEFINITELY-YES! represented by a head tilt to the side, closed eyes and an 'O'-shaped mouth; as well as

Stellenbosch University http://scholar.sun.ac.za

THAT'S-REALLY-INTERESTING represented by an open mouth, then tightly closed lips, half closed eyelids and a nodding head.

1.2.2 Morphological NMS

Morphological NMS are those that add extra meaning to other signs or sign phrases. Modifiers are a class of NMS which generally have a morphological function and most often consist only of mouth configurations. One example is lateral tongue flapping which carries two meanings, depending on the context: *intense enthusiasm and excitement* or *extensive time and distance.* Another is the NMS called 'mm' which involves keeping the lips together and pushing them out without puckering, and means *with relaxation and enjoyment* or *normal and proper* depending on the context. There is also the NMS 'th' produced by pushing the lips out with a protruding tongue into a 'th' configuration and tilting the head. This NMS means *lack of control, inattention* or *unawareness.*

1.2.3 Syntactical NMS

Syntactical NMS are those that are required to produce different sentence forms such as assertions and questions, and most often involve the upper portions of the face. The best example accompanies what are known as wh-questions (who, what, where, when, why and how). In such questions, the eyebrows are drawn together and down, the head is often tilted and the eye gaze is directed at the receiver. Negation has similar NMS; the eyebrows are drawn together and down, the nose is wrinkled, the upper lip raised and the head is shaken from side to side.

Other types of questions are usually accompanied by a raising of the eyebrows. For example yes/no questions are accompanied by raised eyebrows, widened eyes, a forward head tilt and eye gaze directed at the receiver whilst rhetorical questions have raised eyebrows, a head tilt and eye gaze directed at the receiver.

1.2.4 NMS as Discourse Markers

Discourse markers are those actions performed by the receiver used to indicate that the receiver is understanding the signer. Examples of spoken discourse markers are words such as 'uh-huh', 'hmmm', 'really?' or 'yeah?'. NMS used as discourse markers are head nods, head tilts, brow raises and nose twitches.

Stellenbosch University http://scholar.sun.ac.za

1.3 Machine Translation from Spoken to Signed Language

The above discussion relating the importance of NMS in Sign Language shows that any system concerned with the accurate representation of Sign Language must possess the ability to represent both manual and non-manual signs; in particular, facial expressions and movements. It can therefore safely be said that computer facial animation is an essential component of a Sign Language visualization system. The value of such a visualization system in the framework of a machine translation system from a spoken source language to a Sign Language has already been pointed out in section 1.1.

Given that the SASL-MT System [42] is just such a machine translation system it follows that it too must rely heavily on a Sign Language visualization component, which in turn relies heavily on a facial animation component. Given this hierarchy, it is evident that the design of the facial animation component should be determined by the role it plays in the SASL-MT System. The remainder of this section will be devoted to an exposition of the requirements and constraints placed on the facial animation component of the SASL-MT System.

1.3.1 Design Constraints for the SASL-MT Facial Animation System

Two major classes of design constraints can be identified: those placed on the system by the fact that it is a component of the SASL-MT System, and those placed on the system by the fact that it is a component of a Sign Language visualization system.

1. SL Visualization Constraints
    (a) The facial animation component must be able to manipulate a three-dimensional model of a human face in such a manner that localized distortions of the model result in the *recognizable* and *realistic* display of human facial expressions. In this manner the system must be capable of representing a well defined finite set of human *emotions*.
2. SASL-MT Constraints
    (a) The facial animation component must satisfy certain *soft* real-time deadlines due to its role in the SASL-MT System (facial expressions and other body movements must be displayed in a coordinated, synchronous manner).

Stellenbosch University http://scholar.sun.ac.za

(b) In an effort to make the system widely applicable, the facial animation component must be capable of operating on arbitrary NURBS based three-dimensional facial models.

The following two chapters will be devoted to documenting existing approaches which have satisfied some of the constraints introduced above.

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 2

SL Visualization Design Constraints

In section 1.3.1 the primary constraint placed on the design of the SASL-MT Facial Animation System due to its role as a SL visualization system was introduced as being the requirement that the system must be able to manipulate a three-dimensional model of a human face in such a manner that localized distortions of the model result in the *recognizable* and *realistic* display of human facial expressions. In this manner the system must be capable of representing a well defined finite set of human *emotions*. This constraint gives rise to a set of subsidiary constraints which will be discussed below.

Firstly, it indicates that the system operates on a three-dimensional model of a human face. Before considering the nature of this operation, one must therefore consider the nature of the three-dimensional model itself. Secondly, it indicates that the system is capable of generating facial expressions which represent human emotions. This necessitates the quantifiable categorization of human emotive facial expressions. In the following sections of this chapter, these two constraints will be investigated further.

2.1 Three-dimensional Facial Models

WANG [45] defined the *surface model* of a character animation system as "*the method by which the surface of the object is represented*" i.e. the geometric entities constituting the representation of the object. She in turn defined the *structural model* as "*the collection of data representing the physical shape of the object.*" In other words, the structural model of a character animation system can be viewed as an instance of the surface model of that system.

The choice of what geometric representation will be used as the basis for a three-dimensional facial model is critical. This choice significantly influences the techniques used to actuate and render the facial model and hence also the design of the rest of the facial animation system. The criticality of this decision is further compounded by the natural complexity of the human

Stellenbosch University http://scholar.sun.ac.za

face, and the fact that the surface model must be capable of incorporating and representing this complexity. A multitude of approaches exist, the most common of which will be introduced below.

2.1.1 Surface Models

Polygonal Surfaces

By far the most popular surface model [19, 20, 25, 26, 30, 33, 35, 37, 38, 48] used in existing facial animation systems is based on a polygonal surface description, which is generally realized using triangles due to their planarity. The popularity of polygonal surface models stems primarily from the fact that polygons are the only geometric primitives supported by most modern graphics boards. Furthermore, algorithms operating on polygons (for example shading and hidden surface removal) are generally quite simple due to the simple geometric nature of the polygon.

The most significant draw-back of a polygon surface model is its inability to represent a smoothly curved surface. The discrete, planar nature of the polygon implies that it is fundamentally incapable of representing a continuous curved surface. When one considers the nature of the human facial surface, it becomes evident that the use of a polygonal surface model to represent a human face poses a problem.

It should be noted however that whilst polygons cannot truly represent curved surfaces, acceptable approximations can be achieved by using large numbers of small polygons in areas of high surface curvature combined with smooth shading techniques such as Gouraud shading. PARKE and WATERS [27] defined a series of considerations to be noted when developing a realistic polygonal facial model.

Parametric Surfaces

Because of the inability of polygonal surface models to represent continuous curved surfaces, surfaces defined by a set of parametric functions have become a popular alternative. Of the parametric surfaces, the spline surfaces have received the most attention. These include Bézier splines, B-Splines, Catmull-Rom splines and Non-Uniform Rational B-Splines (NURBS) [3]. All of these *tensor-product* surfaces are defined by a set of control values and basis functions.

The primary advantages associated with the use of parametric surfaces are increased flexibility and smoothness as well as a significant reduction in data storage requirements. Further-

Stellenbosch University http://scholar.sun.ac.za

more, the fact that parametric surfaces have a well-defined mathematical basis facilitates their widespread implementation.

The single greatest disadvantage is the lack of hardware support for the rendering of parametric surfaces. Generally, at runtime, the parametric surface representation is converted to a polygonal surface representation for rendering. This implies that rendering partially becomes a software process, a fact which significantly influences performance.

Implicit Surfaces

An interesting, alternative surface representation is based on the use of implicit surfaces [51]. Mathematically, these surfaces are defined by a membership function which divides space into two sets, those which belong to the surface, and those which do not. This definition implies that multiple implicit surfaces can be blended together to form new objects. The blending technique allows implicit surfaces to become powerful, and unique modeling objects.

Practically, implicit surface representations as modeling tools have been realized as *soft objects* and *metaballs*. The greatest disadvantage associated with the use of implicit surfaces is related to the fact that implicit surface modeling is still a relatively new method, and more efficient algorithms for rendering and real-time manipulation are required.

Volume Representations

The two most popular volumetric representations are constructive solid geometry (CSG), and volume element (voxel) arrays [49]. CSG has been widely used in engineering and CAD applications and uses set operations to describe objects as aggregations of simple primitive objects. Volume element array representation has been widely used in medical imaging applications, and describes three-dimensional objects as composites of multiple arrays of parallelpiped volume elements or voxels. To the author's knowledge, neither of the approaches have been successfully applied to facial animation as yet.

2.1.2 Summary

The above discussion highlighted the most popular surface models used in facial animation systems to date. By far the most popular is the polygonal model. This popularity is directly related to the fact that polygonal surfaces have formed the basis of modern three-dimensional computer

Stellenbosch University http://scholar.sun.ac.za

graphics techniques and algorithms. Parametric models were introduced as a major competitor to polygonal models since parametric models provide increased flexibility and smoothness as well as a well-defined mathematical basis for the representation and manipulation of surfaces.

2.1.3 Structural Models

The remainder of this chapter will be devoted to an exposition of some of the structural models used in existing facial animation systems. A large number of facial animation systems have been developed in the past three decades and many of these have followed similar approaches in terms of surface and structural models. Those discussed below represent the most common approaches and will be grouped according to the surface model on which they are based.

Polygonal Structural Models

FASCIA – PARKE

In 1974 PARKE [25, 26], developed the first facial animation system based on a polygonal surface model. The model consists of several separate polygonal surfaces: the facial mask, eye-balls and teeth. Of primary importance is the facial mask, which consists of an arbitrary (in contrast to a regular) network of polygons. The topology of this polygonal network was carefully designed, in a trial-and-error fashion, to provide detail and curvature at appropriate locations on the face.

The age of the FASCIA system, and the original data acquisition techniques used to build the facial model, have meant that it is of a relatively low level of detail, that is to say, it has a low polygon count. This has meant that the system suffers from several common problems associated with polygonal modeling which detract significantly from the visual believability of the facial model. Smooth shading techniques are used to give the appearance of a curved surface, but in profile, the polygonization of the surface is obvious. Also, because of the low polygon count, the average size of each polygon is relatively large. This leads to unrealistic motion, since the only moveable parts of the face are the polygon edges and vertices and hence there are large seemingly static patches on the face during motion. See figure 2.1 on page 17.

PLATT, BADLER & PELACHAUD

In the early eighties, PLATT and BADLER [32, 33, 34, 35] developed a facial animation system which presented the face as a hierarchically structured, regionally defined object. This system

Stellenbosch University http://scholar.sun.ac.za

also animated a polygonal facial mask.

The hierarchical nature of the system included its underlying structural model. The structural model consists of a point database which defines the basic shape of the skin surface. The surface points are connected by arcs which have both geometric (the edges of the polygons constituted by the points and arcs) and semantic information (relating to the elasticity of the area between the two points connected by the arc).

A relatively low level of detail facial model was used; in the initial incarnations [32, 34] a 400 point model was used, but this was later extended to 1200 points [35]. This low level of detail led to many of the same problems associated with polygonization introduced above. The deformation of the facial mesh also suffered from wrinkling problems, specifically around the mouth, due to the physical tissue model used. See figure 2.2 on page 17.

LEE, TERZOPOLOUS & WATERS

In 1990 TERZOPOLOUS and WATERS [37] developed a system which operated on a polygonal structural model. The output from this system was significantly better than any previous attempts at facial animation and included effects such as realistic wrinkling and bulging of the skin.

The structural model of this system was significantly modified in 1991 [38, 48] to enable the use of a radial laser scanner for data acquisition. An adaptive meshing algorithm was applied to the densely sampled scanned data to create a polygonal facial mesh. This mesh was then used to create an accurate, texture mapped model of an actual person. The results were impressive since this was the first coupling of detailed, texture mapped facial models to physically-based animation techniques.

This approach did however still have certain significant flaws which detracted severely from the visual believability of the models, the most significant of which was the fact that the eyes and teeth were not distinct geometric entities. They were in fact simply part of the texture mapped onto the facial geometry. This meant that the eyes and teeth deformed along with the facial tissue. Another problem was related to the inability of laser scanners to produce accurate mappings of the entire human head which resulted in severe distortions of the output generally occurring around the hair, ears and nose.

In 1993, and 1995, LEE, TERZOPOLOUS and WATERS [19, 20] revisited their earlier system,

Stellenbosch University http://scholar.sun.ac.za

and further extended the structural model. In these incarnations, an automatic conformation algorithm was used to adapt a generic polygonal mesh of predetermined topological structure (based on that used by TERZOPOLOUS and WATERS [37] in 1990) to the laser scanned data. This allowed the process of acquiring the data model to become highly automated. Earlier problems were also solved by making the eyeballs, eyelids, teeth, neck, hair and bust of the facial model distinct geometric objects. The process of adapting a complete, generic facial model to the scanned data also solved the problem of corrupt or distorted output from the scanner. See figure 2.4 on page 18.

PIGHIN, SALESIN, SZELESKI, LISCHINSKI, HECKER & AUSLANDER

The drive to create a photorealistic facial animation system lead to the development of an image-based three-dimensional morphing approach by PIGHIN *et al.* [29, 30, 31]. This system merged multiple existing techniques in the field of facial animation to produce results of unprecedented visual quality.

The system uses several uncalibrated photographs of an individual and a photogrammetric algorithm to deform a polygonal, generic facial mesh into a three-dimensional model of that individual. Such models are created for multiple facial expressions or facial poses. Facial animation is then achieved by three-dimensional morphing between key-poses. The fact that all of the key-poses are based on the same generic facial mesh of constant topology implies that the usual problem of correspondence between key-poses is avoided.

It is unclear whether or not the structural model contains distinct polygonal surfaces for the teeth and eyes, but what is clear is that it uses distinct texture maps for the face, hair, eyes and teeth. The visual quality of static facial models achieved with this system are approaching photorealistic, however, certain problems relating to the structural model do exist. The fact that the model data is acquired from multiple photographs means that the textures extracted from these photographs reflect the lighting conditions under which they were taken. Furthermore, artifacts due to misregistrations between different photographs may occasionally occur in the extracted textures. See figure 2.3 on page 17.

Stellenbosch University http://scholar.sun.ac.za

Parametric Structural Models

HOCH, FLEISCHMANN & GIROD

In 1994 HOCH, FLEISCHMANN and GIROD [13] developed a facial animation system capable of automatically adapting a canonical B-spline facial model to data obtained from a three-dimensional laser scanner. The adapted model was then animated using a parametrized actuation model.

The structural model of this system was based on that introduced in 1989 by WAITE [44] which consisted of a rather rudimentary B-spline surface defined by a 12×16 control point grid. This was extended to a 13×16 control point grid to improve the level of detail around the mouth. The spline surface was then texture-mapped to enhance the visual realism.

The results produced by this system serve as excellent examples of both positive and negative side-effects encountered when using a parametric structural model. Compared to a polygonal structural model, the results achieved in terms of surface continuity, using even the simplest parametric surfaces, are staggering. This advantage is further magnified when considering the data storage requirements that would be needed to achieve similar results using polygons.

On the other hand, a noteworthy disadvantage of parametric surfaces is encountered when the need for local refinement of the data model arises. In terms of a facial mask, areas requiring local refinement are the mouth, nostrils, eyelids and ears. The nature of parametric surface necessitates that local refinement of the control point grid can only be achieved by a global refinement i.e. the introduction of a new row or column of control points. This problem is a significant obstacle to the creation of a non-uniformly curved parametric surface (a prime example of which is the human face).

Two general approaches have been developed to overcome this problem. The first introduces the notion of a hierarchical parametric surface in which local refinement is achieved by adding new parametric surface patches of successively greater refinement within specified regions of a surface. An alternate approach is to align multiple distinct parametric surfaces of varying levels of refinement in such a manner that they appear to form a single continuous surface. The first approach will be described in more detail in the following section. The second approach will be discussed in detail in section 4.3.

Stellenbosch University http://scholar.sun.ac.za

Langwidere – WANG

In 1994 WANG [45] developed the facial animation system known as Langwidere. The structural model of this system was based on hierarchical B-splines. This model enables the local refinement of a B-spline surface by the recursive addition of patches of greater refinement known as overlays. These patches are produced by midpoint refinement of a base surface and result in a surface representation with multiple levels of resolution.

This approach achieves local refinement whilst maintaining a relatively economical description of the surface. Rendering efficiency is also not degraded since the overlayed surface can still be rendered as a standard spline surface. The most significant drawback of using hierarchical B-splines is that no standard tools exist for the creation and modeling of such surfaces. Whilst the results produced by the Langwidere system were theoretically quite appealing, the actual models used were not visually appealing. This situation could however be improved, given suitable modeling tools and an experienced 3D modeller to create a realistic facial model.

2.1.4 Summary

The above discussion has served to highlight the differences between the most popular structural models. Whilst the use of polygons generally results in high performance rendering, this advantage must be considered in the light of a series of disadvantages. Representing a human face using polygons implies the use of discrete, planar objects to represent a continuous non-planar surface. The use of smooth-shading algorithms can, to a certain extent, hide this from the viewer, but this is only a partial solution since the polygonal nature of the model is revealed in the profile and borders of the face.

During animation, the coarseness of the polygonal mesh is evidenced by the fact that the only moveable structures (polygon vertices and edges) do not fully capture the structure of a human face. Increasing the visual believability of the model requires significant model refinement which results in models consisting of hundreds of thousands of polygons. Whilst this in itself does not represent a major problem (modern graphics hardware can render millions of polygons per second) the actuation and manipulation of such a large number of vertices significantly degrades performance.

For parametric models, the opposite is true. The sparse nature of the control point set needed to define such a surface implies that the actuation and manipulation of such surfaces is

Stellenbosch University http://scholar.sun.ac.za

fairly inexpensive. Rendering of these surface is on the other hand, expensive since the lack of hardware support requires that they be mapped to polygonal surfaces of arbitrary refinement before being rendered. Whilst the problem of local refinement of parametric surfaces certainly is noteworthy, solutions do exist which to a large extent negate the problem.

2.2 Expression Models

The opening paragraph of this chapter stated that the quantifiable categorization of human emotive facial expressions was a necessity. The mechanism facilitating such a categorization will be referred to as the *expression model* and will be investigated in the remainder of this section. An expression model is almost wholly defined by an underlying expression specification mechanism, and as such, discussion of an expression model is generally in terms of its underlying expression specification mechanism.

Over the years, several expression models have been developed; the two most noteworthy being the Mimic language developed by HJORTSJO [12] and the Facial Action Coding System (FACS) developed by EKMAN and FRIESEN [6].

