THE UNIVERSITY OF CALGARY

Langwidere: A Hierarchical Spline Based Facial Animation System with Simulated Muscles

by

Carol Leon-Yun Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

OCTOBER, 1993

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Langwidere: A Hierarchical Spline Based Facial Animation System with Simulated Muscles" submitted by Carol Leon-Yun Wang in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, David Hill,

Department of Computer Science

Brian Wyvill,

Department of Computer Science

Saul Greenberg,

Department of Computer Science

Ray Arnatt,

Department of Fine Arts

Date ______________________

Abstract

This thesis presents Langwidere, a facial animation system. Langwidere is intended to serve as the basis for a flexible system capable of imitating realistic characters and actions, such as speech or expressing emotion, as well as creating the exaggerated and fantastic characters found in traditional animation.

In the field of computer generated characters, facial animation is difficult because of the complexity of the surface and the fact that humans have entire sections of their brains devoted to facial processing, finding and magnifying any imperfection. Typically, the construction of new characters involves tedious and repetitive labor. An unwieldy level of detail is often needed for a realistic model. Apart from any geometric consideration, animating a new character often requires reprogramming or changing and retuning a large number of arcane parameters.

Numerous small, time consuming details go into creating a complete and visually interesting animated sequence. Much of this repetitive and tedious labour may be automated by a well integrated, user-friendly animation environment. Langwidere is created as the foundation for such a system.

Acknowledgements

First and foremost thanks must go to Dr. David Forsey for the hierarchical B-spline editor, Drag, on which Langwidere is built, not to mention technical advice and general enthusiasm. Thanks, also, to Dr. Keith Waters for a copy of his thesis project code, and especially to Dr. Len Breem for hunting down all those many files, and the thesis itself, and sending them to me.

Many thanks to Vince Demarco, Shauna Hoar, Mark James, and Thomas Orth for being such good sports and letting me glue reflective balls onto their faces. Thanks also to Glenda for lending equipment and expertise in motion tracking (and gleeful application of glue) to the project.

For stalwart proof reading assistance, and the occasional strenuous wording objection, thanks must go to David Hill, Tom, Aslam Mohammed, Norman Yee, Natascha Schuler, Eric Schenk, and Mark, Greg, Peg, and Alan James.

For valiant bureaucratic assistance thanks go to Lorraine Storey and Camille Sinanan.

Thanks to Bev Frangos, Tom, and Verle Henry for frantic last minute (of course) photocopying help.

Thanks for financial assistance and general nagging go to Brian Wang, Mom, Dad, Alvin Wang, and, last but not least, the Computer Science Department of the University of Calgary.

Thanks for the idea in the first place (even if it did drift a tad) go to Dr. Brian Wyvill, and my supervisor, Professor David Hill for guidance along the way.

Finally, thanks to Mark James for being my best friend and my own personal software support staff.

Dedication

To my mother,

Yes, Mom, *now* it's done.

To my father,

Harmony, help, and health, I will remember.

To my brothers,

I'm sorry I could not help more when you needed it.

Contents

List of Tables

List of Figures

Notation

Many aspects of "modeling" were dealt with in this project. It is unfortunate that many of the terms generally used in this area are similar, leading to confusion. The glossary is an attempt to separate and define the types of modeling as they will be described in this paper, explain unavoidable jargon, clarify terminology, and expand frequently used acronyms. Further, several "systems" are referred to in association with the models; the specifics of these systems are given for quick reference. Terms that can be found in the glossary are denoted by the sans serif font when they appear for the first time.

Chapter 1
Introduction

> "Young man," they said, "Your search is through
> You can trust in us, we'll look after you
> We will be your body, we will be your mind
> And we've a face of each and every kind
> Yes, we've a face of each and every kind"
>
> Nik Kershaw
> *Faces*
> from the album "Human Racing"

An environment for human character animation should be as easy as possible for anyone to use whether their goals are artistic, scientific, or commercial. The model should be realistic and generate natural movement and logical behaviour where desired, but create illogical or exaggerated behaviour on demand. This thesis presents Langwidere, a facial animation system that is intended to serve as the basis for such an environment.

1.1. Implementation Goals

Typically, the construction of new characters involves tedious and repetitive labor. An unwieldy level of detail is often needed for a realistic model. Animating a new character often requires recoding or changing and retuning many arcane parameters in software. It is for this reason that researchers are so often associated with a single character. The goal of this thesis is to introduce a system that supports easy creation of new characters without reprogramming.

Specifying an animation is frequently difficult, tedious, and time consuming. Many natural actions are small and periodic, such as blinking or gaze tracking, but their lack can be disturbing to the viewer. Other actions, such as speech, while simple in theory, are so involved that animating them is an arduous task. Both these problems are ripe for automation. There are many possibilities for automation in human character animation that would both simplify user interaction and make for a more realistic and interesting end product. Speech synthesis techniques can be used to automate lip movement (Hill, Pearce, & Wyvill, 1988; Lewis & Parke, 1987; Lewis, 1991; Pelachaud, 1991). Papers on psychology, sociology, and biology (Birdwhistell, 1970; Kendon, 1981; Morris, 1982) can be used to determine behavioural patterns. The foundation for a system using these techniques would be a head model capable of showing emotion and speech accurately.

The source code for two facial animation and modeling systems was publicly available for research purposes: Fascia, Parke's public domain version of his work, and Waters' muscle model process, as implemented for his Ph. D. thesis. An earlier modified version of Fascia had been used locally in a speech animation system, but it produced visible artifacts owing to difficulties with the polygonal representation. Proper lip rounding for vowels could not be achieved with this model.

The source code and data files for Waters' muscle model were obtained and modified for testing purposes. Again, movement in the lip region was unsatisfactory. Further, general artifacts were produced as a result of using only sixteen muscles for the entire face and the polygonal nature of the model that was particularly evident when the face was animated.

For serious animation research, a head model with proper movement is required, which means that a new structural model as well as a new actuation model needed to be created for the animation system. The scope of this project has therefore been limited to creating a basic facial animation system with hooks for proper speech animation as the foremost goal.

1.2. Overview

This introductory chapter outlines the Langwidere's goals and describes a number of applications for the finished system. A short overview of each chapter follows.

Chapter 2 contains a survey of surface modeling methods used by previous facial animation systems. After comparing these techniques and contrasting desirable characteristics with feasibility the elements of Langwidere are selected. Possible implementations that were discarded in favour of the current system are described.

Chapter 3 provides detail on the design elements of Langwidere: hierarchical B-splines, simulated muscle actuation, and the Facial Action Coding System (FACS). Features of the original systems by Forsey, Waters, and Ekman and Friesen are analyzed and a guide to modeling faces with hierarchical B-splines included. The animation scripting language, Ev, is planned out with respect to timing and syntax.

Chapter 4 describes the Langwidere system, which was implemented as an extension to David Forsey's hierarchical B-spline editor, Drag. The various stages of producing "Al", the first head model, and its offspring are detailed. Langwidere's muscle types, muscle contraction types, and muscle interaction types are followed by a brief description of some animations produced with Langwidere.