The Mimic language was developed to describe facial expressions and gestures. The language defines a set of abstract muscles which can be seen as forming the alphabet of a language. By combining letters of the alphabet (i.e. muscle actions), words can be formed. These words then correspond to facial expressions.

Whilst both the Mimic language and FACS were developed with similar goals and objectives, the abstract nature of the Mimic language, and the fact that it takes a more qualitative approach to the representation of facial expressions has meant that it has found little application in the realm of computer facial animation. FACS on the other hand has been readily adopted for use in facial animation for precisely the opposite reasons and despite the fact that it was not designed with facial animation in mind. The following section investigates the Facial Action Coding System in more detail.

2.2.1 Facial Action Coding System

The Facial Action Coding System (FACS) has formed the basis for the expression model used in the majority of existing facial animation systems. The FACS is a "*method of describing facial movement based on an anatomical analysis of facial action*"[6] developed by EKMAN and

Stellenbosch University http://scholar.sun.ac.za

Friesen in the mid-to-late seventies.

The motivation for the development of the FACS was the need to be able to distinguish all possible facial movements. This need was born within the field of psychology which was, and still is, trying to answer questions relating to the significance of facial movements and expressions such as: which facial movements signal emotion? Are certain facial expression universal amongst different cultures, and which facial movements have significance during human conversation?

The FACS is an abstraction of the facial anatomy which has as its basis the set of minimal perceivable facial movements, each of which is represented by an *Action Unit* (AU). The FACS defines fifty-nine action units which constitute all perceivable movements of the human face. Since facial movements are not necessarily the result of the contraction of a single facial muscle some AU's represent the action of multiple muscles. An AU is defined by a unique number and a name and is associated with several muscles, the activations of which result in the action represented by the AU.

An example is AU4, the *Brow Lowerer*, which has the Depressor Glabellae, Depressor Supercilli and Corrugater muscles as its muscular basis[1]. By describing facial expressions or movements in terms of AU's a quantitative representation is achieved.

2.2.2 Summary

The FACS has proven to be a successful basis for the expression model of a facial animation system [4, 13, 28, 34, 45, 47]. Alternative approaches such as the *Minimal Perceptible Actions* (MPA) of Kalra [17] have been developed but these have been similar to the FACS. For this reason, the FACS has been chosen to form the basis of the expression model in the SASL-MT facial animation system.

---

[1] The full list of the FACS action units can be found in Appendix D on page 84.

Stellenbosch University http://scholar.sun.ac.za

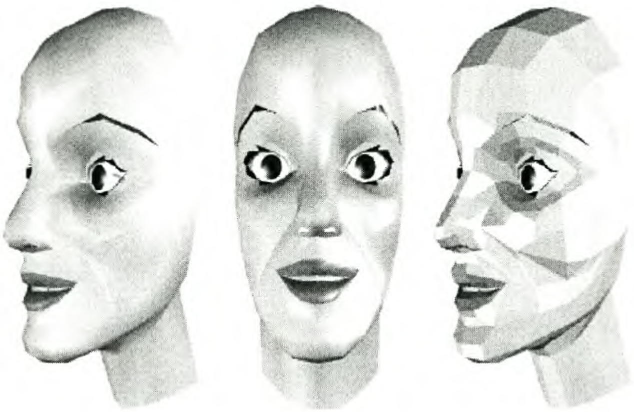
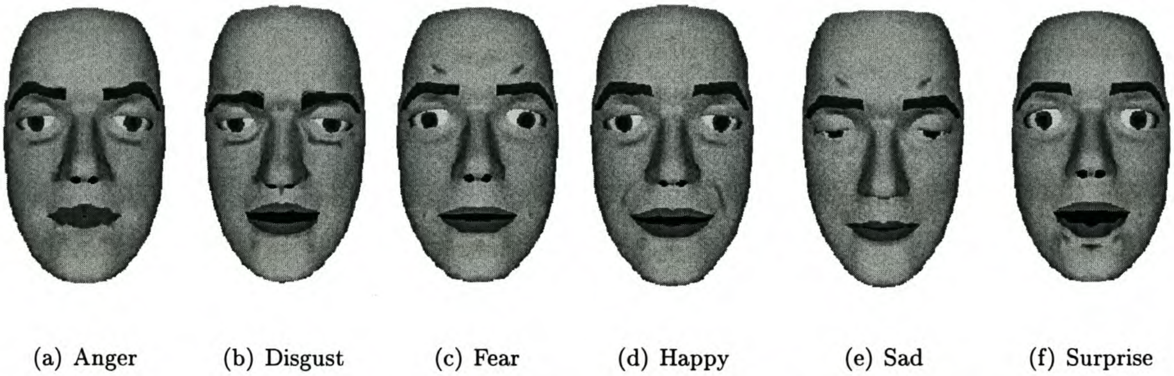


Figure 2.1: Example of the Fascia facial model, from [45].



(a) Anger (b) Disgust (c) Fear (d) Happy (e) Sad (f) Surprise

Figure 2.2: Example facial expressions generated by the system created by Platt & Badler. Images from [35].

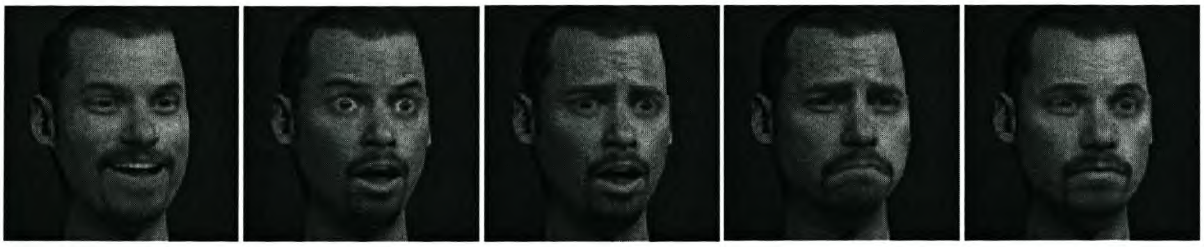


Figure 2.3: Example facial expressions generated by the system created by Pighin *et al.* Images from [30].

Stellenbosch University http://scholar.sun.ac.za

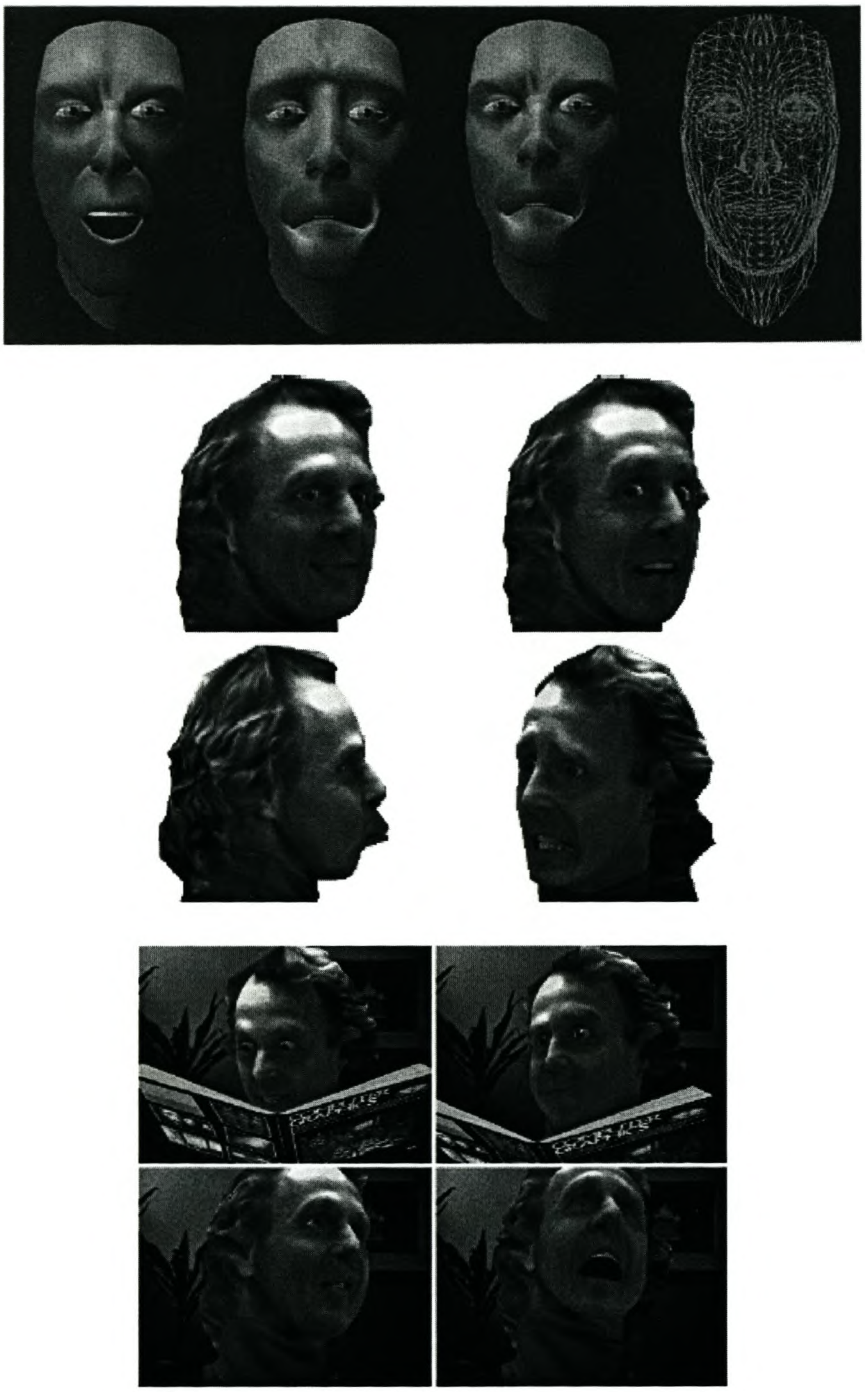


Figure 2.4: Example facial expressions generated by the systems created by Lee, Terzopolous and Waters. Images from [20, 27]

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 3

SASL-MT Design Constraints

In section 1.3.1 the primary constraint placed on the design of the SASL-MT Facial Animation System due to its role as a component of the SASL-MT system was introduced as being the need to satisfy a set of soft real-time deadlines imposed by the SASL-MT system. To elaborate: it is necessary that during the visualization of a certain SL sentence, the manual and non-manual signs which constitute that sentence be generated in a coordinated and synchronous manner. This is due to the fact that manual signs are interpreted in the context of non-manual signs and vice versa (see section 1.2). This constraint gives rise to a set of subsidiary constraints which will be discussed below.

Firstly, it implies that the facial animation system must use a method of model actuation which will enable the generation of facial animation sequences in interactive time. That is to say; the actual three-dimensional facial model must be manipulated at interactive rates and hence the algorithms used to achieve this manipulation must operate at equivalent rates.

Secondly, given that explicit temporal demands are placed on the facial animation sequences, a mechanism allowing explicit and detailed control of these sequences is required. In the following sections of this chapter, these two constraints will be investigated further.

3.1 Actuation Models

WANG [45] defined the *actuation model* of a character animation system as "*a parametrization of the structural model with inherent manipulation methods for the purposes of simulating motion.*" In the case of the SASL-MT Facial Animation System, what is required is an actuation model which will deliver satisfactory results in terms of visual realism and believability whilst remaining computationally inexpensive.

Several diverse approaches have been followed in the development of actuation models. The simplest, and most insightful, categorization of these models is achieved by considering their

Stellenbosch University http://scholar.sun.ac.za

computational complexities. As would be expected, those models which have proven to be most expensive have generally delivered the best results. The most popular actuation models implemented to date are introduced below along with a description of at least one existing facial animation system using that model. The discussion of each model will serve to highlight both the positive aspects of the model, and its associated inadequacies.

3.1.1 Parametric

The concept of a parametrized actuation model was introduced by PARKE [25, 26] in 1974. This model uses sets of parameters, and facial synthesis functions which utilize these parameters, to generate facial animation sequences. Manipulation of the parameters results in a direct manipulation of the facial model. The motivation for the development of this model was the desire to create an encapsulated model that could generate a wide range of faces and facial expressions based on a small set of input control parameters.

Two general classes of parameters were introduced by PARKE: *expression* parameters and *conformation* parameters. As the names suggest, expression parameters are concerned with the generation of facial expressions. Examples of such parameters are eyebrow-separation, jaw-rotation and mouth-corner-position. Conformation parameters on the other hand are concerned with controlling the range of facial conformation and include parameters such as nose-length, chin-shape and eye-size. In general, most facial animation systems are concerned only with facial expression control, and therefore do not attempt to define conformation parametrizations.

The development by PARKE [25, 26], of the first parametrized facial animation system was a significant landmark in the field of computer facial animation as it ushered in the possibility of facial animation generated at interactive rates. Over the years, this system has been extended by several third parties and a version known as FASCIA has been released into the public domain. PARKE'S original system, and all of its derivatives, are based on a polygonal surface model.

The primary advantages associated with the use of a parametrized actuation model relate to the inherent simplicity of the model; the fact that it is computationally inexpensive and the fact that it allows for the varying of facial conformation (i.e parameters controlling the relative size, shape and position of facial features can be used to modify the facial structure). The disadvantages of the model depend largely on the direct coupling which exists between

Stellenbosch University http://scholar.sun.ac.za

parameter sets and facial topology. A specific parametrization therefore can only operate on a specific topology. If a different facial model needs to be used, either its topology must be adapted to that used by the system, or the relation between parameters and topology that exists within the system must be redefined; both of which are non-trivial tasks. Also, the parametrization is generally not related to actual facial anatomy; a characteristic which makes features of realistic facial animation such as skin wrinkling and bulging hard to achieve.

3.1.2 Pseudomuscle-Based

The difficulty of simulating realistic facial motion using computationally inexpensive parametrized actuation models lead to the development of pseudomuscle-based systems. Such systems attempt to achieve a compromise between computational expense and visual realism by abstractly simulating facial anatomy. A pseudomuscle-based approach was followed, in 1988, by MAGNENAT-THALMANN, PRIMEAU & THALMANN [40] in their Abstract Muscle Actions (AMA).

AMA's are procedures which simulate the specific action of single facial muscles, or closely related facial muscle groups, and which are controlled by parameters. These procedures are based on empirical models and not on actual physical simulation. The most renowned use of AMA's was for the simulated characters based on Marilyn Monroe and Humphrey Bogart which appeared in the animated short film *Rendez-vous à Montréal* in 1987 [21].

The best example of a pseudomuscle-based actuation model is that introduced in 1992 by KALRA *et al.* [17, 18] which was an extension of the AMA concept. In this case however, the simulation procedures were based on rational free form deformations[1].

Pseudomuscle-based approaches represent a performance/quality compromise weighted slightly in favour of the performance side of the compromise. This is in contrast to parametrized actuation models which make no compromise and are geared solely towards performance.

---

[1] Free form deformation [36] is a technique for deforming solid geometric models in a free form manner and involves a mapping from $\Re^3$ to $\Re^3$ through a trivariate tensor product Bernstein polynomial. Physically FFD corresponds to deformations applied to a parallelpiped of clear, flexible plastic in which is embedded the geometric object to be deformed. As the parallelpiped is deformed, so too is the object embedded within.

Stellenbosch University http://scholar.sun.ac.za

3.1.3 Muscle-Based

The differentiation between muscle-based and pseudomuscle-based actuation models is not entirely clear. In trying to achieve this categorization the following question should be asked: Does the actuation model try to simulate the actual physical processes involved in facial expression, or does it attempt to simulate an abstraction of these processes? If the model tends towards the former, then it should be considered a muscle-based approach; if towards the latter, then it should be considered a pseudomuscle-based approach. It should be apparent that muscle-based approaches tend to be computationally more expensive than pseudomuscle-based approaches.

In the early eighties, PLATT and BADLER [32, 33, 34, 35] developed a facial animation system which presented the face as a hierarchically structured, regionally defined object. This system animated a polygonal facial mask using an approach which applied forces to an elastic mesh representation of the facial surface.

In 1987 WATERS [47] introduced, what has today proven to be by far the most popular approach to muscle modeling in computer facial animation. His actuation model, like that of PLATT and BADLER, was also based on a simple mass-spring skin tissue model, but it utilized geometrically defined muscles, each determined by a small set of parameters. The model also catered for two different types of muscles: linear and sphincter, in an attempt to simulate the different types of actual human facial muscles.

The popularity of this model was largely due to the fact that now, the actuation model could be decoupled from the structural model (i.e. the facial model topology) meaning that several distinct facial models could be animated using the same methods.

Whilst all muscle-based approaches attempt to simulate facial musculature, most generally neglect to simulate other primary components of human facial anatomy such as the underlying bone structures and the skin, or if they do, they generally use overly simplified models. This decision usually relates to the performance/quality compromise in the sense that a decision is made to invest the greatest proportion of allowable complexity in the muscle models, as the author is of the opinion that this has the greatest influence on the actual quality of the generated facial animation.

This decision can however be questioned as there have been cases where the lack of complexity in the simulation of other facial components has degraded the quality of the facial animation to such an extent that the decision can not be warranted. A case in point is the initial facial

Stellenbosch University http://scholar.sun.ac.za

animation system of PLATT and BADLER, introduced above, where the quality of the facial animation was adversely affected by the over-simplified mass-spring tissue model.

In general however, muscle-based approaches represent the most unbiased compromise between performance and quality, and it is this fact that has lead to the overwhelming popularity of the approach.

3.1.4 Physically-Based

In direct contrast to parametrized actuation models, where performance is the priority, are physically-based actuation models whose sole aim is the realistic simulation of human facial structures and processes. In 1990 TERZOPOLOUS and WATERS [37] developed one of the most ambitious physically-based systems to date. This system included an anatomically-based facial muscle process combined with a physically-based model of human facial tissue, all of which existed within the context of, and operated on, a polygonal facial model. The output from this system was significantly better than any previous attempts at facial animation and included effects such as realistic wrinkling and bulging of the skin, during muscle contraction.

The heart of this system was a tri-layer mass-spring mesh used to model the facial tissue. Simulated muscles were then embedded in the facial tissue mesh. A noteworthy feature of these muscle models is that they included the notions of volume-preservation (a fundamental characteristic of human muscle tissue) and skull penetration. These features greatly improved the visual realism of the facial animation.

As can be expected, this visual realism comes at quite a cost since most of the tissue simulation involves numerical integration of complex differential equations. Generally physically-based systems can only achieve interactive rates on multi-processor systems with significant graphics hardware, where the modeling and rendering processes can run as separate threads of execution.

3.1.5 Summary

The above discussion highlighted the varying computational complexities of actuation models. In particular it pointed out the unavoidable compromise which must be struck between the desire for visual fidelity and the need to provide performance at interactive rates. In terms of the SASL-MT Facial Animation System, the requirement that facial animation sequences must be generated at interactive rates whilst maintaining reasonably high levels of visual realism

Stellenbosch University http://scholar.sun.ac.za

imply that the actuation model should be muscle-based since this represents a fairly unbiased compromise.

3.2 Animation Models

The second subsidiary constraint relates to the need for explicit temporal control of facial animation sequences. WANG [45] defined the *animation model* of a character animation system as "*a method of grouping and modifying the actuation parameters for animation.*" The focus in this section will be on the modification of the actuation parameters. Discussion of the parameter organization and grouping will be left to a later section as these are issues more closely related to user interface design.

Again, several diverse approaches exist, however most of these can be seen as belonging to one of two major classes: *performance driven* animation, or *script driven* animation. These two classes of animation control will be discussed in turn below.

3.2.1 Performance Driven Animation Model

The process of coordinating a facial animation sequence generally requires the manipulation of a multitude of parameters over a certain period of time. If the number of parameters is small, or the period of time is short, this does not generally pose a significant problem. If on the other hand, as is generally the case, the number of parameters is significant, and so too the time period, this becomes a non-trivial task. If there are subsets of parameters each requiring coordination with other subsets the complexity of the task can expand rapidly.

Now, if each parameter must be specified *by hand* for each frame of an animation sequence, the amount of work required to specify an animation sequence of just a few seconds can easily become overwhelming. It is in an attempt to avoid this workload that performance driven animation is usually employed. In this model, parameter values are determined by capturing the actual motion of a live performer and mapping this information onto the parameter set. In its final implementation, the use of this model implies that the time it takes for the live performer to act out the animation sequence is approximately equal to the time it takes to specify the animation sequence. Moreover, in environments with sufficient processing power, it is even possible to map the captured motion of the performer to the parameter set in real-time. This then implies total real-time control of an animation sequence, something which is an ideal

Stellenbosch University http://scholar.sun.ac.za

in most areas of computer animation.

As can be expected, this comes at a significant cost. This cost is two-fold: firstly, the use of a performance driven animation model necessitates the use of an actual live performer; something which is not always a possibility in certain environments. Secondly, it requires the use of motion capture devices such as data-gloves; equipment which can be prohibitively expensive (although, it should be added that this problem is being overcome by using computer-vision techniques to track human motion in video streams). For these reasons, performance driven animation has not been as popular as one would expect, given the significant reduction in workload associated with the use thereof.

Performance driven animation however has an additional significant advantage which is often overlooked. This advantage has to do with the fact that the believability and correct interpretation of a facial expression is closely linked to the temporal flow of that facial expression. A good example of this is the expression of surprise. If the onset and duration of the expression's display are not coordinated correctly, it can easily look like a lazy yawn. This is in no way a problem associated only with computer animation; in fact, of the twelve principles of animation introduced by THOMAS and JOHNSON [41] as the cornerstones of any animation production at Disney Studios, two (*Ease-in & Ease-out*) relate directly to timing and three others (*Follow-Through*, *Overlapping Action* & *Arcs*) indirectly.

When using a performance based animation model, one gets all this for free as a side-effect of using a live performer. Initially this might not seem so significant, but the magnitude of this advantage cannot be appreciated until one has actually tried to create a realistic facial expression by directly specifying the changes in a set of parameters over a period of time. As human beings we have had thousands of hours to learn how to accurately display emotions and expressions on our faces but when we try to discretely quantify and specify those same expressions we more often than not fail miserably. It is in this that the art of character animation can be found and it is the ability to accurately capture the temporal flow of a character which makes an animator an artist.

3.2.2 Script Driven Animation Model

A script driven animation model allows animators to specify *by hand* the values of actuation parameters [26, 40, 45]. More importantly, it allows the animator to describe the change of these

Stellenbosch University http://scholar.sun.ac.za

parameters over time, thereby allowing the specification of an animation sequence. Generally, in a script based approach, the animator is confronted with specifying a set of key-frames by specifying the values for parameters at specific points in time. The computer is then responsible for creating an animation sequence by interpolating between these values.

Given that in the previous section, the need to directly specify actuation parameters was indicated as being a serious problem, it is necessary to motivate what the advantages are of using this approach. One of the biggest advantages relates to the varying level of animation control provided by using a script driven animation model. A performance driven animation model operates at the highest possible level of control and does not provide detailed low level explicit control of timing (you can't tell an actor to smile for exactly 2.34 seconds and then switch to a frown in 0.7 seconds); a script driven animation model allows for this.

More importantly, it allows for hierarchical animation control. This then, is the way around the problem of specifying multiple values for large parameter sets. By allowing the hierarchical specification of parameter values, we once again significantly reduce the workload of the animator. For example, suppose a script driven animation model allows for the control of facial animation by directly specifying *key-frames*, but that it also allows for the naming and definition of *expressions* as the composition of several key-frames. To generate an animation sequence in which a character smiles then only requires the once-off specification of the key-frames which constitute the "smile" expression. The next time that the animator needs to create a smile she need only reference the expression she called "smile."

3.2.3 Summary

In summary, the use of a performance driven animation model allows for quick and simple creation of realistic facial animation sequences, but it requires the presence of a human performer, it requires expensive hardware, and it does not allow low-level control of the animation sequence timing. A script driven animation model on the other hand allows for direct, explicit control of all animation parameters, including timing, but this can be a laborious process and can also result in unrealistic animation sequences.

In terms of the SASL-MT Facial Animation System, the fact that explicit and detailed control over facial animation sequence timing is required, coupled with the fact that much of the labour intensive aspects of script driven specification can be overcome by using well-designed

Stellenbosch University http://scholar.sun.ac.za

user interfaces and tools implies that the animation model should be script-based. Also, since the SASL-MT project is concerned with text-based translation, using a performance driven animation model would not make sense.

3.3 Visualization Models

In section 1.3.1 the secondary constraint placed on the design of the SASL-MT Facial Animation System due to its role as a component of the SASL-MT system was introduced as being the requirement that the SASL-MT system should be developed and delivered on a widely accessible platform. This requirement has far reaching consequences and severely impacts the design of every aspect of the facial animation component.

Applications relying on three-dimensional computer graphics are generally designed in such a manner that the system can be divided into two logically separate (and ideally physically as well) components; one responsible for the management of visualization tasks and the other for the management of all other back-end tasks and processes. This division allows one to separate the design decisions accordingly.

Given the desire to maximize the accessibility of the SASL-MT system it seems a strategically sound decision to develop the system for deployment on the internet since this will ensure the most general applicability. This however introduces a large number of design constraints, the most significant of which is the need to maintain some form of platform independence.

When providing content on the internet, the use of a widely accepted standard such as HTML generally delivers the required platform independence. This however changes as soon as one tries to include three-dimensional graphics in this content. HTML was traditionally designed to include only static two-dimensional content, but as the internet has evolved the requirement for dynamic three-dimensional content has become apparent.

The decision was made to use the Virtual Reality Modeling Language (VRML) [43] as the underlying platform upon which all visualization tasks would be implemented. This would be supported by a Java-based back-end platform. The motivation for these decisions will be exposed in the remainder of this section.

Stellenbosch University http://scholar.sun.ac.za

3.3.1 Virtual Reality Modeling Language

The Virtual Reality Modeling Language (VRML) [43] is a text-based international standard for describing three-dimensional graphics content on the internet. In contrast to low-level graphics API's like OpenGL [24] or DirectX [23] which are interfaced at a source code level, VRML is a standalone graphics library.

To view a scene created using VRML, a VRML browser is required, which is a software component taking as input a standard text file containing VRML source code and generating as output the three-dimensional scene described by the file. What makes VRML particularly useful in the context of three-dimensional graphics on the internet is the fact that a VRML browser can generally be embedded in a standard HTML document as a plugin.

VRML is however more than just a static three-dimensional scene renderer. The language has an event based architecture and incorporates detailed timing mechanisms which allow animation of objects. VRML is a hierarchical language, and the elements of the hierarchy are abstract data structures known as nodes. There are various types of nodes, and associated with each node is a set of input events and output events. The dynamic nature of the language is achieved by routing output events from one node to the input events of other nodes.

Creating complex systems is however difficult using VRML alone. For this reason, VRML can be extended by using Java. Support for Java, both scripts and applications, is built into most VRML browsers. In this way then it is possible to "connect" a Java applet to a VRML browser via an API, known as the External Authoring Interface (EAI).

3.3.2 Summary

In an effort to satisfy the requirement that the SASL-MT Facial Animation System be delivered on a widely accessible platform the decision was made to build the system on a two-tier architecture. The computational processes would be managed in Java whilst the visualization processes would be implemented in VRML.

During the development of the SASL-MT Facial Animation System, the new standard for three-dimensional computer graphics on the internet known as X3D [52] came to the fore. The fact that X3D represents an evolution of VRML, combined with the fact that X3D compatible browsers are not yet commonplace has meant that the decision to use VRML instead of X3D seems reasonable. The process of making the SASL-MT Facial Animation System X3D

Stellenbosch University http://scholar.sun.ac.za

compatible is seen as being achievable in the future.

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 4

SASL-MT Facial Animation System

4.1 Introduction

The SASL-MT Facial Animation System implements a NURBS based surface model. The structural model supports the definition of a facial model as the union of multiple separate NURBS surfaces and supports the deformation of these surfaces as a single perceived surface. Facial model actuation is achieved using a muscle model process controlled via a FACS based parametrization.

4.2 Surface Model

4.2.1 Overview

The SASL-MT Facial Animation System implements a NURBS based surface model. Section 2.1.1 introduced a set of surface models that could be used in a computer facial animation system and highlighted some of the advantages and disadvantages of each approach. This section will describe the parametric surface model implemented in the SASL-MT Facial Animation System starting with a high-level overview of NURBS surfaces in general and ending with an investigation of NURBS in VRML.

4.2.2 Non-Uniform Rational Basis-Splines

NURBS[1] surfaces [3, 11] can be characterized as interpolated surfaces determined by a set of control points/vertices and a set of basis functions which are in turn determined by a set of what are commonly referred to as knot vectors (which are simply vectors of floating point numbers). The control points define a convex hull for the interpolated surface which is generated in the context of the basis functions and knot vectors.

---

[1] See Appendix A on page 68 for a detailed mathematical description of NURBS.

Stellenbosch University http://scholar.sun.ac.za

Control points have non-uniform local control of the generated surface. That is, each control point only affects the shape of a localized region of the generated surface and the magnitude of the region of influence can vary between control points. The knot vectors determine this area of influence for each control point. The non-uniform nature implies that NURBS surfaces have a parameter space in addition to the three-dimensional space in which they are displayed. This parameter space has two dimensions generally referred to as the U and V dimensions.

Control points are expressed in four-dimensional coordinates (x, y, z, w). The fourth coordinate w, commonly referred to as the weight of the control point, defines the magnitude of the influence that a specific control point has on the shape of the generated surface. This is in contrast to the knot vectors which determine the magnitude of the region of the generated surface influenced by the control point. Control points are generally represented in homogenous form.

NURBS surfaces have several important mathematical properties:

- Control points have local control of the surface.
- NURBS do not change under the standard geometric affine transformations or under perspective projections.
- The lattice connecting the control points of the surface forms a convex hull.
- NURBS surfaces are continuous up to the second derivative.
- Different segments of a NURBS surface can have different continuities. This is known as the multiplicity property. Placing multiple control points at the same position in space reduces the continuity of the surface at that point. A consequence of this is that by placing three points at the same position in space, one can force a discontinuity in the surface without affecting the continuity of the rest of the surface.
- Surface normals can be computed at any point on a NURBS surface. This characteristic facilitates high quality shading of NURBS surfaces.

The primary motivation for choosing NURBS as the basis of the surface model for the SASL-MT Facial Animation System was based on the following points:

Stellenbosch University http://scholar.sun.ac.za

- NURBS are ideally suited to representing complex curved surfaces such as the human face [14, 46]. NURBS also seem to be popular in this regard within the three-dimensional modeling community [39].
- Many professional tools exist for modeling and rendering NURBS surfaces [1, 5].
- A sparse control point grid can define a relatively complex NURBS surface. This in turn makes the actuation and manipulation of a NURBS surface fairly inexpensive. Also, given that the SASL-MT Facial Animation system will be used over the internet, any reduction in data set size represents a significant advantage.

4.2.3 Implementation

In 1999 the VRML97 specification was extended to include support for NURBS curves and surfaces [11]. Given that this extension does not form part of the core specification, not all existing VRML browsers support NURBS, but the Blaxxun Contact [15] browser which was chosen to serve as the recommended browser for use with the SASL-MT Facial Animation System provides almost complete support for the specified VRML extension. Blaxxun Interactive, the creators of the Blaxxun Contact VRML browser, have also created a 3D Studio Max [5] VRML export plug-in which can be used to export NURBS surfaces created in 3D Studio Max to VRML format.

4.3 Structural Model

4.3.1 Overview

The structural model of the SASL-MT Facial Animation system operates on a NURBS based surface model. The structural model supports the definition of a three-dimensional facial model as the union of multiple separate NURBS surfaces and supports the deformation of these surfaces as a single perceived surface.

A significant problem with parametric surfaces is encountered when the need for local refinement of a surface arises [45, 46]. In terms of a facial mask, such a requirement is evidenced when one considers the variation in surface complexity between areas of the face such as the cheeks and the ears. It is obvious that a sparse set of control points can be used to represent the surface of the cheek whereas a much denser set of control points is required to represent

Stellenbosch University http://scholar.sun.ac.za

the ear. Areas of the human face requiring local refinement are the mouth, nostrils, eyelids and ears. The nature of parametric surfaces necessitates that local refinement of the control point grid can only be achieved by a global refinement; that is, the introduction of a new row or column of control points. This means that if a facial mask is to be represented by a single parametric surface, then the density of the control point grid must be defined by the region of the surface with maximal complexity. This problem is a significant obstacle to the creation of a single non-uniformly curved parametric surface representation of a human face, since the density of the control point grid becomes unmanageable.

Two general approaches have been introduced to overcome this problem. The first introduces the notion of a hierarchical parametric surface, in which local refinement is achieved by adding new parametric surface patches of successively greater refinement within specified regions of a surface. This is the approach that was followed by WANG [45, 46] in the development of the Langwidere facial animation system (see section 2.1.3).

An alternate approach is to align multiple distinct parametric surfaces of varying levels of refinement in such a manner that they appear to form a single continuous surface. This allows areas of varying complexity of the facial mask to be created from independent NURBS surfaces [42]. This is the approach that has been adopted for the SASL-MT Facial Animation System.

This approach seems to be a popular method employed in the three-dimensional modeling community for representing human faces with NURBS surfaces [50]. It however gives rise to a set of challenging problems, the most significant of which becomes evident when actuating the NURBS surfaces during animation. The static alignment of different NURBS surfaces is relatively simple to achieve in a heuristic fashion; maintaining this alignment during animation is however not as simple. If the alignment between surfaces is lost, cracks will appear at the joining edges destroying the continuity and realism of the face.

The primary motivation for choosing the approach of multiple aligned surfaces for the SASL-MT Facial Animation System was based on the following points:

- Modeling and rendering support for hierarchical parametric surfaces is not widespread and there is no VRML support for such surfaces.

- Modeling a human face using multiple aligned NURBS surfaces is significantly simpler than trying to model the same face using a single NURBS surface. In fact, single surface

Stellenbosch University http://scholar.sun.ac.za

facial masks all typically suffer from wrinkling [46] (due to excessive control point refinement in areas of minimal curvature) and pole problems [27] (visual discontinuities in the surface in areas where several edges of the surface meet; typically at the top of the head).

Facial Models

A major design goal for the SASL-MT Facial Animation System was the ability to animate an arbitrary three-dimensional, NURBS based, facial model. Many existing facial animation systems, whilst being able to create realistic facial animations for a specific facial model or a specific facial topology, were restricted in their flexibility with regards to adaptation to arbitrary facial models. In fact, some of the earlier systems were designed to operate on a single model [25, 33, 34, 44].

The animation of arbitrary three-dimensional NURBS based facial models is supported in the SASL-MT Facial Animation System. Given that NURBS are well-defined mathematical objects, three-dimensional models based on NURBS can be defined, modified and interpreted independent of any proprietary software packages. This combined with support for NURBS in the VRML specification make it possible to take any NURBS surface created using any three-dimensional modeling package and include it in a VRML scene.

The SASL-MT Facial Animation System makes several assumptions about the facial model defined in a specific VRML scene. These assumptions primarily serve to define the format of the VRML file containing a facial model. This is necessary because certain regions of the facial model must be distinguished from others (for example the eyeballs) and the components of the facial model must be distinguished from all other elements of the scene.

Adapting a VRML scene containing a three-dimensional facial model for animation with the SASL-MT Facial Animation System requires that certain naming conventions be followed with regards to node naming; that certain nodes be grouped according to certain standards and that some additional information such as eyeball rotational centers be specified. Given that VRML is designed around human readable/editable files, achieving the above adaptation is relatively simple.

Stellenbosch University http://scholar.sun.ac.za

4.3.2 Implementation

To illustrate the process of adapting an arbitrary facial model for animation via the SASL-MT Facial Animation System, consider the example of Adam, a three-dimensional model created by WILSON [50] using Maya 4 [1]. To prepare the model for animation using the SASL-MT Facial Animation System and to avoid manual editing of the original source file (in a proprietary format) we automatically converted the file to IGES format using Maya 4 and subsequently imported it into 3D Studio Max [5]. From here the model was automatically exported as a VRML scene using the Blaxxun VRML export plug-in. Finally the VRML source file was manually edited to ensure that all naming conventions were adhered to and that all grouping requirements were satisfied.

The facial mask is depicted in figure 4.1 on page 35. In the figure, the mask has been coloured in such a manner that the twelve NURBS surfaces constituting the right hand side of the face are clearly visible. The left hand sided of the face has been coloured naturally and appears as a single continuous surface.

Figure 4.1: The Adam facial mask coloured so that all 12 individual NURBS surfaces constituting the right half of the facial model are clearly visible. (The surface constituting the philtrum and inner nostrils is difficult to discern in the figure.)

Stellenbosch University http://scholar.sun.ac.za

4.4 Actuation Model

4.4.1 Overview

The SASL-MT Facial Animation System utilizes a muscle model based on that developed by WATERS [47]. The primary motivation for choosing WATERS'S muscle model approach was based on the requirement that facial animation sequences must be generated at interactive rates whilst maintaining reasonably high levels of visual realism.