Chapter 5 summarizes the design elements of Langwidere and evaluates their characteristics relative to their development during the course of the thesis project.

Concluding remarks and possible directions for future work are given in Chapter 6.

FACS' list of Action Units and their component muscles are compiled in Appendix A.

Appendix B contains a list of the expressive muscles of the face that were implemented in Langwidere and their reference codes. Each muscle is annotated with location and function described in non-technical language.

1.3. Applications

Imitating a human face realistically is important for many applications. Structural precision, realistic motion, exacting speech simulation, or simple visual believability are of varying importance to the various applications in user interface design, education, health sciences, forensics, media, and art. Specific examples of applications for these areas follow.

1.3.1. User Interface. Many researchers in Human-Computer Interfaces would like to use animated faces to create a comfortable and informal environment through which phobic or non-technically inclined users can interact with computers. An interface of this sort would require speech animation as well as speech recognition and natural language processing that is not entirely within current software capabilities. Some researchers at Sony are working on bridging this gap (Franks, 1993).

1.3.2. Education. There are many educational uses for a facial animation system with speech capability:

- Teach pronunciation and other basic language skills
- Teach new languages
- Speech therapy
- Correcting learning disabilities
- Language training for the hearing impaired

Precisely animating the lip positions, teeth and tongue would be essential, but other information would also be needed. Speechreading, which is more popularly known as lipreading, is more than just reading lips (O'Neill & Oyer, 1981; Woodward & Barber, 1960; Tatoul & Davidson, 1961). The information that can be garnered by just lip movement is not as much as the layman expects. Proper emotive clues and facial gestures are critical if the reader is to understand what is being said.

1.3.3. Health. Health services cover a wide range of activities, such as psychology, psychiatry, speech therapy, and medical imaging.

The field of psychology, in trying to understand human perception, has many uses for a facial animation system:

- Perception of age.
- Perception of identity (Bruce, 1988).
- Identifying emotional states (Ekman & Friesen, 1978a, 1984).

The field of psychiatry, which often uses human substitutes as a psychiatric evaluation or therapy device, might also find such systems useful.

Speech therapy is often needed for victims of disease or trauma, such as stroke.

A system for medical imaging requires a high degree of accuracy with respect to the physical properties and interaction between flesh and bone for reference in reconstructive surgery planning or anatomical studies. Surgeons could show their clients options or projected results in such situations as dental reconstruction and plastic surgery (Pieper, 1991). Likely only the most rudimentary animation capabilities such as moving one joint or muscle at a time would be needed for these purposes.

1.3.4. Forensics. A system for forensic work would have even less to do with animation, but tools and techniques for constructing and specifying faces — as are useful for police and courtroom work — would also be indispensable to a facial animation system. Criminal identification requires constructing a face from verbal descriptions. When a victim has been lost for some time, it may be necessary to update images by artificial aging (Bruce, 1988). In even more extreme cases, an unknown victim's appearance might have to be reconstructed with tissue thickness statistics (Krogman, 1962) from the skull out. These applications are all for static head models, but as it becomes less expensive and better known, computer graphics is being used in the courtroom to illustrate technical details for lawsuits (Iversen, 1991) and it would not be at all surprising to see facial animation used in this setting.

1.3.5. Media and Art. The use of computer graphics in media and art increases daily. Applications in motion pictures, television shows, and advertising are now commonplace, but few computer generated *human* likenesses are seen and these generally take the form of modified or processed actors (see Section 2.4.2.1 for examples).

Feature films such as *Terminator 2: Judgment Day*, *Robocop II*, and *The Lawnmower Man* used computer animated figures for important aspects of key characters.

Realistic facial animation has largely been out of reach for artists to experiment with because of high equipment cost and requirements for serious technical knowledge and abilities.

1.4. Summary

The applications for an appropriately flexible and realistic facial animation system are many and varied. A survey of previous systems is needed to see what possibilities already exist, what characteristics are desirable, and what elements are feasible within the scope of a Masters project.

Chapter 2
Modeling and Animating Talking Heads

> Once the whole is divided, the parts need names.
> There are already enough names.
> One must know when to stop.
> Knowing when to stop averts trouble.
>
> Gia-Fu Feng and Jane English (trans)
> page 34 of
> *Lao Tsu Tao Te Ching*

2.1. Terminology

A complete human character animation system has many elements — facial animation, human character animation, speech synthesis, human-computer interface, biological simulation, and behavioral simulation — and borrows from many disciplines — psychology, sociology, biology, robotics, physics, and, of course, computer science. The basic building blocks of such a system as defined in this thesis are:

Surface Model: the method by which the surface of the object is represented.

Structural Model: the collection of data representing the physical shape of the object.

Actuation Model: a parameterization of the structural model with inherent manipulation methods for the purposes of simulating motion.

Muscle Model: a simulation of the physical processes that affect the flesh when a muscle is contracted.

Animation Model: a method of grouping and modifying the actuation parameters for animation.

Speech Model: a method of generating of speech synchronization data for animation.

Aspects and detailed examples of each of the above will be discussed beginning with a short comparison of surface modeling techniques.

2.2. Modeling Surfaces

As with any computer graphics project, the choice of surface representation when animating a human face critically affects the result in many ways. Since some of the surface modeling techniques discussed in the surveyed systems are not widely known, a description of several methods follows.

2.2.1. Polygons.

The surface modeling discussion begins with polygons and their more powerful variant, winged-edge polygons.

2.2.1.1. *Plain Polygons.*

Polygons, or in most cases triangles, are an extremely popular method of surface modeling. Routines to manipulate and render polygons can be found in virtually any three-dimensional graphics toolkit or rendering package. Dedicated graphics workstations such as those produced by Silicon Graphics often have polygon rendering built into the hardware making them fast and easy to use. When using polygons, the primitives are usually intended to be planar, which is why triangles are most often used since they are guaranteed to be planar. When using polygons to represent a curved surface, smaller polygons are used to approximate the surface more closely. Shading techniques such as Phong or Gouraud are used to create the illusion of a smoothly curved surface. Frequently, objects that are stored in other formats for easy manipulation are polygonized for rendering (van Overveld & Wyvill, 1993), but in animation this often causes shimmering (minute inconsistencies in the polygonization as the model moves and transforms) that is difficult to avoid.

2.2.1.2. *Winged-Edge Polygons.*

In the winged-edge polygon representation, only the vertices of the polygons are explicitly stored. Edges are represented by pointers to the appropriate vertices. Each vertex has pointers to the edges it terminates. Each edge has pointers to the two faces it separates (Foley, van Dam, Feiner, & Hughes, 1990; Glassner, 1991). A face, or polygon, keeps a list of the edges that define it, and each polyhedron a list of the faces that define it. In this way, manipulating an object can proceed in an hierarchical manner and data redundancy can be eliminated, unlike the traditional polygonal mesh which potentially must search the whole data set to find adjacent vertices or edges that share a vertex (Foley et al., 1990, p. 545).

Symbolics' *S-Geometry* system, which facilitates hand-editing of forms based on winged-edge polyhedra, takes the extra step of remembering successive generations of base parent and refined, or "smoothed", child objects. Control from the parent object will propagate to the child (Elson, 1990). This allows a lower resolution model to be used for quick rendering and motion testing.

2.2.2. Splines. Mathematical splines are piecewise polynomial curves with specific properties of continuity and smoothness or fairness. The cubic spline is generally used as a compromise between smoothness, ease of specification and control, computation time, and numerical stability. Extending the cubic spline curve into three dimensions gives bicubic patches. Splines generally allow fewer data points to be stored at the cost of extra manipulation and computational overhead for rendering. For a more complete discussion see (Rogers & Adams, 1990; Foley et al., 1990).

There are many different types of splines, but for the sake of brevity, discussion in Section 2.2.2.1 will be limited to cubic splines. Section 2.2.2.2 deals with hierarchical splines.

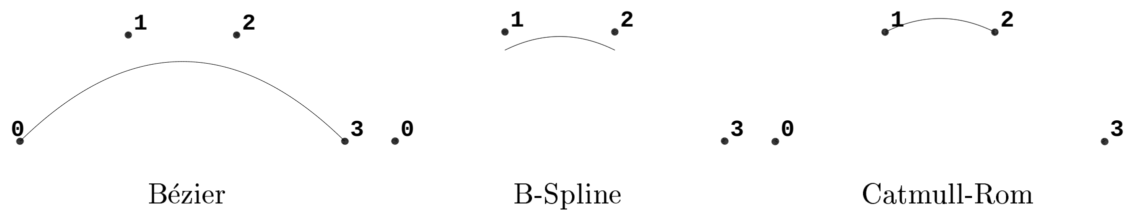


FIGURE 2.1. Cubic spline curves

2.2.2.1. *Cubic Splines.* Three different types of splines will be mentioned in the structural modeling discussion: Bézier, B-Spline, and Catmull-Rom. For each of these, four control points, or vertices, define a curve segment[1] (see Figure 2.1). Bézier curves pass through the first and last control points but only pass near the middle control points. The entire curve fits within the polygon defined by the control points, a phenomenon known as the convex-hull property; this is also true of B-splines but not of Catmull-Rom. The B-spline curve segment starts near the second control point and ends near the third, passing through none. The Catmull-Rom curve segment starts at the second control point and ends at the third. When joining segments together, control points can simply be added in sequence for Catmull-Rom to define a curve that is connected and continuous in the first derivative, B-splines are continuous in the second derivative as well. With Bézier splines this is more complex, curve segments do not "naturally" connect. The first control point of the succeeding segment must repeat the last control point of the preceding curve segment in order to force the curve to be continuous. To get a first order continuous curve, the second last and last control points of the preceding segment must be collinear with the first and second control points of the succeeding segment.

Generally, splines are only used to represent smooth surfaces; if you want to represent a cube you are much better off with polygons. Also, splines cannot match every topology. For instance, bicubic splines cannot properly represent spheres, ellipsoids, or other quadric surfaces. If you imagine trying to sew together a round ball from a single square piece of cloth, it is easy to see that there will be many problems with seams and wrinkles in the attempt to coax a round shape from the mess. There is simply no way to construct a *quadric* surface from bi*cubic* patches. Forcing the ends of a cylinder or sheet of bicubic patches together in order to close or seal the surface inevitably results in a discontinuity of the first and second order derivatives to the surface at that point. This is known as the

[1] For quadric splines three control points are needed, for quartic splines five, and for quintic splines six.

"north pole" problem. On the other hand quadric, quartic, or other even-order polynomial tensor surfaces can be closed without any disruption of continuity.

2.2.2.2. *Hierarchical Splines.* Another shortcoming of splines is that while an existing surface can be given more detail in a local area by splitting the patch and inserting new control points, all patches contiguous to the affected patch must also be subdivided, thus adding more detail and data where none was needed or desired.

Hierarchical splines allow local refinements to remain local by creating new, more detailed patches that exactly conform to the surface at the specified location. The new patches are defined in terms of the parent surface so that editing of the whole model can continue, and changes that affect the parent surface will propagate to the child surface ensuring that the surface remains continuous and the local refinement remains fixed in place (for a more detailed description see Section 3.2.1 and (Forsey & Bartels, 1988b, 1988a; Forsey, 1990, 1991)).

2.2.3. Other. Surface modeling methods are many and continually evolving. Three more which may be considered "major" are presented for further comparison purposes.

2.2.3.1. *Implicit Surfaces.* Implicit surfaces — also known as isosurfaces, soft objects, metaballs, and blobbies — are defined by a field function around a skeleton object or frame. The surface is defined as the set of points within a certain distance of the frame. A point in the influence of two or more points of the frame is subject to blending, thus forming a single continuous surface (see Figure 2.2). As might be expected from the plethora of names, this is a relatively young modeling method, and new techniques for blending, rendering, manipulation, and animation are developing rapidly. For more information see (Bloomenthal & Wyvill, 1990).

2.2.3.2. *CSG.* Constructive solid geometry, perhaps better known as CSG, is a technique whereby closed surfaces, usually Platonic solids and quadrics, can be combined with Boolean operators to produce new objects. Typically, objects produced with these methods are used

TDI Blob Modeler (Belleville, Associate Editor, 1993)

Metaball Ballerina (Yoshimoto, 1992)

FIGURE 2.2. Examples of implicit surface characters

for rigid bodies such as machine parts, not flexible surfaces like the face. However, free form deformation will work on any surface and one generated with CSG would be as good as any other.

2.2.3.3. *Particle Systems.* Particle systems deal with "particles" (a term more commonly associated with dust and sand but which are often as complex as leaves or birds) and functions that govern the behavior of the particles. This technique can be applied to such varied tasks as modeling schools of fish, blood, fire (Reeves, 1983), hair (Kajiya & Kay, 1989; Perlin & Hoffert, 1989; Lewis, 1989), grass, trees (Orth, 1993), cloth (Breen, House, & Getto, 1991), fireworks, and waterfalls (Sims, 1988).

2.3. Modeling Structures

The structural model contains the data representing the physical shape of the face (and head, if applicable). Problems that accompany the structural model include model construction via scanning or editing, and data representation (see Section 2.2). Previously existing modeling methods are discussed in terms of their surface modeling technique in this section. Hooks for data manipulation techniques must frequently be built into the data structures which store the model. These will be dealt with in the actuation model discussion in Section 2.4. The actual data structures used in this implementation will be described in Chapter 4.