WATERS'S original model was extended to include four muscle types: linear, sphincter, sheet and eyelid muscles. The eyelid muscles have no real physiological basis and are therefore purely abstract. In a human face, the muscles controlling the movement of the eyelids are a complicated set of linear and sphincter muscles. Realistic movement of the eyelids in a facial animation system using only the other three muscle models is difficult to achieve because of the intricate interleaving of the muscles around the eye. For this reason, a separate model was developed for the actuation of the eyelids. Before proceeding with a detailed discussion of each of the muscle types an introduction to terminology relating to, and general discussion of, the muscle models will be given.

Each muscle is defined by a *muscle vector* which is determined by the *attachment* and *insertion* of the muscle. The muscle vectors determine the direction of muscle contraction; in general, towards the attachment.

Each facial muscle affects a specific region of the face. Since the surface model of the system consists of NURBS, each muscle influences a region of one or more of the NURBS surfaces constituting the face. The region of influence can therefore be specified in terms of the union of subsets of the control points of several NURBS surface (since a single muscle influence can span several surfaces).

The set of control points influenced by a specific muscle will be referred to as the *activation set* of that muscle. The influence of a muscle's contraction on a specific point in the activation set varies depending on that point's position relative to the attachment of the muscle. A function is therefore defined for each muscle type to determine the fall-off of a muscle's influence.

The activation set of a muscle, and the fall-off function, are determined by a set of parameters associated with that muscle. It is this set of parameters and associated fall-off function which differentiates the four muscle types. The determination of the activation set based on these parameters is done as a preprocessing step before actual animation.

Stellenbosch University http://scholar.sun.ac.za

To achieve realistic muscle contraction, a muscle must have a contraction limit. This limit is a property of each individual muscle. Typically human facial muscles have a contraction limit of 25% of their neutral length, but experimentation has shown that, since the muscle model is only a simulation of physical muscles, a mechanism for overriding this limit is required. Determining this contraction limit without a visual reference is impossible. For this reason, specification of a muscle contraction limit is done external to the process of muscle creation, and therefore the notion of such a limit is not included directly in the muscle model.

4.4.2 Muscle Model Implementation

The Linear Muscle Model

The linear muscle model described below is derived from that developed by WATERS [47]. Linear muscles are the most prolific in the facial anatomy. Examples of linear muscles are the zygomatic major muscle, which raises the corner of the lips in a smile, and the corrugator supercilli, which raises the inside of the eyebrow.

A linear muscle vector is determined by an attachment point and an insertion point, both of which are specified by single control points on some NURBS surface. The primary parameters characterizing a linear muscle are its angle of influence and its insertion and dissipation radii.

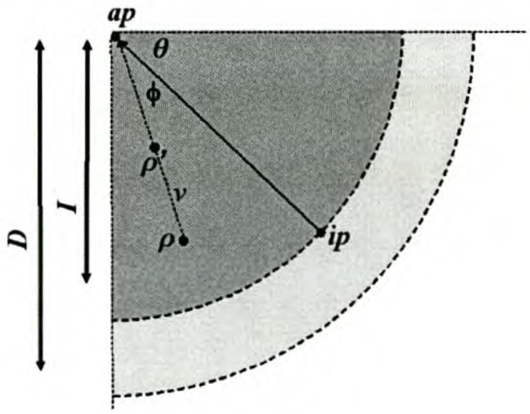


Figure 4.2: The linear muscle vector model.

The angle of influence θ of a linear muscle specifies the zone of influence of the muscle in terms of an angular deviation from the muscle vector determined by the attachment point ap, and insertion point ip, as illustrated in figure 4.2. The insertion radius I is the magnitude of the muscle vector, that is to say, the distance between the attachment point and the insertion point. The dissipation radius D is a directly specified distance and specifies the zone of influence of the muscle in terms of a radial displacement from the attachment point.

A point on the surface of the face, say ρ, is in the activation set S of a specific linear muscle if that point falls within the dissipation radius of the muscle, and the angle between the muscle vector and the point is less than or equal to the angle of influence. Furthermore, the two radii I and D divide the activation set into two distinct subsets: the insertion set S_I and the dissipation set S_D. S_I consists of all points whose radial distance from ap is less than or equal to I, and S_D consists of all points whose radial distance is greater than I and less than or equal to D.

Let $v = \rho - ap$ be the vector between the point ρ and the attachment point ap, let $m = ip - ap$ be the muscle vector and let P be the set of all control points constituting the face. Then

$$\begin{aligned} S &= \left\{\rho \in P \ : \ \|v\| \leq D \ \text{ and } \ \frac{v \cdot m}{\|v\|\|m\|} \geq \cos(\theta)\right\}, \\ S_I &= \{\rho \in S \ : \ \|v\| \leq I\}, \ \text{and} \\ S_D &= \{\rho \in S \ : \ I \leq \|v\| \leq D\}. \end{aligned}$$

When a linear muscle contracts, all the points in the activation set are displaced towards the attachment point. The attachment point stays fixed, so we assume it has zero displacement, while the insertion point is the point of maximum displacement. All other points are displaced according to a fall-off function, dependent on the position of those points relative to the attachment point.

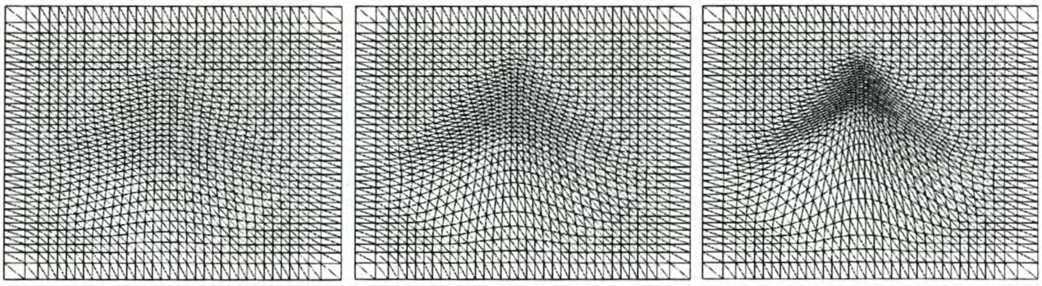


Figure 4.3: A linear muscle with 45° angle of influence contracted at 30%, 60% and 100%.

Therefore a point $\rho(x, y, z)$ in the activation set is displaced in the direction of the attachment point to the new point $\rho'(x', y', z')$ with this displacement being determined by the fall-off function. The fall-off function is dependent on the two variables $\|v\|$ and $\cos(\phi) = \frac{v \cdot m}{\|v\|\|m\|}$. For

Stellenbosch University http://scholar.sun.ac.za

linear muscles then, the displacement of ρ to ρ' is found using the following expression:

$$\rho' = \rho + ar\frac{v}{\|v\|},$$

where

$$a = \cos(\phi) = \frac{v \cdot m}{\|v\|\|m\|}$$

is the cosine of the angle between v and m and where

$$r = \begin{cases} \cos\left(\frac{I-\|v\|}{I} * \frac{\pi}{2}\right) & \text{for } \|v\| \leq I \\ \cos\left(\frac{\|v\|-I}{D-I} * \frac{\pi}{2}\right) & \text{for } I < \|v\| \leq D \end{cases}$$

is a radial displacement parameter and $\frac{v}{\|v\|}$ is the unit vector in the direction of v. See figure 4.3 for an example of a contracting linear muscle.

The Sheet Muscle Model

Unlike linear muscles, sheet muscles do not contract towards a localized point. Sheet muscles consist of a group of parallel aligned fibers spread over an area. An example of a sheet muscle is the frontalis muscle which covers most of the forehead, and is responsible for raising the eyebrows.

A sheet muscle is also defined by an attachment point ap and insertion point ip as illustrated in figure 4.4. The muscle vector $m = ip - ap$, however, determines the direction of contraction for all points in the activation set of a sheet muscle. The primary parameters characterizing a sheet muscle are the parallel insertion radius I, the parallel dissipation radius $D_{||}$ (the definitions of which are the same as for the linear muscle model) and the perpendicular dissipation radius $D_{\perp}$ which specifies the zone of influence of the muscle in terms of a perpendicular displacement from the attachment point. A point on the surface of the face, say ρ, is in the activation set S of a specific sheet muscle if that point falls within both the parallel and perpendicular dissipation radii of the muscle. Furthermore, the two radii I and $D_{||}$ divide the activation set into two distinct subsets: the insertion set S_I and the dissipation set S_D. S_I consists of all points whose parallel radial distance $r_{||}$ from ap is less than or equal to I, and S_D consists of all points whose parallel radial distance is greater than I and less than or equal to D.

Let $v = \rho - ap$ be the vector between the point ρ and the attachment point ap, let $m = ip - ap$ be the muscle vector, let $r_{||} = \|(\|v\| \cos(\phi))\|$ be the magnitude of the vector component

Stellenbosch University http://scholar.sun.ac.za

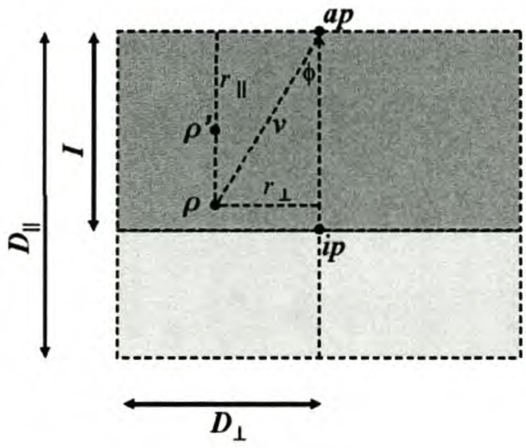


Figure 4.4: The sheet muscle vector model.

of v parallel to m, let $r_{\perp} = \|(\|v\| \sin(\phi))\|$ be the magnitude of the vector component of v perpendicular to m and let P be the set of all control points constituting the face. Then

$$\begin{aligned} S &= \left\{\rho \in P \ : \ r_{\perp} \leq D_{\perp} \text{ and } r_{\|} \leq D_{\|}\right\}, \\ S_I &= \left\{\rho \in S \ : \ r_{\|} \leq I\right\}, \text{ and} \\ S_D &= \left\{\rho \in S \ : \ I \leq r_{\|} \leq D_{\|}\right\}. \end{aligned}$$

When a sheet muscle contracts, all the points in the activation set are displaced parallel to the direction of the muscle vector. The attachment point stays fixed, so we assume it has zero displacement, while the insertion point is the point of maximum displacement. All other points are displaced according to a fall-off function, dependent on the position of those points relative to the attachment point.

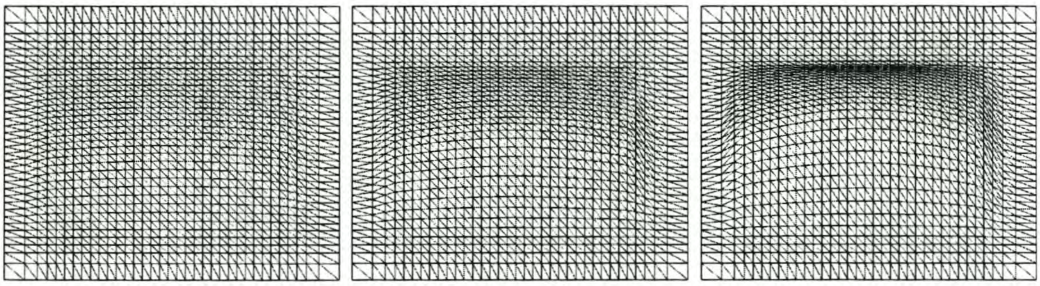


Figure 4.5: A sheet muscle contracted at 30%, 60% and 100%.

Therefore a point $\rho(x, y, z)$ in the activation set is displaced in a direction parallel to that of

Stellenbosch University http://scholar.sun.ac.za

the muscle vector to the new point $\rho'(x', y', z')$ with this displacement being determined by the fall-off function. The fall-off function is dependent on the two variables $r_{||}$ and $r_{\perp}$. For sheet muscles then, the displacement of ρ to ρ' is found using the following expression:

$$\rho' = \rho + a_{||} a_{\perp} \frac{m}{||m||}$$

where

$$a_{\perp} = \cos\left(\frac{r_{\perp}}{D_{\perp}} * \frac{\pi}{2}\right)$$

and

$$a_{||} = \begin{cases} \sin\left(\frac{r_{||}}{||m||} * \frac{\pi}{2}\right) & \text{for } r_{||} \leq I \\ \sin\left(\frac{r_{||}}{D_{||}} * \pi\right) & \text{for } I < r_{||} \leq D \end{cases}$$

and where $a_{||}$, $a_{\perp}$ are radial displacement parameters and $\frac{m}{||m||}$ is the unit vector in the direction of m. See figure 4.5 for an example of a contracting sheet muscle.

The Sphincter Muscle Model

Sphincter muscles, unlike all other facial muscles, have both attachments in skin. When they contract they squeeze the surrounding tissue together like the mouth of a string bag. An example of a sphincter muscle is the orbicularis oris which controls the puckering of the lips.

A sphincter muscle is also defined by an attachment point *ap* and insertion point *ip* as illustrated in figure 4.6. The primary parameters characterizing a sheet muscle are the semimajor r_{maj} and semiminor r_{min} radii, the values of which are a function of the insertion radius I. The semimajor and semiminor radii determine the elliptical zone of influence of a sphincter muscle. A point on the surface of the face, say ρ, is in the activation set S of a specific sphincter muscle

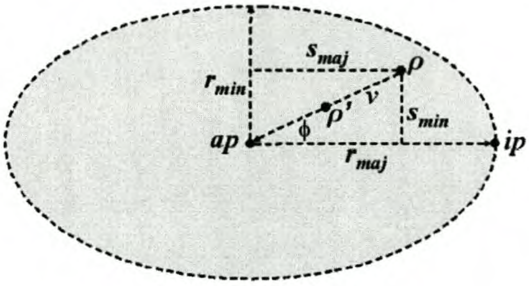


Figure 4.6: The sphincter muscle vector model.

Stellenbosch University http://scholar.sun.ac.za

if that point falls within the ellipse with semimajor and semiminor radii determined by r_{maj} and r_{min} respectively.

Let $v = \rho - ap$ be the vector between the point ρ and the attachment point ap, let $m = ip - ap$ be the muscle vector, let $s_{maj} = \|(\|v\| \cos(\phi))\|$ be the magnitude of the vector component of v parallel to m, let $s_{min} = \|(\|v\| \sin(\phi))\|$ be the magnitude of the vector component of v perpendicular to m and let P be the set of all control points constituting the face. Then

$$S = \left\{ \rho \in P \ : \ \left(\frac{s_{maj}^2}{r_{maj}^2} \right) + \left(\frac{s_{min}^2}{r_{min}^2} \right) \leq 1 \right\}.$$

When a sphincter muscle contracts, all the points in the activation set are displaced towards the attachment point. The attachment point stays fixed, so we assume it has zero displacement. All other points are displaced according to a fall-off function, dependent on the position of those points relative to the attachment point.

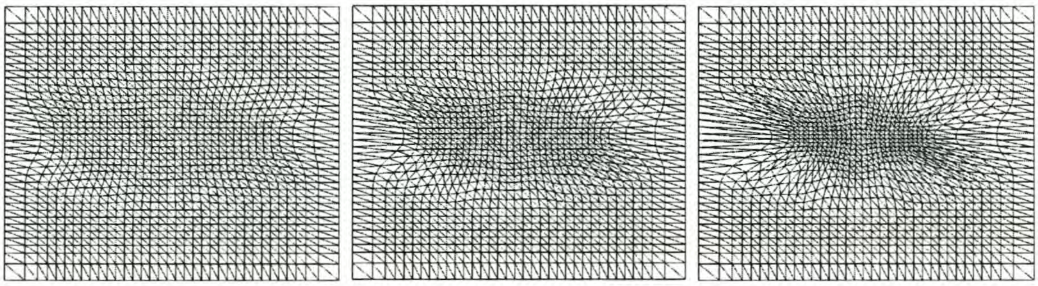


Figure 4.7: A sphincter muscle contracted at 30%, 60% and 100%.

Therefore a point $\rho(x, y, z)$ in the activation set is displaced in the direction of v to the new point $\rho'(x', y', z')$ with this displacement being determined by the fall-off function. The fall-off function is dependent on the two variables s_{maj} and s_{min}. For sphincter muscles then, the displacement of ρ to ρ' is found using the following expression:

$$\rho' = \rho + a \frac{v}{\|v\|}$$

where

$$a = 1 - \frac{\sqrt{r_{maj}^2 s_{maj}^2 + r_{min}^2 s_{min}^2}}{r_{maj} r_{min}}$$

and a is a displacement parameter and $\frac{v}{\|v\|}$ is the unit vector in the direction of v. This model might initially seem incorrect and unnatural. The motivation for the fall-off function is based entirely on experimentation. Several fall-off functions were tested, and the above function delivered the most realistic, three-dimensional, visual sphincter muscle contraction. See figure 4.7 for an example of a contracting sphincter muscle.

The Eyelid Muscle Model

As was stated earlier, the eyelid muscle model implemented in this system has no real physiological basis. There are several reasons which lead to the development of a separate model for the eyelids. The most significant is the fact that the muscles controlling the eyelid are difficult to model using the other three muscle models. The other issue concerning eyelid motion which requires separate attention is the fact that to simulate realistic eyelid motion it is essential that the eyelid, during motion, must track the surface of the eyeball. The other three muscle models cannot ensure that this occurs.

Developing a realistic muscle model for the eyelids proved to be difficult, not only because of the two reasons already mentioned, but also because of the nature of the eyelid itself. The eyelid is an extremely thin section of skin, and many people have been tempted to model eyelids as a single flat surface with no volume. This however results in an unrealistic appearance. The eyes are the focal points of the face, and any imperfections in eyes are blatantly obvious.

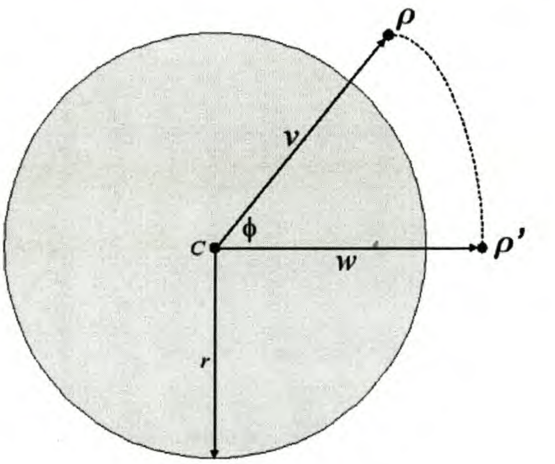


Figure 4.8: The eyelid muscle vector model.