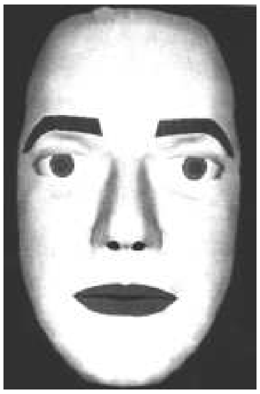
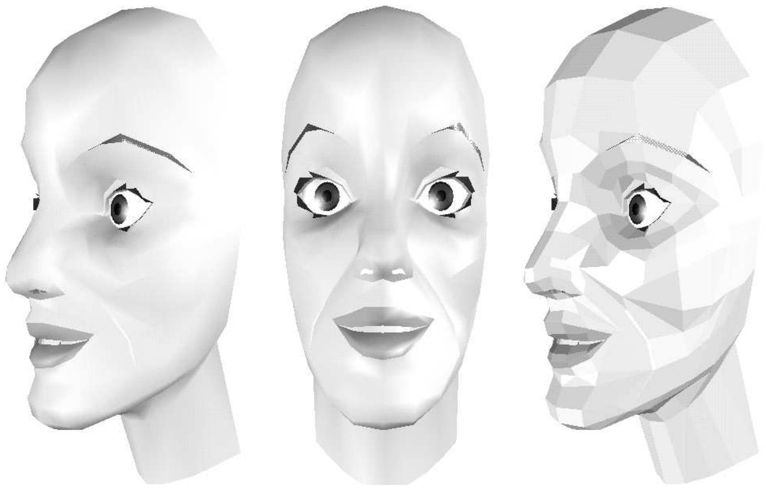


FIGURE 2.3. Catherine Pelachaud's head model (Pelachaud, 1991)

2.3.1. Polygonal Models. As might be expected from the popularity of modeling with polygons, many facial animation systems are based on polygonal meshes, including Frederick Parke's Fascia, Steve DiPaola's FAS, Matt Elson's work at Symbolics, Stephen Platt and Catherine Pelachaud's work at the University of Pennsylvania (see Figure 2.3), Nadia Magnenat-Thalmann and Daniel Thalmann's famous actor collection, Keith Waters' work at Middlesex Polytechnic, Tony de Peltrie created by researchers in Montreal, Nestor Sextone created by Berkeley researchers, Mike Normal by de Graf/Wahrman, and Don Quichotte created at Videosystem in France. The first six of these will be discussed in further detail in this section; the rest, while significant works, do not greatly differ in theory or technique.



(a) (b) (c)

FIGURE 2.4. T-Square, rendered by Fascia

2.3.1.1. *Fascia.* Frederick Parke's face model[2] is one of the premier works in the field of human character animation. The model consists of a polygon mesh with carefully placed vertices to add detail and provide curvature where needed. While careful shading of the polygons creates the illusion of a smooth surface (see Figure 2.4 (a) and (b)), this cannot conceal the reality of their flatness in the profile and borders of the face (see Figure 2.4 (c)). In animation this problem is further highlighted since the only points of flexibility are the vertices and edges of the polygons, which are not always the points at which the face should be bending for an expression or lip position. Also, actuating a polygonal model requires "explicit, direct control of specific facial vertices. Such direct control, while exhibiting certain favourable qualities, is restricting in terms of flexibility and creating generalized manipulation approaches." (DiPaola, 1991, p. 130).

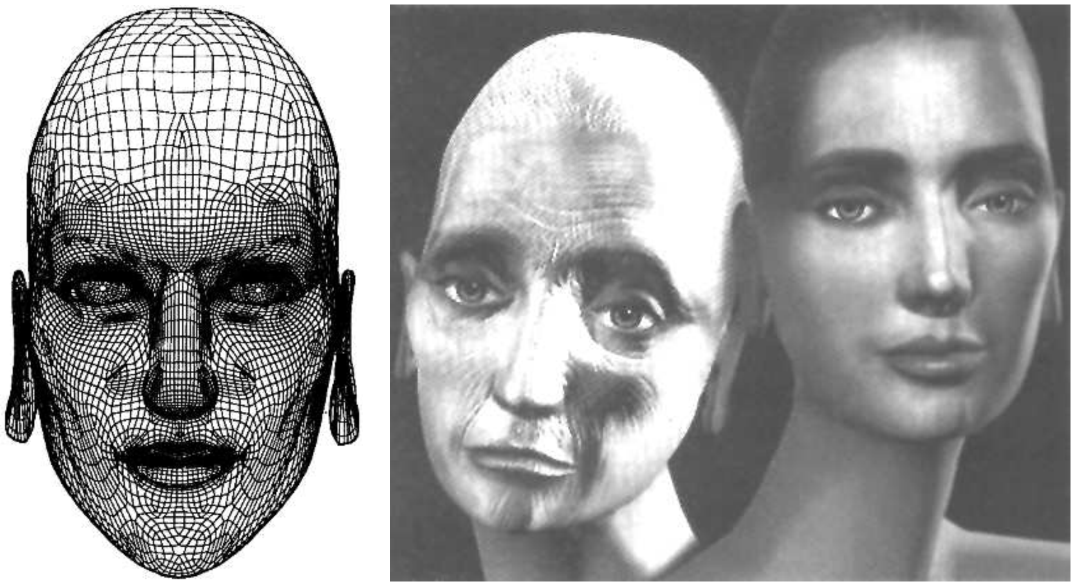
[2]Called *Fascia* in its current public domain release.

FIGURE 2.5. FAS creations (DiPaola, 1991)

2.3.1.2. *Facial Animation System.* Steve DiPaola's Facial Animation System (FAS) is an extension of Parke's model (see Figure 2.5). The purpose of FAS is to provide an environment for the easy creation, manipulation, and animation of faces. New facial models are generated by digitizing people or sculpted forms and by manipulating old computer models with free form deformation, stochastic noise deformation, and simple vertex editing. New parameters for manipulating and controlling the face model are constantly being generated. FAS is based on Parke's polygonal model[3] and suffers from similar drawbacks, but its creator suggests that "Hierarchical patch representations, such as those presented by Forsey and Bartels (1988b), appear to offer what might be at the present state of animation research the best solution for obtaining indirect control of a facial model, providing variable resolution and high quality surface curvature." (DiPaola, 1991, p. 131)

2.3.1.3. *Lotta Desire.* Matt Elson from Symbolics uses Winged-edge-Polyhedra to model "Lotta Desire" from the animation *The Little Death* and *Virtually Yours* (Elson, 1990). When a model is no longer clearly imaginary, or a caricature, it suffers intense scrutiny from the viewer. The smallest detail can be used as an excuse to label it as unreal or wrong

[3]All of Parke's faces have the same number of vertices with the same topological arrangement.



Wireframe (Elson, 1990) Lotta, old and young (Robertson, 1991)

FIGURE 2.6. Lotta Desire, modeled with winged-edge polyhedra

(Waters, 1988, p. 204), especially with an object so familiar as the human face. Lotta's hair is just a bit of colour on its scalp (it actually looks like a bald person who has used eyebrow pencil to colour in hair), its ears flare out unnaturally from the sides of its head, and it is far too thin. On the other hand, the sequence in which it ages complete with wrinkles and appropriate colouration changes in *Virtually Yours* is quite impressive (see Figure 2.6).

2.3.1.4. *Miscellaneous Polygonal Models.* Platt implemented a structurally based model which is segmented according to the Facial Action Coding System's *Action Units* (see Section 2.4.3.1 and 3.5.1). This model was also used by Pelachaud in a speech animation system (Pelachaud, 1991).

Nadia Magnenat-Thalmann and Daniel Thalmann use extremely detailed polygonal models to counterfeit famous actors such as Marilyn Monroe and Elvis Presley. The models are created in a number of ways such as:

- Scan in orthogonal photographs and interpolate the given points to generate a 3D model.

- Create a physical 3D model in clay or plaster.
  - Use a 3D digitizer to gather data.
  - Slice the model at regular intervals and use a 2D digitizer to gather the data from the cross-sections which is assembled in the computer to create a 3D model.
- Edit or modify a pre-existing 3D model.

It is the difficulty in modeling complex 3D objects that makes digitizing sculpted work so popular. Further, in this manner computer scientists can get help from artists to produce models with more character and life. Also popular is gathering basic structural data from real people since the realism is inherent, and it is usually cheap to find volunteers. Keith Waters was no exception in this regard, and has generated a number of polygonal models through digitizing people and sculptures. During the course of his Ph. D. research, however, he experimented with spline patches and had this to say about them: "For the modelling of the face patches have two advantages over direct polygonal descriptions: firstly the resolution can be selected for the structure, and secondly doubly curved surfaces can be generated." (Waters, 1988, page 230)

2.3.2. Spline Models. Since the skin is a smooth, flexible surface, many researchers have taken the approach of using spline patches in their models. Examples include Billy created by a team at Pixar lead by William Reeves, the Facial Action Control Editor created by Clea Waite at MIT, the model created by Komatsu for the "Caron's World" animation, and a model created by a team of researchers at the University of Paris. Further details on these models will be given in this section.

2.3.2.1. *Billy.* Pixar has an unwritten rule to "only use polygons for flat things" (Reeves, 1990, currently written at the bottom of p. 98) hence the choice of bicubic Catmull-Rom spline patches in constructing Billy, the baby in "Tin Toy" (see Figure 2.7). Triangular Bézier patches were also attempted, but severe wrinkling problems precluded their use. Wrinkling remained problematic, although less so after the switch from triangular patches

(a) (b) (c)

FIGURE 2.7. Billy, from "Tin Toy" (Reeves, 1990)

to Catmull-Rom, and Billy still had a "north pole" problem at the top of his head as well as major problems with the neck. In Reeves opinion, "Hierarchical patch representations such as that proposed by Forsey (1988b) present a solution to this problem. At the time of 'Tin Toy', these had not been published." (Reeves, 1990, p. 99)

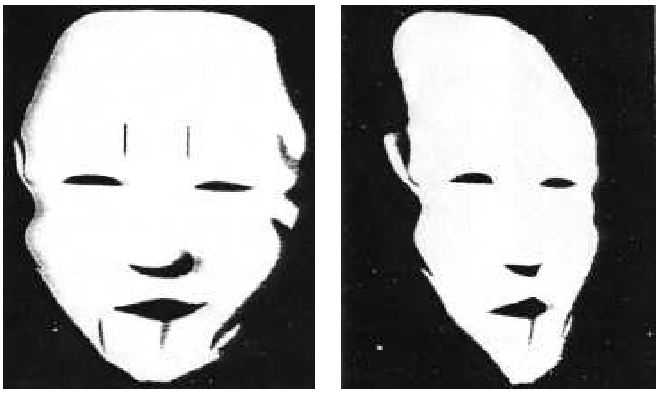


FIGURE 2.8. Examples of *face* (Waite, 1989)

2.3.2.2. *Facial Action Control Editor.* Waite's Facial Action Control Editor (*face*) used B-splines with a total of 16×12 control points. Openings for the eyes, nostrils and mouth are cut into the surface with a technique called *trimming* (Foley et al., 1990, p.528). As may be expected from the number of control points, this is not a very detailed model; there are no eyelids nor are there any eyes (see Figure 2.8). The author considers this "enough ... for general face shaping with no additions. ... On the other hand, horns, warts or other anomalies would not negatively affect the model. With refinement (Forsey & Bartels, 1988a) some interesting 'monsters' would be possible." (Waite, 1989, page 53–54)

2.3.2.3. *Caron's World.* In his 1988 paper, Komatsu models a human body with biquartic Bézier splines. This method was chosen for the reasons given below from the paper: (Komatsu, 1988, page 265)

(1) The shape of the model must be smooth everywhere.
(2) The three-dimensional form of the body is defined by the skeleton, which must change flexibly at the connecting angles between bones.
(3) The shape must reflect the motions of the skeleton in accord with action, and the shape must always keep its smoothness.
(4) A local change in the skeleton must affect only a small part of the shape.
(5) Change of shape such as swelling of muscles must be simulated. Condition 2 is required to produce animation. Condition 4 means that changes of the shape are restricted to parts directly related to movements of the skeleton.

Sixty-four Bézier patches and sixty Gregory[4] patches go into the making of a basic body shape with rudimentary hands, feet and head. A later, "improved" model with fingers had 156 Bézier patches and 172 Gregory patches, more than double the number of patches used in the original model. No reason was given for selecting biquartic patches over the more common bicubic patches. Possibly they were chosen so that the surface could be sealed at terminating points without the need to worry about "north pole" difficulties or the irregularities of human topology.

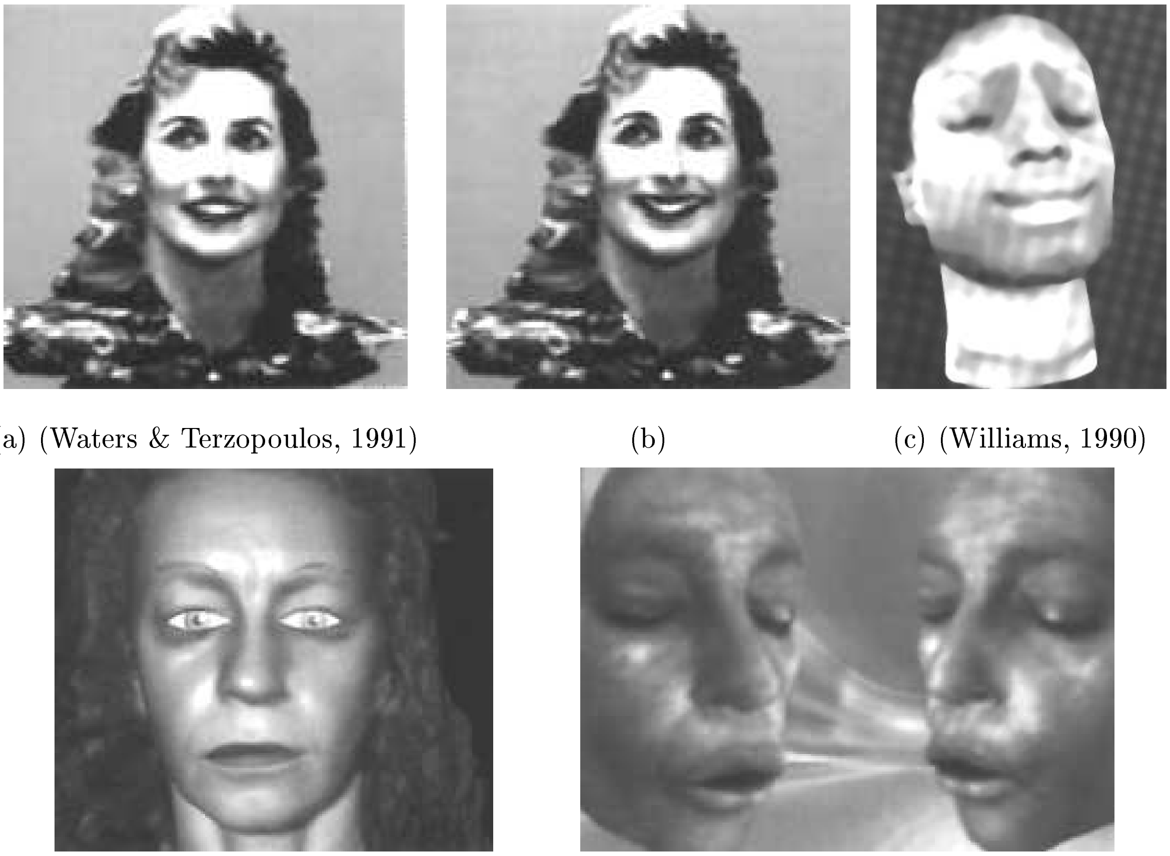
[4]Pentagonal patches constructed from five Bézier patches which were used at the joints in order to deal with branching structures (such as arms) in a more graceful manner.