The model that was finally implemented was kept as simple as possible. Unfortunately the model does not have the same mathematical grounding of the other three. The eyelid muscle model has no determining parameters. The activation set of an eyelid muscle must be specified

Stellenbosch University http://scholar.sun.ac.za

by the user. This is generally not a problem because the eyelid should consist of only a few control points.

Contraction of an eyelid muscle is achieved by simply rotating the control points in the activation set around the center of the eyeball, and then ensuring that the control points lie on the surface of the eyeball. Suppose ρ is a point in the activation set of an eyelid muscle. When the muscle contracts ρ must be displaced to some other point ρ'. Let the point c be the center of the eyeball with radius r and let the vectors $v = \rho - c$ and $w = \rho' - c$ be the vectors from c to ρ and ρ' respectively.

If $\|v\| \geq r$ then certainly it must hold that after rotation about the center of the eyeball c, $\|w\| \geq r$. It is therefore assured that no control point outside of the eyeball can ever intersect the eyeball surface. It can however occur that after rotation, the control point is far enough away from the eyeball surface that it will look unrealistic. For this reason, once points are rotated about the center, they are displaced by means of a simple vector operation to be as near as possible to the eyeball surface.

4.5 Animation Model

4.5.1 Overview

In section 3.2 two different animation models were introduced: performance driven and script driven. It was stated that within the context of the SASL-MT Facial Animation system a script driven animation model is required. This decision was motivated primarily by the fact that a script driven animation model can provide absolute and explicit temporal control of animation sequences. This decision is supported by the animation model inherently supported by VRML.

Obviously the ultimate goal of any facial animation system is the generation of facial animation sequences. Within the SASL-MT Facial Animation System the most fundamental building block of a facial animation sequence is a facial muscle contraction. Facial muscles, in the context of the SASL-MT Facial Animation System, are elements of the muscle model (i.e. linear, sheet, sphincter or eyelid muscles) associated with elements of the expression model (i.e. action units).

A facial animation sequence is however much more than a series of facial muscle contractions. The significance of temporal control over facial animation sequences has already been discussed. Since facial muscle contractions have no intrinsic temporal characteristics additional constructs

Stellenbosch University http://scholar.sun.ac.za

must be defined.

Within the SASL-MT Facial Animation System, a facial animation sequence is defined as a sequence of facial expressions. Facial expressions are in turn defined as series of facial poses which are in turn defined as a set of muscle contractions.

Facial Poses

In the SASL-MT Facial Animation System a facial pose is defined as an instantaneous snapshot of the facial model under the contraction of one or more facial muscles. A facial pose therefore is determined by and specified in terms of percentage contractions of action units. For example, a certain facial pose could be defined as

60% AU12 - Lip Corner Puller +
40% AU11 - Nasolabial Furrow Deepener +
20% AU20 - Lip Stretcher.

Facial Expressions

Evidencing the significance of temporal control is the fact that the most fundamental property of a facial expression is its duration. All other aspects of a facial expression are determined relative to the expression duration.

Facial expressions are determined by a duration and a series of *facial movements*. A facial movement is an ordered instance of a facial pose. Each facial movement is associated with a duration defined relative to the associated facial expression duration. For example, suppose three facial poses have been defined as say

$$P_1 = \{60\%\ AU12,\ 40\%\ AU11,\ 20\%\ AU20\},$$
$$P_2 = \{80\%\ AU12,\ 75\%\ AU11,\ 40\%\ AU20\},\text{ and}$$
$$P_3 = \{10\%\ AU12,\ 5\%\ AU11,\ 7\%\ AU20\}.$$

A new facial expression E with duration $d = 2.34$ seconds can then be defined as

$$10\%P_1 + 80\%P_2 + 10\%P_3.$$

This should be interpreted as follows: suppose that at time $t = 0$ the facial model is in pose P_0. Then, at time $t = 0.10 * d = 0.234$ the facial model will have changed from facial pose P_0 to P_1. At time $t = 0.80 * d + (0.234) = 1.872 + 0.234 = 2.106$ the facial model will have changed from

Stellenbosch University http://scholar.sun.ac.za

facial pose P_1 to P_2. And finally at time $t = 0.10 * d + (2.106) = 2.34$ the facial model will have changed from facial pose P_2 to P_3.

In this manner, facial expressions can be constructed as sequential facial movements. Since each facial pose is defined independently of any facial expression, facial poses can be reused, thereby decreasing the specification effort. Moreover, since facial movement durations are defined relative to facial expression duration, an expression can be modified either by modifying the percentage durations of individual movements, or by modifying the total expression duration. Full animation sequences are constructed by scheduling sequentially executed facial expressions.

It should also be noted that facial expressions and poses are independent of the underlying muscle specifications. That is facilitated by the fact that facial poses are defined in terms of action unit names and not muscle instances. This means that facial expressions can be used across different muscle sets and more importantly across different facial models as long as the underlying muscle set contains the same action units. This too serves to significantly reduce the configuration effort since facial expressions and poses only need to be defined once and can then be applied to multiple facial models. The process of animating a new facial model then reduces to importing that facial model and defining an equivalent muscle set for the model. In fact, if two facial models share the same topology even the muscle sets can be reused, reducing the configuration effort even further.

4.5.2 Implementation

It is obvious that in terms of computer animation, the facial poses described in the previous sections represent keyframes in an animation sequence. Within the SASL-MT Facial Animation System facial animation sequences are generated using a keyframed animation approach natively supported by VRML. The generation of an animation sequence depends on two independent phases. The first *offline* phase consists of muscle, facial pose and facial expression definition. This is a once-off phase, and once a facial expression has been defined it can be displayed at any time. This phase could be described as *keyframe definition*, but that could lead to confusion since it is not the actual keyframes (i.e. NURBS surface states) that are being defined. This phase only involves the specification of a parameter set, which is later used as input to the realtime animation sequence generation process.

Stellenbosch University http://scholar.sun.ac.za

The actual display (realtime rendering) of a facial expression represents the second *online* phase. During this phase, actual keyframes are determined and constructed and fed to the VRML browser for animation. A brief overview of the *online* phase will be given below. The *offline* phase will be discussed in section 4.6 on page 48.

Generating Animation Sequences

In this section the generation of an animation sequence consisting of the display of a single facial expression will be described. An animation sequence consisting of multiple expressions would be generated by repeating the process described below for each expression in the sequence.

1. For each of the facial movements in the facial expression do the following:
    (a) Contract each of the muscles in the facial pose. This results in the displacement of the control points within the activation set of the muscle and therefore in a localized distortion of the facial model.
    (b) Record the current state of the facial model as a keyframe in the animation sequence according to the temporal parameters configured for the facial movement.
2. Interpolate between keyframes using standard bilinear interpolation according to the temporal parameters configured for the facial expression.

The above algorithm can be easily implemented in VRML using `CoordinateInterpolator` and `TimeSensor` nodes. With each NURBS surface constituting the facial mask is associated a `CoordinateInterpolator` node which is a simple data structure used to store keyframes for that specific node. Physically, the generation of a keyframe (see 1 (a) above) therefore consists of generating keyframes for each of the NURBS surfaces constituting the facial mask, i.e. for each of the `CoordinateInterpolator` nodes. An animation sequence is rendered by activating a `TimeSensor` node connected to all `CoordingateInterpolator` nodes. The `TimeSensor` generates as many interpolation steps as the system will allow between keyframes with the guarantee that each keyframe will be displayed at the configured time. This implies that animation sequences can never suffer from lagging, i.e. the requirement for absolute, explicit temporal control of animation sequences is satisfied. This in turn means that the facial animation sequence can be synchronized with other processes such as sign generation.

Stellenbosch University http://scholar.sun.ac.za

4.6 Expression Model

4.6.1 Overview

As discussed in section 3.2, a script driven animation model can be prohibitively expensive from a configuration point of view. In terms of the SASL-MT Facial Animation System this also holds true. Each muscle requires the specification of several parameters; the definition of facial poses requires the specification of individual muscle contractions; and the configuration of facial expressions requires the configuration of timing parameters for all facial movements constituting an expression. At first glance this process seems to represent a significant configuration effort. However, well designed user interfaces and reuse of existing configuration can significantly reduce the effort.

4.6.2 Implementation

The SASL-MT Facial Animation system is intended to be used as a component of a larger Sign Language visualization system. To this end, the system must be self contained and portable and support integration with other unrelated systems. To achieve this the system has been designed around a core animation engine responsible for producing animation sequences based on a predefined configuration and for interaction with a VRML browser. Extensibility is supported by a plug-in framework forming part of the core which allows plug-in modules to be added to extend the core functionality.

Plug-in modules are used to provide the configuration interfaces, and it is these interfaces which serve to reduce the configuration effort required to generate facial animation sequences due to the fact that the system utilizes a script driven animation model. Four configuration plug-ins have been implemented, all of which have been implemented as wizards which interactively guide users through the related processes. This approach significantly decreases time required to familiarize oneself with the system.

Muscle Creator/Editor

Two plug-ins exist which can be used to create and edit facial muscles. The Muscle Creator plug-in aims to simplify the process of muscle creation and guides the user through the parameter configuration for a muscle. However, the algorithm used to determine muscle activation sets

Stellenbosch University http://scholar.sun.ac.za

does not take into account the reality of human facial structure. For this reason, a Muscle Editor plug-in has been developed as well and facilitates the modification of existing muscles. Muscle parameters can be interactively modified and muscle activation sets can be manually edited. Muscles can also be manually specified in configuration files but the use of the plug-ins decreases the time required for muscle specification by several orders of magnitude.

Muscle Set Manager

Given the abstractions designed into the system with regards to facial pose and facial expression specification, there are no dependencies between facial expressions and poses and muscles (see section 4.5.1). That is to say, muscle sets can be switched in or out without redefining facial expressions. This means that facial expressions need only be defined once and can then be applied to different facial models (i.e. muscle sets) or different muscle sets can be applied to the same facial model without affecting the expressions defined for that model. The Muscle Set Manager plug-in provides the functionality to load/unload/edit muscle sets.

Muscle Controller

The Muscle Controller plug-in has been implemented to provide functionality to interactively contract individual muscles. This plug-in exposes a basic interface consisting of a slider bar for each muscle contained in the active muscle set and thereby allows the interactive contraction of each muscle. The Muscle Controller plug-in should be used in conjunction with the Muscle Creator and Editor plug-ins to view the action of muscles.

Expression Creator/Editor

A single plug-in has been implemented to support the configuration of facial movements and expressions. This plug-in exposes an interface familiar to anyone who has some knowledge of GUI based computer animation software packages. Facial expressions are defined in terms of facial movements which are represented on a graphical timeline. The timeline elements (facial movements) can be interactively modified using the interface and facial expressions can thus be configured in an intuitive manner.

All of the plug-ins combine to form a user-friendly graphical front-end to the SASL-MT Facial Animation System. The design is such that plug-ins can be dynamically loaded or unloaded

Stellenbosch University http://scholar.sun.ac.za

and the functionality of a specific instance of the system can therefore be increased or decreased depending on the environment in which it is being used. For example, when the SASL-MT Facial Animation System is used in conjunction with the encompassing SASL-MT System the configuration of muscles and facial expressions will be a once-off process. During configuration time it would be necessary to load all the plug-ins, but once the system is being utilized in a production environment there is no longer any need for all of the configuration plug-ins and these can be unloaded. This mechanism can also be used to limit access to certain features. A publicly accessible website might only load certain plug-ins while a website restricted to local access could expose all plug-ins.

4.7 Summary of the System

This chapter discussed the theoretical basis of the SASL-MT Facial Animation System and introduced various aspects relating to the implementation of the designs. The SASL-MT Facial Animation System was introduced as implementing a NURBS based surface model. The advantages of this decision were discussed in section 4.2.

Section 4.3 described the implementation of a NURBS based structural model and highlighted the problems of surface refinement and the solution employed in SASL-MT Facial Animation System. Next the implementation of the muscle model was discussed, in section 4.4, and the different muscle types were discussed in detail. Finally the animation and expression models were discussed, in sections 4.5 and 4.6. The approaches to expression and muscle configuration were introduced and the implementation of the system user interfaces were described.

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 5

Results

This chapter discusses the results achieved using the implementation of the SASL-MT Facial Animation System. It looks at the use of NURBS as a basis for a facial animation structural model and further investigates the muscle models used in the system. The use of VRML as basis for an animation system is also discussed. The applications of a facial animation system in the context of a Sign Language visualization system are revisited and finally a series of examples of facial expressions generated using the system are discussed.

5.1 NURBS Based Facial Models

Throughout the duration of this research effort, the author spent countless hours following various approaches in attempts to create realistic three-dimensional human faces. Several three-dimensional geometries were tested and none seemed better suited to the modeling of three-dimensional facial models than NURBS. As has been stated several times throughout this document, NURBS are mathematically well-defined objects and as such have been thoroughly researched and documented from various perspectives. Their nature is such that they are ideally suited to the modeling of organic objects and they are intuitive and easy to understand and use. The use of NURBS surfaces implies the use of easily manageable data sets and nullifies many problems often encountered with systems implemented in low-bandwidth environments.

Unfortunately, certain definite disadvantages can also be identified. There is as yet no hardware support for the rendering of NURBS surfaces. This has an immediate impact on any system utilizing NURBS as a surface. Fortunately the tesselation process whereby a NURBS surface is converted to an equivalent polygonal surface is not complex and given a mid-range CPU this impact can be avoided to a large extent. Within the SASL-MT Facial Animation System, this problem has proven to have little impact. The performance improvements gained by using reduced sized data sets far outweigh the penalties incurred by using a partially software

Stellenbosch University http://scholar.sun.ac.za

based rendering process. Interactive frame rates were achieved with the initial implementation and little optimization was ever required.

The other major drawback of a NURBS based facial model was indicated as being the inability to achieve local refinement of a NURBS surface. During the course of this research, many efforts were made at modeling a facial model using a single NURBS surface and none delivered realistic results. Several single surface facial models do exist, but most are either caricatures or suffer from noticeable side-effects. Whilst the use of hierarchically defined spline surfaces overcomes this problem to a large extent, the current lack of support for modeling, using popular, established, three-dimensional modeling packages, and rendering of such surfaces, in the author's opinion, detracts from their usability.

The approach of aligning multiple independent NURBS surfaces was encountered and immediately identified as a viable approach to the static modeling of facial models. Whether or not an actuation model could be successfully implemented within the context of such an approach was however questioned. The obvious problem was maintaining the alignment of the individual surfaces during actuation. Some preliminary attempts were made at defining and implementing a mechanism which would guarantee alignment during actuation but none of these could be scaled up to a fully generic solution for surfaces of arbitrary refinement.

Despite this, a prototype was developed to investigate the feasibility of the approach in the absence of a rigorous solution to the surface alignment problem. The initial investigations were startling in that it seemed possible to maintain visual perception of surface continuity with little effort. In the large majority of cases if muscle activation sets did not end directly on the boundaries of adjacent surfaces then alignment was maintained to a satisfactory degree. In the exceptional cases that boundary control points of one adjacent surface were in the activation set whilst the boundary control points of the other adjacent surface were not, the problem could be fairly simply resolved by extending the activation set to include additional control points. There were however several cases in which surface continuity could be broken and this occurred mostly around regions where more than two surfaces are adjacent.

Taking the above into account one can reduce the incidence of surface discontinuities by designing facial models in such a manner that surfaces overlap to a certain degree instead of having them perfectly aligned since this will reduce the occurrence of the boundary control point problem. Results could also be improved by designing the topology such that areas where

Stellenbosch University http://scholar.sun.ac.za

more than two surfaces are adjacent occur in largely static areas of the face.

5.2 Muscle Models

The muscle models introduced by WATERS [47] in 1987 have been reimplemented in several facial animation systems [8, 45]. This implies that the models have met with much success. The modeling of the major facial muscles such as the zygomaticus major (linear muscle) and the frontalis (sheet muscle) using these models has indeed been successful. In fact most true linear and sheet muscles can be modeled effectively using the original models. The human face however has an incredibly intricate structure and it is often the interaction of multiple muscles that we perceive as facial movements or features. These detailed interactions can not be effectively modeled using gross macro models. The ideal examples of this are the eyelids and mouth. In these regions several muscles interact to achieve singular results.

It is notable that these are also the areas where sphincter muscles exist. Furthermore, these are areas where muscles interact not only with other muscles but also with other structures such as the skull, eyeballs and teeth. To achieve realistic results these interactions must form part of the model. Unfortunately without implementing a physically-based actuation model this is nearly impossible to achieve.

Specific problems with Waters's sphincter models relate to the fact that the model is defined as contracting all points in the activation set towards a single point, i.e towards a singularity. Whilst this is not in itself invalid, it does not truly represent the behaviour of human facial tissue. Factors such as volume preservation and the compressibility limits of facial tissue imply that the behaviour of the model fails to accurately capture the realistic behaviour. The implementation of the sphincter muscle in the SASL-MT Facial Animation System attempts to take the above issues into account by defining a fall-off function which prevents the formation of a contraction singularity. Whilst the results achieved indicate an improvement of the original model, they still do not represent an ideal solution. If the contraction limits of sphincter muscles are not carefully controlled, wrinkling of the facial surface can occur.

The addition of eyelid muscles is seen as a significant enhancement of the original muscle model introduced by WATERS. Many facial animation systems simply ignore the need to simulate eyelid muscles, or choose to do so in a rudimentary manner. Given that the eyes and surrounding regions of the face are often the focal point of human interaction, it is important

Stellenbosch University http://scholar.sun.ac.za

that they be modeled realistically. The models implemented in the SASL-MT Facial Animation system deliver satisfactory results. The major movements of the eyelids are realistic, but close inspection during actuation has revealed that in certain circumstances, depending largely on the facial model design, overlapping of surfaces can occur near the corners of the eyes or when the eyelids are closed. The ideal solution to the problem of eyelid actuation would include a tissue model able to support the movement of facial tissue over static structures.

5.3 Facial Animation using VRML

5.3.1 VRML and Java

The decision to use VRML as the underlying platform upon which all visualization tasks would be implemented was primarily motivated by the need to maximize the accessibility of the SASL-MT Facial Animation system. Whilst this decision was sound in the context of accessibility and portability, it introduced major obstacles in terms of performance and ease of implementation. Computer facial animation is a high-performance domain and for any system to succeed within this domain it must perform at interactive rates.