FIGURE 2.9. Biquintic B-spline faces (Nahas, Huitric, & Saintourens, 1988)

2.3.2.4. *Nahas, Huitric, and Saintourens.* As the Komatsu model demonstrates, adding detail to one area of a continuous spline surface causes a dramatic increase in the number of patches needed in the surface.[5] Despite these possible complications, a team of researchers at the University of Paris built a model of the face out of B-splines (Nahas et al., 1988). Nahas, Huitric, and Saintourens use biquintic B-splines to model head data obtained with a 3D scanner. The problem with traditional splines is that only a single level of detail can be chosen. Areas of low detail (back of the head, sides of the cheeks) are over-represented, and areas of high detail may not have the required amount of curvature. Thus their 1988 model has no nostrils and the lips look blurry (see Figure 2.9).

2.3.3. Facial Masks. Feeling that the structural model cannot or need not be further improved — perhaps due to the difficulty in reproducing skin texture algorithmically — many researchers have turned to scanning real faces that they use as texture map masks over their structural model. Laser scans provide extremely detailed 3D data as well as the colour or grey level of the scanned points. This data can be combined to create an amazingly realistic static model. The problem then becomes one of finding points that correspond to the appropriate parts of the face in the large welter of data generated by the

[5]The Komatsu model has what might charitably be called a vestigial face.



(a) (Waters & Terzopoulos, 1991) (b) (c) (Williams, 1990)

(d) (Nahas, Huitric, Rious, & Domey, 1990) (e)

FIGURE 2.10. Examples of Facial Masks.

scanning process, and ensuring that the texture mask moves appropriately when the facial surface is deformed to simulate movement.

Among those that have employed this technique are Nahas, Huitric, Rious, and Domey (Nahas et al., 1990) (see Figure 2.10 (d) and (e)); Waters and Terzopoulos (Waters & Terzopoulos, 1991) (see Figure 2.10 (a) and (b)); and Lance Williams of the Apple Computer Advanced Technology Group (Williams, 1990) (see Figure 2.10 (c)).

A different type of mask has been created by researchers in Japan who performed "brute force" animation of a Kabuki mask, a 2D model deformed to appear 3D (Franks, 1993;

Gross, 1991).

2.4. Modeling Actuation

Once a structural model is complete, there are various ways to animate it. Some opt to animate the entire model directly, but most use intermediate techniques that parameterize possible actions, simulate biological processes, or otherwise move the model to action. This section deals with that step leading to full-fledged animation.

2.4.1. Interpolation Techniques. Constructing virtual 3D objects on a computer is difficult. One is at the mercy of the medium and the programmer who wrote the tools being used to provide the desired view and allow manipulation of the point required. Even *finding* this point may be a difficult task since one point on the screen has a tendency to look much like any other point on the screen, and a sense of volume or location is lacking. Further, when the model begins to get large or complex there are often too many other points or lines obscuring the focus. There is no physical feedback to the process that can aid intuition and, more often than not, the model in progress is not being rendered in its final format. Thus the actual appearance of the model may also be in question. As programs and interaction techniques become more sophisticated, these and other difficulties can be ameliorated. Constructing a model would then be reduced to a test of one's skill and artistry. Many computer scientists have no artistic ability and often obtain the help of artists to sculpt physical 3D models for them. The computer scientist's contribution is then to scan in or otherwise digitize the model for computer use.

Generating movement can have some of the same problems experienced in modeling. Which point should be moved? How does this really look? If there is no system in place for generating motion in a procedural manner, the artist may be called upon once more to recreate the model in several different positions. Alternatively, the scientist may work to create specific target positions between which the computer would interpolate to simulate motion. This is known as *keyframing* or *inbetweening*.

Some works that have used this method include *Tony de Peltrie* by researchers from Centre de Calcul, Université de Montréal (Bergeron, 1989), *Don Quichotte* by Videosystem of France (Videosystem, 1991), and by Nahas, Huitric, et al in their two papers (Nahas et al., 1988, 1990).

Another case which may be considered rotoscoping (see Section 2.4.2.1), but more easily falls into the category of interpolation, is ILM's T1000 character from the movie *Terminator 2*. The face of the character in the movie rarely deviates from a scanned state except when transforming or "morphing" between shapes, but cannot be considered actuation in the sense of a process that causes an action to occur. Other than transformation, the only computer generated action involving the face of the character was the delivery of two words. The phonemes for the phrase "Get out." were each digitized, requiring the actor to hold his face perfectly still and in position for fifteen seconds at a time so that he could be scanned (Duncan, 1991, p. 46). Similarly, when the pseudopod in *The Abyss* morphed into Mary Elisabeth Mastrantonio and Ed Harris, their faces and expressions were digitized, and used as raw data for the shape of the pseudopod. Interpolation between the configurations was used to provide the "animation".

These control methods are extremely cumbersome, requiring extra equipment and/or drawing and gluing things onto people. It is also difficult to blend actions in sequence together properly, much less change or correct some small part of the action without redoing most of the work that has already been completed. At the least, requiring an actor/performer to generate animations is inconvenient.

2.4.2. Performance Techniques. Another way to gather information from the real world is to record the actions of a real person manipulating direct input devices.

2.4.2.1. *Rotoscoping.* As applied to traditional animation, rotoscoping usually takes the form of filming the subject in motion and later tracing or copying the image onto animation cels. With computers the job can be even easier. Frames from film or video tape can be directly input to the computer and image processing techniques applied to distort or enhance the image. Techniques such as these were applied to actor Matt Frewer in heavy makeup and prosthetic makeup appliances to produce Max Headroom.

Computers allow further abstraction of this technique. Cameras properly calibrated and set up can, in three dimensions, track dots applied to a subject and record the motion. The motion once processed can be used to drive a properly configured model. This is the technique used by Robert Abel and Associates to animate the "sexy robot" in their canned goods commercial (Abel, 1988). This technique was also used by Waters and Terzopoulos to drive their dynamic tissue model of the face (Terzopoulos & Waters, 1990).

Apple's Lance Williams uses a mask generated from scanned data that is warped into motion by high contrast points applied to the performer's face. The moving these points then displaces corresponding points on the mask, simulating motion. A large plate on the forehead of the actor is used to provide a frame of reference for the moving points as well as data on the movements of the entire head (Williams, 1990).

David Forsey's *A Public Service Announcement* dragon was animated by attaching infra-red light-emitting-diodes to a hand and moving it as if manipulating a puppet (Forsey, 1990).

2.4.2.2. *Puppetry.* While the distinction between rotoscoping and puppetry is quite clear in the "real world", there is significant overlap of technique as applied to computer animation.

Mike Normal, created by the deGraf/Wahrman company, is controlled through combining most of the above methods. Expressions and speech positions are derived from a live subject using laser scanned "stills". Mike can then be performed using a variety of devices including DataGlove, a head waldo, and the keyboard. Its actions are controlled through a puppeteer interface.

2.4.3. Parameterization Techniques. For generating motion, one extreme is scanning and digitizing all motion and actuation data from external sources, the other is specifying each vertex point location in the model for each frame, between these would be *parameterization* techniques.

By using formulæ or lookup tables that specify the position of a group of vertices given a small number of parameters, a structural model can be actuated with much less effort.

2.4.3.1. *Facial Action Coding System.* FACS is a system for recording emotional states based on muscular configuration. This system was developed by psychologists Ekman and Friesen (1977) using all the muscles of the head and some from the neck. The 268 muscles of the face (Hill et al., 1988) are further grouped into 58 Action Units (see Appendix A and Table A for more details) that provide a convenient way to refer to muscles that tend to work together. This notation is so natural that it has become the de facto standard basis for the parameterization of facial animation systems despite the fact that the action units of expression do not necessarily lend themselves to speech generation.

2.4.3.2. *Platt.* Platt implemented a structurally based model with muscle model type control, which was recently used by Pelachaud to generate speech (Pelachaud, 1991). The physical model is based on a rather sparse polygon mesh and polygons in the mouth area appear to go the wrong way when speaking.

2.4.3.3. *Fascia.* Parke has parameters for the shape and size of the head. A different routine is created for each part that will move.

2.4.3.4. *Facial Animation System.* FAS has more than a hundred parameters that have been programmed for this model to ease the burden on the animator with respect to control, but most are surface-based shape conformation tools. Proper motion for fine actions as required for speech and emotion depend on the talents and artistry of the manipulator.

Other tools that were created for FAS include free form deformation methods (Haman, 1992) that can be applied locally to create hair for beards and eyebrows.

2.4.4. Local Deformation Methods. Both Fascia and FAS are actuated by routines specifically tailored to the area that they are moving, such as a routine that moves the corner of the mouth up for a smile, another routine that moves it down for a frown, and yet another to purse the lips. A different method of actuation is to write a small family of routines that are defined by the surface they are called to actuate. In this way the same code can be used to actuate different areas of the face without the need to constantly change the code for proper effect.

2.4.4.1. *Muscle Model.* Motion generated by a muscle model appears more natural than that of most hand animated systems. Some artifacts of this method are easily overlooked, but worth mentioning. There are no allowances for bones in the control model which affect areas where the bones are prominent, such as cheek and jaw, or where the fasciæ and muscle layers are thin, such as the forehead and jaw. A partial solution to this problem is suggested by rotational muscles as implemented by Bill Reeves at Pixar (see Section 2.4.4.3 for more details).

Another solution to the bonelessness problem is Waters' and Terzopoulos' work on a model of facial tissue (Terzopoulos & Waters, 1990) that gives the flesh thickness and physical properties, although the bones of the skull beneath the flesh are not really accounted for. Another artifact of this model is not so easily overlooked if one is interested in simulating speech properly (which requires lip-rounding and puckering). Mouth actions, other than smiling, frowning, and jaw rotations do not seem quite right. This is because Waters models the mouth area with a simulated sphincter muscle despite the fact that "The Orbicular oris … is *not*[6] a sphincter muscle" (Gray, F.R.S., 1977, p. 308).

> … the muscle consists of independent quadrants (upper, lower, left and right) that co-operate in many contrasting ways in the great variety of labial actions. There is some interlacing and transmedian overlap of corresponding right and left quadrants, while there is, at least, close juxtaposition of upper

[6]Emphasis was applied by this author.

> and lower quadrants at their lateral (modiolar) attachments. ... Formally, then, orbicularis oris as a whole is composed of eight segments ... each segment resembles a fan with its stem at the modiolus. Each fan is open in peripheral segments and almost closed in marginal segments. The region of apposition of marginal and peripheral parts is indicated by the lines of junction between the red-lip area and the skin ventrally and the mucosa posteriorly.

2.4.4.2. *face.* Waite applied Waters' muscle model to a simplified B-spline face. Four Action Units were implemented with what appears to be twenty-six muscles, thirteen muscles reflected bilaterally.

2.4.4.3. *Billy.* To provide actuation for Billy, Pixar used Waters' muscle model with modifications. In addition to the linear, sheet and sphincter muscles developed by Waters, rotational muscles were created. The area of influence for a rotational muscle was defined as a partial torus. The pulled muscle was simulated by rotating the torus to simulate flesh warping around the jaw and the eyeball. Forty-seven muscles were used in Billy's face.

2.4.4.4. *Lotta Desire.* Elson, the creator of "Lotta Desire", cites Waters' work and credits him as "inspirational" but does not clearly state the method of actuation.

2.4.4.5. *Pelachaud.* Pelachaud uses a facial modeling system derived from Platt's work. The structure is divided into sections analogous to the areas moved by FACS Action Units, but the actual algorithms or functions used to determine how these areas are moved are never stated (Pelachaud, 1991).

2.4.4.6. *Deformation Kernels.* Lance Williams uses "deformation kernels" to move his scanned facial masks based on the tracking points. High contrast points are applied to the face of a subject allowing the motion produced by a performance to be recorded digitally and then applied to a prescanned facial mask through deformations that are based on the location of these points (see Figure 2.11 and Figure 2.10(c)).