Because VRML alone is not powerful enough to satisfy the requirements of a facial animation system of the scale of the SASL-MT Facial Animation System, its functionality had to be extended using the Java External Authoring Interface (EAI). This approach meant that all functional requirements of the SASL-MT Facial Animation System could be satisfied using VRML, but it also introduced certain performance restrictions. The Blaxxun Contact VRML browser used by the SASL-MT Facial Animation System (and most other VRML browsers as well) has been implemented as an ActiveX [22] component. The SASL-MT Facial Animation System implemented in Java must therefore be connected to this ActiveX component via the VRML Java EAI, and both of these are embedded objects in an HTML page.

The implication of this architecture is that data transfer between the components is not optimal. Given that the actuation of a facial model represents the modification of what could be a relatively large data set (the activation sets of all muscles involved) the bottle-neck between the VRML browser and the Java applet represents a significant problem. This bottle-neck influenced the design and implementation of the animation model and meant that the desire for a true real-time facial animation system could not be satisfied. The adoption of a low-

Stellenbosch University http://scholar.sun.ac.za

level three-dimensional graphics API such as OpenGl [24] or DirectX [23] would mean that the bottle-neck described above could, to a large extent, be avoided.

5.3.2 The Future of VRML

During the development of the SASL-MT Facial Animation System, the new standard for three-dimensional computer graphics on the internet known as X3D [52] came to the fore. The fact that X3D represents an evolution of VRML, combined with the fact that X3D compatible browsers are not yet commonplace has meant that the decision to use VRML instead of X3D seems reasonable. The process of making the SASL-MT Facial Animation System X3D compatible is seen as being achievable in the future.

5.4 Sign Language Visualization

The use of the SASL-MT Facial Animation System as a component of an SL visualization system has imposed certain unique requirements on the system. NMS in Sign Language extend beyond the universal emotional expressions such as happiness, sadness and disgust. As discussed in section 1.2, NMS have significant lexical, morphological and syntactical importance in SL and as such the facial animation component of an SL visualization system should be capable of all representing all possible NMS.

This goal was not achieved by the SASL-MT Facial Animation system. No model of the tongue was included; non-muscle based facial motions such as puffing of cheeks were not implemented and motion of the jaw is not supported. Whilst these represent serious shortcomings of the system, the design is such that additional muscle models could easily be developed to support the unsatisfied requirements described above. Furthermore, a large set of facial expressions can be easily and realistically generated.

The intricacies of movements of the mouth have also not been captured sufficiently. To animate speech as well as certain SL NMS involving the mouth a more comprehensive solution is required. The basic muscle models do not support the refined control needed to accurately simulate the motions of the lips and mouth.

Stellenbosch University http://scholar.sun.ac.za

5.5 Facial Expression Examples

See figures 5.1, 5.2, 5.3, 5.4 and 5.5.

5.6 Results Summary

The use of aligned NURBS surfaces to represent human facial models in computer facial animation were shown to represent a simple and satisfactory alternative to hierarchical splines, although a mechanism, either heuristic or mathematical, to strictly enforce adjacent surface alignment will improve the results. Furthermore, whilst, on the whole, the basic set of muscle models developed by Waters produced satisfactory results, a much richer set of muscle models is required to achieve believable, realistic facial animation.

Finally, with regards to SL visualization it was shown that, due to the importance of NMS in SL, a computer facial animation system is a definite requirement for any system attempting to visualize SL.

Stellenbosch University http://scholar.sun.ac.za

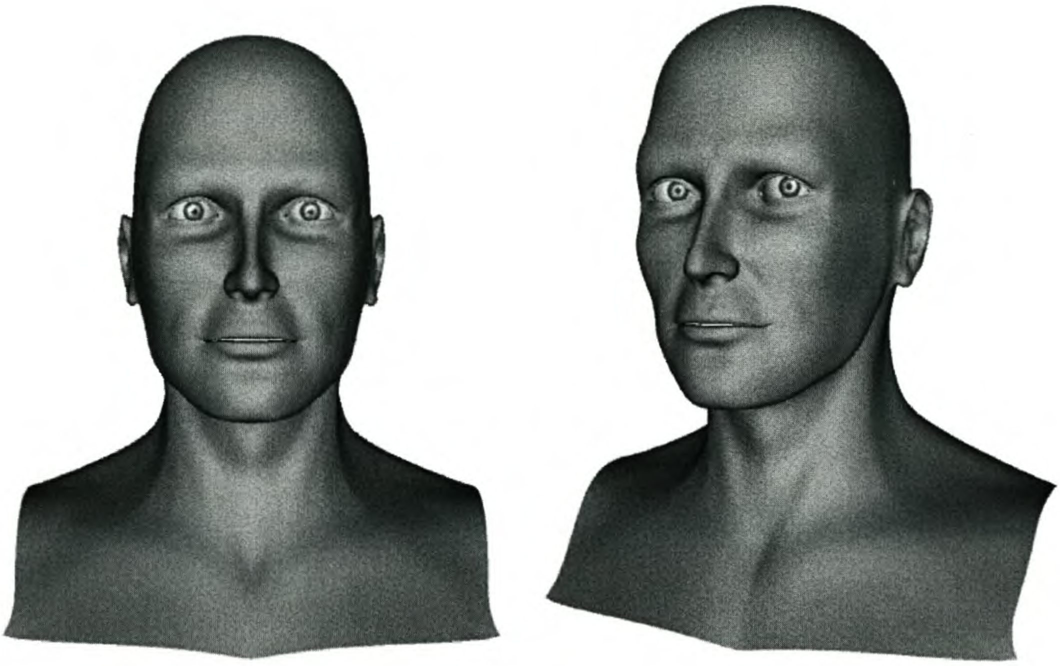


Figure 5.1: The neutral Adam facial model.

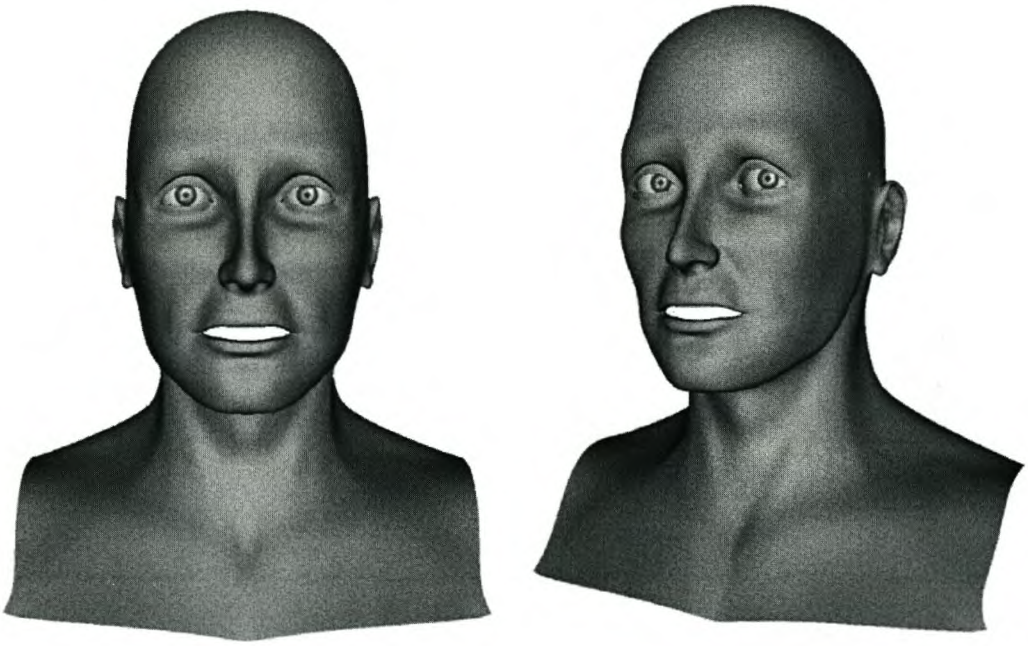


Figure 5.2: The Adam facial model displaying a surprised facial expression.

Stellenbosch University http://scholar.sun.ac.za

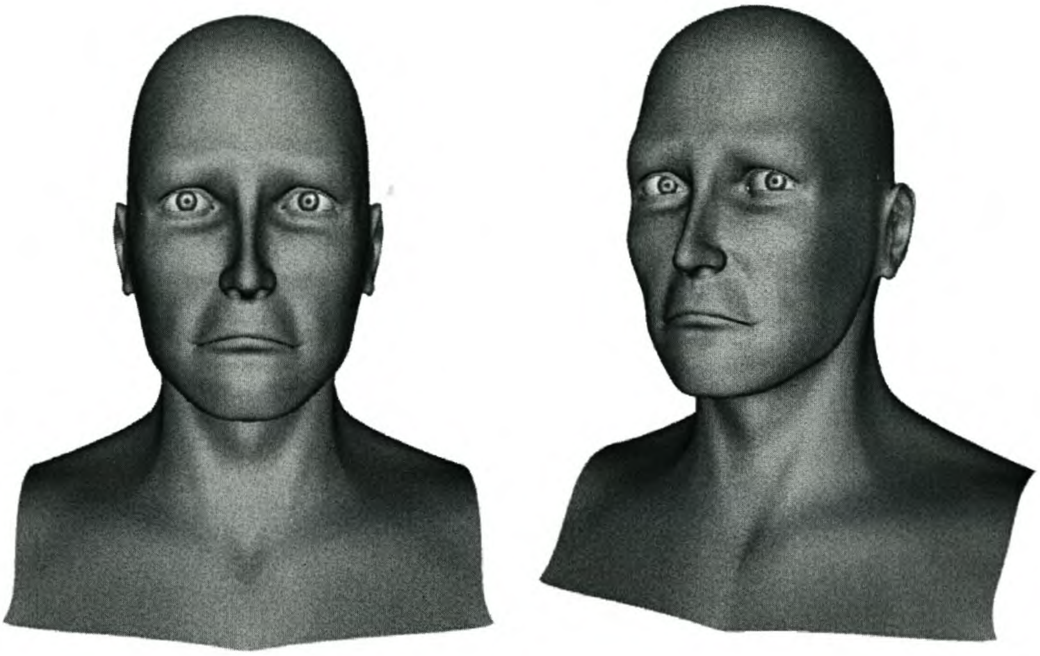


Figure 5.3: The Adam facial model displaying a sad facial expression.

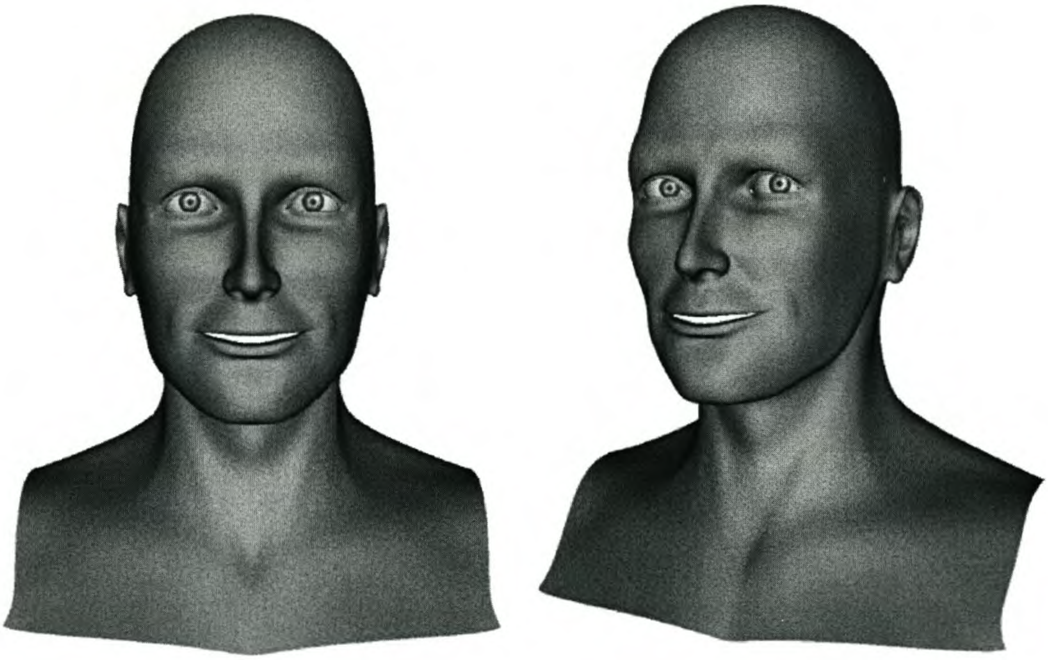


Figure 5.4: The Adam facial model smiling.

Stellenbosch University http://scholar.sun.ac.za

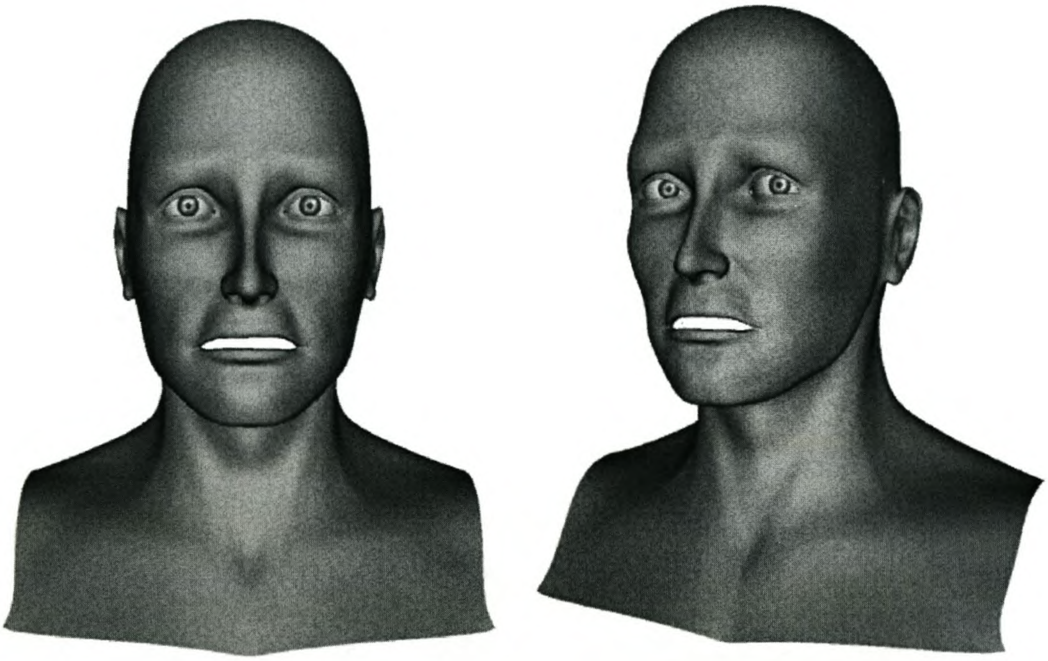


Figure 5.5: The Adam facial model displaying a fearful expression.

Stellenbosch University http://scholar.sun.ac.za

CHAPTER 6

Conclusions and Future Work

A successful implementation of a computer facial animation system, intended to function as a component of the SASL-MT visualization system, has been achieved. In section 1.3.1, the design constraints placed on the facial animation component of the SASL-MT system were described as:

1. The facial animation component must be able to manipulate a three-dimensional model of a human face in such a manner that localized distortions of the model result in the *recognizable* and *realistic* display of human facial expressions. In this manner the system must be capable of representing a well defined finite set of human *emotions*.
2. The facial animation component must satisfy certain *soft* real-time deadlines due to its role in the SASL-MT System (facial expressions and other body movements must be displayed in a coordinated, synchronous manner).
3. In an effort to make the system widely applicable, the facial animation component must be capable of operating on arbitrary NURBS based three-dimensional facial models.

The remainder of this chapter will describe what conclusions can be drawn from this implementation of a facial animation system and will indicate several possibilities for future research in the area.

6.1 Surface and Structural Models

The SASL-MT Facial Animation System is capable of operating on arbitrary NURBS based facial models. The facial model may consist of one or more aligned NURBS surfaces. Although the results achieved in this implementation have been satisfactory, detailed analysis of facial models during actuation has revealed that on occasion, the surface alignment deteriorates to

Stellenbosch University http://scholar.sun.ac.za

such an extent that a mechanism of enforcing surface alignment is warranted and should be considered as a possible future enhancement.

The use of a NURBS based facial model along with the approach of aligning surfaces of varying refinement to form a single continuous surface has facilitated the application of the SASL-MT Facial Animation System to highly realistic facial models. In the context of the SASL-MT project, this has meant that we can strive toward realistic character animation, thereby enriching the user experience (an invaluable side-effect in an SL visualization system).

6.2 Actuation Model

The NURBS based facial model is actuated using a muscle model process. Four muscle models have been implemented all of which delivered satisfactory results. The sphincter muscle model can cause unwanted surface wrinkling if the contraction limits are not carefully controlled, and as such a solution to the problem should be sought.

To achieve realistic facial animation of all FACS action units a richer set of muscle models are required. Abstract models are required for the non-muscular facial movements such as the puffing of the cheeks. A model for the tongue and teeth are also required. Models for jaw and head rotations and tilts are required as well. A significant problem around the mouth relates to collision detection for the lips. Currently, it is possible for the surfaces of the lips to intersect and overlap. Finally, a significant improvement in realism can be achieved by including a model for the underlying skull and defining interactions between the skull and the facial muscles.

6.3 Expression Model

The SASL-MT Facial Animation System utilizes an expression model based on the FACS which provides an intuitive and well-defined interface to the muscle model process. The expression model allows one to define facial animation sequences in terms of facial expressions which are in turn defined in terms of abstract percentage AU contractions. The approach facilitates the reuse of facial animation sequences and expressions across different facial models.

The current implementation enforces a one-to-one mapping between FACS AU's and muscles. By extending this relationship to allow the association of multiple muscles with a single AU the configuration effort could feasibly be reduced.

Stellenbosch University http://scholar.sun.ac.za

6.4 User Interfaces

A comprehensive, flexible and extensible user interface has been developed for the SASL-MT Facial Animation System. The user interface supports the creation and editing of muscles and facial expressions and facial animation sequences. Possible improvements in the user interface should focus on the muscle creation process. Currently the process of creating muscles is largely automated, but in certain circumstances muscle activation sets need to be manually edited. This process of manual activation set editing can become cumbersome in areas of high complexity such as the corners of the mouth and the eyes. A mechanism for efficiently specifying control points for inclusion/exclusion in/from the activation set would greatly reduce the configuration effort.

Furthermore, facial animation sequences can currently only be specified on a single time-line. The author is of the opinion that by supporting the notion of parallel time-lines, the specification of an animation sequence can be made more intuitive (for example, certain facial motions can be grouped in different concurrent time-lines so that motions of say, the eyes, can be specified on a separate time line to say, motions of the mouth).

6.5 Summary of Conclusions & Future Work

Many enhancements can be made to improve the functionality, performance, visual realism and usability of the SASL-MT Facial Animation System. The purpose of the project, namely, to create a foundation for the facial animation component of the SASL-MT visualization system, was fulfilled.

Stellenbosch University http://scholar.sun.ac.za

Bibliography

[1] Alias. Maya. `http://www.alias.com/`.

[2] Charlotte Baker and Carol Padden. Focusing on the Nonmanual Components of American Sign Language. *Understanding Language Through Sign Language Research - Perspectives in Neorolinguistics and Psycholinguistics*, pages 27–57, 1978.

[3] R. Bartles, J. Beatty, and B. Barsky. *Introduction to Splines for Use in Computer Graphics and Geometric Modeling.* Morgan Kaufmann, 1987.

[4] Michael D. Bondy, Marius D. Cordea, Nicolas D. Georganas, Dorina C. Petriu, Thomas E. Whalen, and Emil M. Petriu. Model-Based Face and Lip Animation for Interactive Virtual Reality Applications. In *Proceedings of the 9th ACM International Conference on Multimedia*, October 2001.

[5] Discreet. 3D Studio Max. `http://www.discreet.com`.

[6] Paul Ekman and Wallace V. Friesen. Background of the Facial Action Coding System. *Journal of Environmental Psychology and Nonverbal Behavior*, 1(1):56–75, 1976.

[7] Paul Ekman, Thomas S. Huang, Terrence J. Sejnowski, and Joseph C. Hager. Final Report to NSF of the Planning Workshop on Facial Expression Understanding. In *Planning Workshop on Facial Expression Understanding*, 1992. `http://face-and-emotion.com/dataface/nsfrept/nsf_contents.html`.

[8] Fatih Erol. Modeling and Animating Personalized Faces. Master's thesis, Bilkent University, Turkey, 2002. `http://vrlab.epfl.ch/~ferol/mscthesis.pdf`.

[9] Louie J. Fant. *The ASL Phrase Book.* Contemporary Books, 1994.

Stellenbosch University http://scholar.sun.ac.za

[10] Bill Fleming and Darris Dobbs. *Animating Facial Features and Expressions.* Charles River Media, Inc., Rockland, Massachusetts, 1999.

[11] Holger Grahn, Thomas Volk, and Hans Wolters. NURBS in VRML. In *Proceedings of the 5th symposium on Virtual Reality Modeling Language*, pages 35–43, Monterey, California, United States, 2000.

[12] C.H. Hjortsjo. *Man's Face and Mimic Language.* Lund, Sweden, 1970.

[13] Michael Hoch, Bernd Girod, and Georg Fleischmann. Modeling and Animation of Facial Expressions based on B-Splines. *The Visual Computer*, 11(2):87–95, 1994.

[14] H. S. Horace and C. S. Chan. Script-based Facial Gesture and Speech Animation Using a NUBS Based Face Model. *Computers and Graphics*, 20(6):881–891, November 1996.

[15] Blaxxun Interactive. `http://www.blaxxun.com`.

[16] Ruth T. Jackson, Rebecca Orton, and Peter Un. A Comparison of Non-Manual Signals Among Hearing, Hard of Hearing and Deaf Signers. `http://www.geocities.com/rebaorton/rough.htm`.

[17] Prem Kalra, Nadia Magnenat-Thalmann, Angelo Mangili, and Daniel Thalmann. Simulation of Facial Muscle Actions Based on Rational Free Form Deformations. In *Proceedings Eurographics '92*, pages 59–69, 1992.

[18] Prem Kalra, Angelo Mangili, Nadia Magnenat-Thalmann, and Daniel Thalmann. 3D Interactive Free Form Deformations for Facial Expressions. Technical report, MIRALab – University of Geneva, Computer Graphics Lab – Swiss Federal Institute of Technology Lausanne, Switzerland, 1998. `http://vrlab.epfl.ch/Publications/pdf/Kalra_and_al_Compugraphics_91.pdf`.

[19] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Constructing physics-based facial models of individuals. In *Proceedings of Graphics Interface '93*, pages 1–8, Toronto, Ontario, Canada, 1993. `http://citeseer.ist.psu.edu/lee93constructing.html`.

[20] Yuencheng Lee, Keith Waters, and Demetri Terzopolous. Realistic Modeling for Facial Animation. *ACM SIGGRAPH Computer Graphics*, pages 55–62, 1995.

Stellenbosch University http://scholar.sun.ac.za

[21] Nadia Magnenat-Thalmann and Daniel Thalmann. The Direction of Synthetic Actors in the Film Rendez-vous a Montreal. *IEEE Computer Graphics and Applications*, 7(12):9–19, 1987.

[22] Microsoft. ActiveX. `http://msdn.microsoft.com/workshop/components/activex/activex_node_entry.asp`.

[23] Microsoft. DirectX. `http://www.microsoft.com/windows/directx/`.

[24] OpenGL. `http://www.opengl.org`.

[25] Frediric I. Parke. *A Parametric Model for Human Faces.* PhD thesis, University of Utah, 1972.

[26] Frediric I. Parke. Parameterized Models for Facial Animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, 1982.

[27] Frediric I. Parke and Keith Waters. *Computer Facial Animation.* A K Peters, Ltd., 1996.

[28] Catherine Pelachaud, Norman I. Badler, and Mark Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1):1–46, 1996. `http://citeseer.ist.psu.edu/pelachaud94generating.html`.

[29] Frederic Pighin, Joel Auslander, Dani Lischinski, David H. Salesin, and Richard Szeliski. Realistic Facial Animation Using Image-Based 3D Morphing. Technical Report UW-CSE-97-01-03, University of Washington, 1997. `http://ftp.cs.washington.edu/tr/1997/01/UW-CSE-97-01-03.PS.Z`.

[30] Frederic Pighin, Jamie Hecker, Dani Lischinski, David H. Salesin, and Richard Szeliski. Synthesizing Realistic Facial Expressions from Photographs. Technical report, University of Washington, 1998. `http://smg.www.media.mit.edu/classes/VirtualSociety99/temp/p75-pighin.pdf`.

[31] Frederic Pighin, David H. Salesin, and Richard Szeliski. Modeling and Animating Realistic Faces from Images. *International Journal of Computer Vision*, 50(2), November 2002.

[32] Stephen Platt and Norman Badler. Animating Facial Expressions. *ACM SIGGRAPH Computer Graphics*, 15(3):245–252, August 1981.

Stellenbosch University http://scholar.sun.ac.za

[33] Stephen .M. Platt. A System for Computer Simulation of the Human Face. Master's thesis, The Moore School, University of Pennsylvania, 1980.

[34] Stephen M. Platt. *A Structural Model of the Human Face.* PhD thesis, University of Pennsylvania, 1985.

[35] Stephen M. Platt, Aaron Smith, Francisco Azuola, Catherine Pelachaud, and Norman Badler. Structure-Based Animation of the Human Face. Technical Report MS-CIS-91-15, University of Pennsylvania, Department of Computer and Information Science, 1991. `http://citeseer.ist.psu.edu/platt91structurebased.html`.

[36] Thomas W. Sederberg and Scott R. Parry. Free-Form Deformation of Solid Geometric Models. *ACM SIGGRAPH Computer Graphics*, 20(4):151–160, 1986.

[37] Demetri Terzopolous and Keith Waters. Physically-based Facial Modelling, Analysis, and Animation. *The Journal of Visualization and Computer Animation*, 1(1):73–80, 1990.

[38] Demetri Terzopolous and Keith Waters. Techniques for Realistic Facial Modeling and Animation. *Computer Animation '91*, pages 59–74, 1991.

[39] Demetri Terzopoulos and Hong Qin. Dynamic NURBS with Geometric Constraints for Interactive Sculpting. *ACM Transactions on Graphics*, 13(2):103–136, April 1994.

[40] D. Thalmann, E. Primeau, and N. Magnenat-Thalmann. Abstract Muscle Action Procedures for Human Face Animation. *The Visual Computer*, 3(5):290–297, 1988.

[41] F. Thomas and O. Johnson. *Disney Animation: The Illusion of Life.* Abbeville Press, New York, 1981.

[42] L. van Zijl and D. Barker. A Machine Translation System for South African Sign Language. In *Proceedings Afrigraph 2003*, ACM SIGGRAPH Conference Proceedings Series, 2003, Cape Town, South Africa, February 2003.

[43] Virtual Reality Modeling Language. `http://www.web3d.org/`.

[44] C.T. Waite. The Facial Action Control Editor, FACE: A parametric facial expression editor for computer generated animation. Master's thesis, Massachusetts Institute of Technology, Media Arts and Sciences, 1989.

Stellenbosch University http://scholar.sun.ac.za

[45] Carol Leon-Yun Wang. Langwidere: A Hierarchical Spline Based Facial Animation System. Master's thesis, University of Calgary, 1993. `http://www.arghc.com/docs/thesis.ps.gz`.

[46] Carol Leon-Yun Wang and David R. Forsey. Langwidere: A New Facial Animation System. Technical Report TR-94-34, University of Calgary, 1994. `http://www.cs.ubc.ca/cgi-bin/tr/1994/TR-94-34.pdf`.

[47] Keith Waters. A Muscle Model for Animating Three-Dimensional Facial Expressions. *ACM SIGGRAPH Computer Graphics*, 21(4), July 1987.

[48] Keith Waters and Demetri Terzopolous. Modeling and Animating Faces using Scanned Data. *The Journal of Visualization and Computer Animation*, 2(4):123–128, 1991.

[49] Alan Watt. *3D Computer Graphics*. Addison-Wesley, third edition, 2000.

[50] Jeffrey I. Wilson. `http://www.zoorender.com/freetrade/jeff_ian/zoo_adam.zip`.

[51] J.R. Woodwark. Blends in Geometric Modeling. In *Proceedings 2nd IMA Conference on the Mathematics of Surfaces*, pages 255–297, Cardiff, September 1986.

[52] X3D. `http://www.web3d.org/`.

[53] Liwei Zhao, Karin Kipper, William Schuler, Christian Vogler, Norman Badler, and Martha Palmer. A Machine Translation System from English to American Sign Language. In *Proceedings of the 4th Conference of the Association for Machine Translation in the Americas*, pages 54–67, Cuernavaca, Mexico, 2000.

Stellenbosch University http://scholar.sun.ac.za

APPENDIX A

Non-Uniform Rational Basis-Splines

The information contained in this section is a summary taken from [11]. A point on a NURBS surface is defined by

$$Q(u,v) = \frac{\sum_{i=0}^{m_u}\sum_{j=0}^{m_v} B_{i,k_u}(u)B_{j,k_v}(v)w_{i,j}V_{i,j}}{\sum_{i=0}^{m_u}\sum_{j=0}^{m_v} B_{i,k_u}(u)B_{j,k_v}(v)w_{i,j}}$$

where

u, v	are the parameters of the surface,
B_i	are the basis functions,
k_u	is the order in the u direction,
k_v	is the order in the v direction,
V	is the mesh of control points, and
$w_{i,j}$	are the weights.

The basis functions are defined as follows:

$$\begin{aligned} B_{i,1} &= 1 \text{ for } u_i \le u < u_{i+1} \text{ and } 0 \text{ otherwise}, \\ B_{i,r}(u) &= \frac{u - u_i}{u_{i+r-1} - u_i} B_{i,r-1}(u) + \frac{u_{i+r} - u}{u_{i+r} - u_{i+1}} B_{i+1,r-1}(u), \\ U &= \{u_0, \ldots, u_{m_u}\}, \end{aligned}$$

where U is the knot vector containing a nondecreasing sequence of real numbers.

Surface normals can be calculated by taking the cross product of the surface derivatives at a point:

$$n = \frac{\partial}{\partial u} Q(u,v) \times \frac{\partial}{\partial v} Q(u,v).$$

By stepping through the u and v domains and evaluating the equation for points on the surface, a grid of sample points can be produced. Triangle strips can be generated by stepping through the u domain at two fixed v values. This method of evaluating the surface in terms

Stellenbosch University http://scholar.sun.ac.za

of triangles is known as tessellation. Dynamic tessellation is achieved by varying the step size through the domains. The practical implications of this are that rendered NURBS surfaces can have varying levels of detail, without increasing or decreasing the number of control points used to specify the surface.

Stellenbosch University http://scholar.sun.ac.za

APPENDIX B

Facial Anatomy

The information contained in this section is a summary taken from [27]. To fully understand, appreciate, and motivate the different approaches to computer facial animation, some basic knowledge of human facial anatomy is required. This chapter will therefore be devoted to delivering a broad overview of facial anatomy.

The anatomy of the head and face is a complex assembly of bones, cartilage, muscle, nerves, blood vessels, glands, fatty tissue, connective tissue, skin and hair. For the purposes of three dimensional computer modeling of the human head, three distinct primary components can be identified: the skull and underlying bone structure; the facial musculature; and the skin. Each of these will be discussed in turn, with particular focus given to the facial muscles.

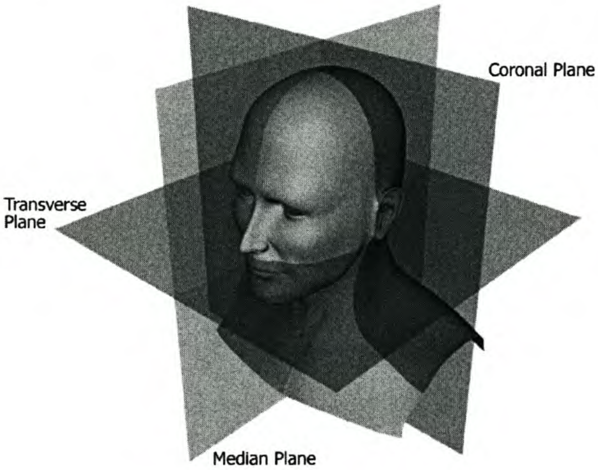


Figure B.1: Anatomical reference planes.

When discussing human anatomy the use of a specific, unambiguous vocabulary is required. In general, the location of body parts are described relative to three imaginary planes:

- **Median Plane.** This vertical plane bisects the body into equal right and left halves (see

Stellenbosch University http://scholar.sun.ac.za

figure B.1 on page 70). A structure is said to be *medial* (or *lateral*) to another, if that structure is nearer (or further) to (or from) the median plane than the other. For example, the eye is medial to the ear, but lateral to the nostril.

- **Coronal Plane.** This vertical plane, perpendicular to the median plane, divides the body into front (*anterior*) and back (*posterior*) halves (see figure B.1 on page 70). A structure is said to be *anterior* (or *posterior*) to another when it is nearer to the *anterior* (or *posterior*) surface of the body. For example, the eyes are anterior to the ears, but posterior to the tip of the nose.
- **Transverse Plane.** This horizontal plane, perpendicular to both the median and coronal planes, divides the body into upper (*superior*) and lower (*inferior*) portions (see figure B.1 on page 70). *Superior* and *inferior* refer to relative positions with respect to the upper and lower body. For example, the lips are superior to the chin, but inferior to the eyes.

For the remainder of this document, it will be assumed that the reader is familiar with the vocabulary introduced above.

B.1 The Skull

The human skull has a dual role: to serve as a protective casing for the brain; and to provide the structural foundation for the face. Accordingly, the skull can be separated into two divisions: the *cranium*, which encloses and protects the brain; and the *facial skeleton*, which provides the structural basis for the face. Each of these divisions are aggregates of several other bone structures.

B.1.1 The Cranium

The cranium can be separated into two divisions: the calvaria and the cranial base. The calvaria consists of the frontal, parietal and occipital bones (see figure B.2). The frontal bone forms the forehead, brow ridges and superior portions of the eye sockets. The paired parietal bones form the majority of the cranial roof and sides. The occipital bone forms the posterior base of the cranium, and articulates with the atlas, the first cervical vertebra.

The cranial base consists of the temporal and sphenoid bones (see figure B.2). The paired temporal bones form the lateral sides of the cranium. The temporal bone houses the external

Stellenbosch University http://scholar.sun.ac.za

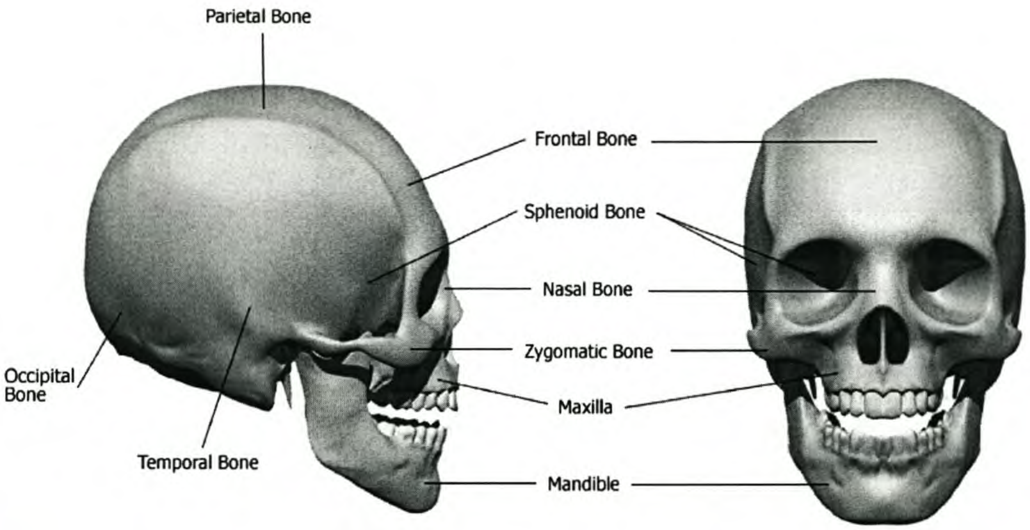


Figure B.2: Frontal and lateral views of the skull, from [10].

auditory meatus (the opening to the middle ear), and provides an articulate region for the mandible (the jaw bone). The sphenoid bone has a central body, forming the anterior base of the cranium, as well as paired greater and lesser wings spreading laterally to form portions of the posterior walls of the eye socket and the anterolateral walls of the cranium.

B.1.2 The Facial Skeleton

The facial skeleton is of particular interest in three dimensional facial modeling since it provides the foundation for the facial skin and muscles. The bones constituting the facial skeleton are the ethmoid bone, the palatine bone, the maxillae, the inferior nasal concha, the zygomatic bones, the nasal bones, the lacrimal bone, the mandible, the hyoid bone and the vomer (see figure B.2 on page 72). Not all these bones are of equal significance in the context of facial modeling, and therefore only those deemed worthy of discussion will be considered below.

The palatine bone is an irregularly shaped bone which forms the posterior part of the hard palate within the mouth. This bone also has several substructures which contribute to the internal complexity of the nasal cavity. The maxillae, which are the second largest bones of the facial skeleton, are located in the lower portion of the skull and house the upper teeth. The *frontal process* of the maxilla forms the lower, inner rim of the eye socket. The *zygomatic process* in conjunction with the zygomatic bone, form the prominence of the cheek. The *palatine*

Stellenbosch University http://scholar.sun.ac.za

process forms the majority of the hard palate and the *alveolar process* houses the the teeth.

The zygomatic bone forms part of the lateral wall of the eye socket, and in conjunction with the zygomatic process of the maxilla forms the prominence of the cheek. The nasal bones form the superior foundation of the nose. The mandible, the largest facial bone, and only freely jointed structure in the face, forms the jaw and houses the lower teeth. The horizontal portion of the mandible is known as the *mandibular body* and the lateral vertical sections are known as the *rami*.

B.2 The Facial Muscles

Muscles are the organs responsible for bringing about the motions of the human body. Muscles are aggregate structures consisting of bundles of muscle fibers. Muscle fibers also have a finer microscopic substructure composed of threads called myofibrils that run the length of the fiber. Along the longitudinal axis of each myofibril is a repeating pattern of filaments called sarcomeres. The sarcomere is the actual functional unit muscle contraction.

In general, muscles are secured between two moving parts such as two bones, bone and skin, two different areas of skin, or two organs. Muscles are attached to other structures by tendons, aponeuroses and fascia, and it is these connective tissues which determine the direction of the force applied by a muscle.

The contraction of a muscle is an active process and occurs due to the electrical stimulation, via a nerve, of the muscle. The macroscopic contraction is realized by the microscopic shortening and thickening of the individual muscle fibers. This is in turn realized via the shortening of the sacromeres along the longitudinal axis of the myofibrils. The relaxation of a muscle is a passive process and occurs due to a cessation or lack of stimulation.

When a muscle is suspended between two points, one fixed and one moveable, the attachment of the muscle to the fixed point is known as the *origin*, and that to the moveable point, the *insertion*. In general, the area most affected by a muscle's contraction is the insertion. The muscles responsible for facial expressions in the human are all superficial muscles and all have their insertion in a subcutaneous layer of fat and skin. In general the origins of the facial muscles are on the facial bones, but some have both their origin and insertion in skin.

The muscles of facial expression form a complex, interweaved network across the face. These muscles work in well coordinated groups, and discerning individual muscle boundaries is dif-

Stellenbosch University http://scholar.sun.ac.za

ficult. If we try to classify muscles in terms of muscle location and focus, two classes can be discerned; the muscles of the upper and lower face. The upper facial muscles are responsible for manipulating the eyebrows, forehead and eyelids. The lower facial muscles are primarily responsible for the manipulation of the mouth, lips and surrounding regions. The muscles of the lower face have an extremely complex interaction.

Attempting to classify the facial muscles according to their actions results in five classes:

- *Vertical contractors* displace portions of the face roughly perpendicular to the transverse plane (either up, towards the brow, or down, towards the chin).
- *Horizontal contractors* displace portions of the face roughly perpendicular to the medial plane (either medially or laterally).
- *Oblique contractors* displace portions of the face in angular directions relative to the reference planes.
- *Orbital contractors* displace all the portions of the face surrounding a point towards that point.
- *Sheet contractors* displace wide areas of the face.

Another means of classifying facial muscles is to consider the orientation of the muscle fibers which constitute the muscle. In this manner, three types of facial muscles can be discerned:

- *Sheet*: A quadrilateral muscle in which the muscle fibers run roughly parallel to each other. Sheet muscles generally have wide areas of attachment and insertion.
- *Linear/Triangular*: A fan shaped muscle in which the muscle fibers tend to converge to a point.
- *Elliptical/Sphincter*: An elliptical muscle in which the muscle fibers are arranged in curved bundles.

The remainder of this section will be devoted to listing the muscles involved with facial expression. The muscles will be listed and grouped according to the areas of the face affected by those muscles.

B.2.1 Muscles of the Eye

The primary muscles responsible for motion of the eyelids and surrounding regions are: the orbicularis oris, corrugator supercili and levator palpebrae superioris (see figure B.3). See APPENDIX C for more detailed information relating to each of these muscles.

Stellenbosch University http://scholar.sun.ac.za

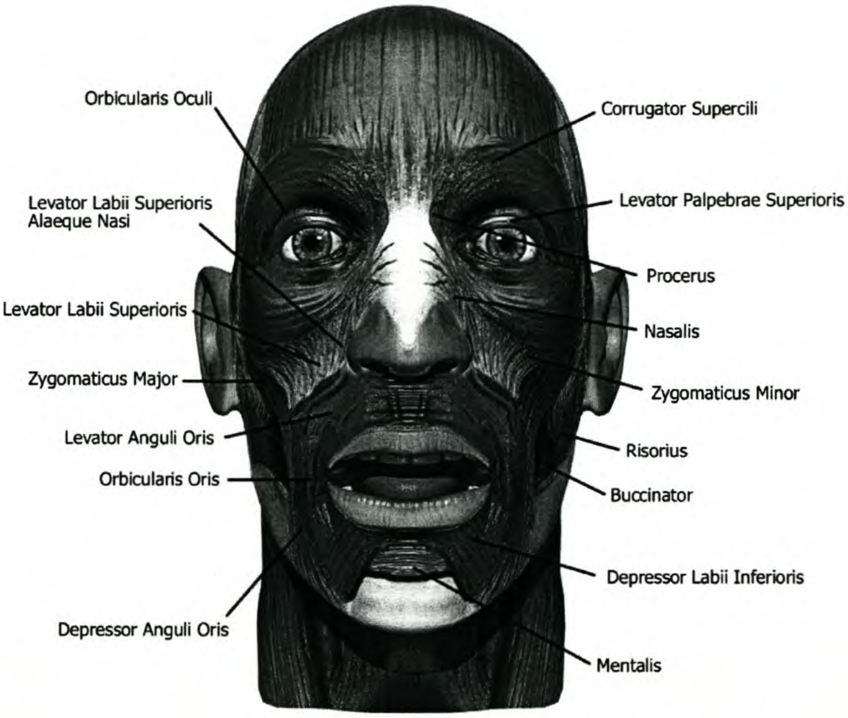


Figure B.3: The muscles of facial expression, from [10].

B.2.2 Muscles of the Nose

The primary muscles responsible for motion of the nose and surrounding regions are: the procerus, nasalis, depressor septi and levator labii superioris alaeque nasi (see figure B.3). See APPENDIX C for more detailed information relating to each of these muscles.

B.2.3 Muscles of the Mouth

The primary muscles responsible for motion of the mouth, lips and surrounding regions are: the orbicularis oris, buccinator, levator labii superioris alaeque nasi, levator labii superioris, zygomaticus major, zygomaticus minor, levator anguli oris, depressor anguli oris, depressor labii inferioris, risorius and mentalis (see figure B.3). See APPENDIX C for more detailed information relating to each of these muscles.

B.2.4 Muscles of the Mandible

Four basic movements of the mandible can be identified. These movements are brought about by the complex interactions of several muscle all attached to the mandible.

Stellenbosch University http://scholar.sun.ac.za

- *Protraction* is the act of pulling the mandible forward. The muscles responsible for this action are the lateral and medial pterygoid muscles along with the masseter.
- *Retraction* is the act of pulling the mandible backward and involves the digastric and the geniohyoid muscles.
- *Elevation* is the act of closing the mouth. The muscles involved with this action are the masseter, the medial pterygoid and the temporalis.
- *Depression* is the act of opening the mouth and involves the digastric, the geniohyoid, the mylohyoid, and the platysma muscles.

See APPENDIX C for more detailed information relating to each of these muscles.

B.2.5 Muscles of the Tongue

There are two groups of muscles associated with the tongue. The *intrinsic* muscles lie within the tongue and can be further subdivided into longitudinal, transverse and vertical groups.

The *extrinsic* muscles originate outside the tongue and attach to the intrinsic muscles. The extrinsic muscles of the tongue are the genioglossus which retracts and protrudes the tongue; the hyoglossus which depresses the sides of the tongue; and the styloglossus which retracts, and draws the tongue upwards.

B.2.6 Muscles of the Scalp

The two frontalis, and two occipital muscles constitute the muscles of the scalp. See APPENDIX C for more detailed information relating to each of this muscle.

B.2.7 Muscles of the Neck

The primary superficial muscles of the neck are the paired trapezius, sternocledomastoid and platysma muscles. See APPENDIX C for more detailed information relating to each of these muscles.

B.3 The Skin

Human skin is a complex, nonhomogenous, multilayered network covering the entire surface of the body. The three primary tissue layers constituting what is collectively called the skin are:

Stellenbosch University http://scholar.sun.ac.za

- *Epidermal Tissue* : This is the outermost layer of dead cells composed primarily of a protein called keratin. There is a continual replacement of cells in this outermost layer.
- *Dermal Tissue* : The dermis consists of a collagen network nonhomogenously interspersed with elastin fibers, blood vessels, lymphatic vessels, microscopic glands and hair follicles.
- *Subcutaneous Tissues* : This is the deepest layer, consisting primarily of fatty tissue distributed within a network of collagenous connective tissue.

The thickness of facial tissue varies across the face, being thinnest around the eyes, and thickest around the mouth and lips. Wrinkles are a prominent feature in facial tissue. Wrinkles occur partly due to the reduction in fatty tissue with age, and the loss of elasticity in the skin.

Stellenbosch University http://scholar.sun.ac.za

Appendix C

The Facial Muscles

The information contained in this section is a summary taken from [27].

Stellenbosch University http://scholar.sun.ac.za

Table C.1: Muscles of the eye.

NAME	TYPE	ORIGIN	INSERTION	ACTION
Orbicularis Oculi	Sphincter	Inner corner of the eye	NA	The orbital and palpebral parts of this muscle are able to act independently. Activation of the orbital part draws the skin of the brow and the cheek toward the inner corner of the eye. The palpebral part has fine control over the individual eyelids.
Corrugator Supercili	Linear	Medial end of superciliary arch	Posterior lateral end of superciliary arch	Activation of the muscle draws the brow inward and down and often produces vertical wrinkles between the eyebrows.
Levator Palpebrae Superioris	Linear	Within the orbit	Upper eyelid	Activation of this muscle results in the lifting of the upper eyelid.

Stellenbosch University http://scholar.sun.ac.za

NAME	TYPE	ORIGIN	INSERTION	ACTION
Procerus	Linear	Nasal Bone	Medial brow and forehead	Depresses the medial end of the eyebrow and produces horizontal wrinkles
Nasalis	Linear	Alveolar eminence over the lateral incisor	Bridge of the nose	Compresses the nasal aperture at the junction of the nostril and the nasal cavity.
Depressor Septi	Linear	On the maxilla, above the central incisor	Mobile part of the nasal septum	Assists the nasalis in widing the nasal aperture.
Levator Labii Superioris Alaeque Nasi	Linear	Upper part of frontal process of maxilla	Medial slip: Wing of the nose. Lateral slip: Orbicularis Oris, near philtrum	Raises and inverts the upper lip and deepens the nasolabial furrow. The medial slip dilates the nostril.

Table C.2: Muscles of the nose

Stellenbosch University http://scholar.sun.ac.za

NAME	TYPE	ORIGIN	INSERTION	ACTION
Orbicularis Oris	Sphincter	An aggregate muscle derived from other facial muscles that converge on the mouth	NA	Controls lip shape during speech and nonverbal communication.
Buccinator	Linear	On the opposite the first molar, and on the mandible from the pterygomandibular raphe	Lateral portions of upper and lower lips	Compresses the cheeks against the teeth.
Levator Labii Superioris Alaeque Nasi	Section C	Section C	Section C	Section C
Levator Labii Superioris	Linear	Wide attachment to the zygomatic bone and the maxilla near the inferior rim of the orbit	In the top lip, between the levator anguli oris, and the levator labii superioris alaeque nasi	Raises the upper lip, and deepens the nasolabial furrow.
Zygomaticus Major	Linear	Anterior surface of the zygomatic bone	The corner of the mouth	Elevates the lip corners superolaterally as in smiling.
Zygomaticus Minor	Linear	Anterior surface of the zygomatic bone, medial to the zygomatic magor's origin	The skin of the upper lip, lateral to the midline.	Elevates the upper lip, exposing the upper teeth, and deepens the nasolabial furrow.
Levator Anguli Oris	Linear	Canine Fossa	Corner of the mouth	Raises the corner of the mouth displaying the upper teeth and deepens the nasolabial furrow.
Risorius	Linear	Anterior border of the masseter muscle	Under the skin and mucous membrane of the upper lip near the corner of the mouth	Pulls the angle of the mouth laterally.
Mentalis	Linear	Circular area above the mental tuberosity	In the skin of the chin	Elevates the skin of the chin.
Depressor Anguli Oris	Linear	Near the insertion of the platysma	Tendonous node at the corner of the mouth	Depresses the corner of the mouth laterally.
Depressor Labii Inferioris	Linear	Slightly superior and medial to the origin of the depressor anguli oris	The skin of the lower lip	Pulls the lower lip down and laterally.

Table C.3: Muscles of the mouth

Stellenbosch University http://scholar.sun.ac.za

NAME	TYPE	ORIGIN	INSERTION	ACTION
Masseter	Sheet	Zygomatic arch & Zygoma	Lateral surface of the mandibular ramus at the angular region	Elevates and retracts the mandible.
Medial Pterygoid	Linear	Pterygoid fossa	Lateral surface of the mandibular ramus, slightly superior to the attachment of the masseter	Elevates and protracts the mandible.
Temporalis	Linear	Temporal fossa, a wide are on the lateral surface of the skull.	Anterior border of the coronoid and mandibular ramus, as well as the medial region of the lower third molar.	Rapid elevation and retraction of the mandible.
Digastric	Linear	Mastoid Notch	Greater horn of the hyoid bone	Raises the hyoid bone – an important action during swallowing.
Geniohyoid	Linear	Anterior end of the mylohyoid line	Upper half of the body of the hyoid bone	Elevates and fixes the hyoid bone. Also involved in elevating the tongue.
Lateral Pterygoid	Linear	Inferior head – lateral surface of the lateral pterygoid. Superior head – infratemporal surface of the great wing of the sphenoid bone	Anterior surface of the condylar neck	Depresses the jaw.
Mylohyoid	Linear	Mylohyoid line on the mandible	Body of the hyoid bone, and mylohyoid raphe	Primarily, it elevates the tongue, but it also influences the position of both the hyoid bone and the mandible.
Platysma	Sheet	Fibrous connective tissue under the skin of the clavicle and shoulder	Inferior border of the mandible	Raises the skin of the neck and depresses the mandible.

Table C.4: Muscles of the mandible

Stellenbosch University http://scholar.sun.ac.za

NAME	TYPE	ORIGIN	INSERTION	ACTION
Frontalis	Sheet	Galea aponeurotica	Root of the nose, and skin of the eyebrow	Lifts the eyebrow and furrows the skin of the forehead.

Table C.5: Muscles of the scalp

NAME	TYPE	ORIGIN	INSERTION	ACTION
Trapezius	Linear	External occipital protuberance	Shoulder girdle	Pull the head backward.
Sternocledo-mastoid	Linear	Posterior sternum	Mastoid process of the temporal bone	Bends the neck laterally. When both contract together, the head is pulled forward.

Table C.6: Muscles of the neck

Stellenbosch University http://scholar.sun.ac.za

Appendix D

Facial Action Coding System

The information contained in this section is a summary taken from [6]. Note that the AU numbering is arbitrary and as such the fact that certain numbers in the sequence 1 – 66 are missing is insignificant. No motivation was given for this by the creators of the FACS.

Note further that for AU's which have no muscle association in the table below either the AU has no muscular basis (for example AU33 - Cheek Blow) or the muscles related to the AU are not facial muscles (for example AU51 - Head Turn Left).

Stellenbosch University http://scholar.sun.ac.za

Action Unit	**Name**	**Associated Muscles**
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Raiser	Depressor Glabellae, Depressor Supercilli, Corrugator
5	Upper-Lid Raiser	Lavator Palpebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
8	Lips Together	Orbicularis Oris
9	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10	Upper-Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11	Nasolabial Furrow Deepener	Zygomatic Minor
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinator
15	Lip Corner Depressor	Triangularis
16	Lower-Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris, Incisivii Labii Inferiouis
19	Tongue Out	
20	Lip Stretcher	Risorius
21	Neck Tightener	
22	Lip Funneler	Orbicularis Oris
23	Lip Tightener	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
25	Parting of Lips	Depressor Labii, Mentalis, Orbicularis Oris (Relaxation)
26	Jaw Drop	Masettter, Pterygoid (Relaxtion)
27	Mouth	Stretch Pterygoids, Digastric
28	Lip Suck	Orbicularis Oris
29	Jaw Thrust	
30	Jaw Sideways	
31	Jaw Clencher	
32	Lip Bite	
33	Cheek Blow	
34	Cheek Puff	
35	Cheek Suck	
36	Tongue Bulge	
37	Lip Wipe	
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa, Depressor Septi Nasi
41	Lid Droop Levator	(Relaxation), Palpebrae Superioris
42	Eyelid Slit	Orbicularis Oculi
43	Eyes Closed	Levatro Palpebrae Superioris (Relaxation)

Stellenbosch University http://scholar.sun.ac.za

Action Unit	**Name**	**Associated Muscles**
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Levator Palpebrae (Relaxation), Orbicularis Oculi (Contract), Pars Palpebralis
46	Wink	Orbicularis Oculi
51	Head Turn Left	
52	Head Turn Right	
53	Head Up	
54	Head Down	
55	Head Tilt Left	
56	Head Tilt Right	
57	Head Forward	
58	Head Back	
61	Eyes Turn Left	
62	Eyes Turn Right	
63	Eyes Up	
64	Eyes Down	
65	Walleye	
66	Cross-eye	

Table D.1: FACS Action Units

Stellenbosch University http://scholar.sun.ac.za

APPENDIX E

SASL-MT Facial Animation System

A functional version of the SASL-MT Facial Animation System can be found on the internet at: `http://www.cs.sun.ac.za/~dbarker/SASLMTFace/index.html`. Installation and usage instructions are available on the website.