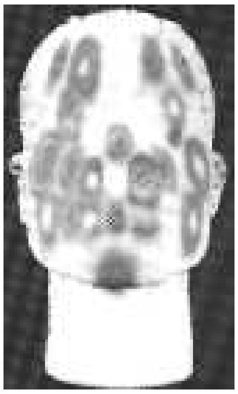


FIGURE 2.11. Deformation Kernels (Williams, 1990)

2.4.5. Dynamic Simulation Methods. Dynamics, in the context of computer graphics, is simulating motion based on physical laws. While Waters' muscle model process can be considered a dynamic simulation of the effects of a muscle movement on skin, no equations of force or elasticity are actually used in the calculations. Other projects which apparently use dynamics are Waters' and Terzopoulos' 1990 collaboration which presents a detailed simulation of skin (Terzopoulos & Waters, 1990), Steven Pieper's computer-aided plastic surgery system (Pieper, 1991), and David Chen's biomechanical simulation of the human gastrocnemius and bicep (Chen, 1991; Chen & Zeltzer, 1992).

2.4.5.1. *Simulated Skin.* Waters' and Terzopoulos' 1990 paper describes a three-layer, physically- and anatomically-based skin model. A deformable spring-mass lattice represents the flesh and can more realistically propagate forces, as generated by Waters' original muscle model process, through the simulated skin which now has volume and differential compressibility.

Waters stated that this type of model is difficult to control (Waters, 1992), obtaining precise movements as is necessary for speech animation is an extremely difficult exercise. The beauty of dynamics is that equations describe object motion so that this need not be specified by hand. The user simply sets up the conditions required and the results are generated automatically. The difficulty is that it is rarely possible to predict the exact results that will be generated. If all that is required is a particular type of motion, then this it not a problem. On the other hand, if exact actions or control are needed then dynamics may not be the answer.

Another model for simulating skin, this one intended for modeling the results of plastic surgery, was written by Steven Pieper. Finite element methods are used to model the effects of forces and strain as applied to the skin by surgery, gravity, and muscles (Pieper, 1991).

2.4.5.2. *Finite Element Muscles.* David Chen's 1991 thesis describes a detailed, finite element method biomechanical simulation of the human calf and bicep with precise data of muscle, bone and tendons obtained through MRI scans. The bones are actuated by the forces generated by the muscles. The muscles deform around the bones and each other, as well as being affected by gravity. Exact volumetric data is used and maintained to produce accurate muscular shapes under contraction (Chen, 1991; Chen & Zeltzer, 1992).

2.5. Animating the Models

A good actuation model obviates point by point control of the model by replacing that type of control with a much smaller number of parameters. Fascia uses 65 controls, compared to Waters' 16 muscles. These are relatively small numbers, but, in an animation, there are 24 frames per second for film (or 30 frames per second on video tape). Even a half minute film will require 720 individual frames to be specified. Yet another kind of model is needed to help the user generate actuation control parameters and to deal more easily with the time aspect of animation.

As with actuation and motion, timing can be set absolutely, by hand, or in a more high-level, relative manner, that allows easy additions, modifications, and coordination. Depending on the way the animation is specified, conflicting actions may need to be resolved or simply blended. Sequences of animation also need to be linked and then the whole animation generated.

A simple but important animated behavior is speech, and speech is used as an example test case in this section to help differentiate the methods of performance driven animation, keyframing, scripted specification, and digital actors. The use of speech synthesizers to drive animation will be covered in Section 2.6.3.

2.5.1. Rotoscoping/Performance/Puppetry. Because the amount of data required to produce an animation can be staggering, many people choose to generate an animation by performance to obtain subtle nuances of detail and sequencing. Then it is often difficult to change details when required, except by rerecording. Segments of performance that need to be collected into a single scene probably do not naturally flow together without significant processing (Duncan, 1991).

2.5.2. Keyframing. In traditional animation, keyframing means that an animator times and draws major (or framing) moments of action with great care and precision, and then fills in the "inbetweens" more quickly using the keyframes as a guide. With computers, the animator sets all the control variables for key moments of a sequence, and then allows the computer to interpolate the "inbetweens".

2.5.3. Scripted Specification. In a basic type of script, all actions are specified in text, by hand, as input for the animation or rendering output program. Keyframes can be considered a subset of scripting. Locations, objects, transformations, motion, timing, and types of interpolation need to be generated by the scripter for the output program. Most often, scripts are entirely text, but they may not be completely hand generated. What a computer program finds easy to read is rarely the same as what a human finds easy to read or write. Intermediate systems to translate or expand a human-generated scripts are common.

2.5.4. Digital Actors. While most of the complexities of human behavior are mental, enough of them have physical effects that psychology, sociology, and physiology texts become useful references for the human character animator. Gestures are used to punctuate verbal statements (Birdwhistell, 1970). Sociological studies of interaction tell us that the North American speaker often avoids looking at the listener preferring to make eye contact only as punctuation or to indicate that the other may speak at this point (Morris, 1982). The physiological effects of emotion can produce distended veins and flaring nostrils due to anger or lack of muscle tone and a loss of precise motion control due to sadness (Bezooyen, 1984).

Programmed rules of behavior could not possibly apply to all people in all circumstances, but they can certainly provide a context for the main, or specialized, actions determined by the animator, creating a less stark, more realistic, scenario and requiring less direct effort on the part of the programmer. Behavioral models in the form of artificial intelligence databases, or object-oriented programming "agents", would be the most likely vehicles for this type of implementation. This would be especially useful to help set timing and synchronize actions for animating dialogue. Catherine Pelachaud is pursuing this subject for her post-doctoral work.

2.6. Visible Articulation

The visual component of speech is an object of study for many reasons, some of which are listed in Section 2.6.2. Visual criteria applied to phonemes results in a new classification system called *visemes*. Previous examples of speech animation are cited according to technique.

2.6.1. Visemes. When modeling speech, it is fruitful to consider exactly what objectives the research is attempting to reach. Certainly, when the end result is the visual component of speech, the literature on speechreading, must be taken into account. Studies have shown that comparatively few phoneme classes are visually distinguishable. The resulting categories are called *visemes*. Visemes are distinguished by lip position and mouth shape; teeth and the visible part of the tongue also provide valuable cues.

Traditional animators often ignore lip synchronization completely, preferring to communicate with the viewer through head moves, expressions, and gestures (Thomas & Johnston, 1981). Speaking characters can be animated using "body English", with the complete omission of the mouth producing barely a perceptible difference (Culhane, 1988, p. 208). When they choose to do so, however, animators have developed a classification system, determined by studying mirrors and movies of live actors, which, although unscientific, is nonetheless useful as a reference. They break phonemes into the following categories:

/m, b, d, p/

/o, u, w/

/f, v/

/e, i/

/s, t, z/

Woodward and Barber (Woodward & Barber, 1960) visually classify the consonants thusly, the groupings being bilabial, rounded labial, labiodental, and nonlabial respectively:

/p, b, m/

/W, w, r/

/f, v/

/t, d, n, l, θ, ð, s, z,č, ǰ, š, ž, y, k, g, h/

The nonlabial group — including alveolar, dental, alveopalatal, velar, and glottal — were visually indistinguishable.

Vowels are largely separated by lip rounding and height of lip opening (Fromkin, 1964). But, as any ventriloquist will tell you, the height of lip opening for vowels can be as small as you want at the time (Ritchard with Kathleen Moloney, 1987), making the height of the lip opening a poor distinguishing characteristic. The visual vowel classifications below are largely derived from a linguistics text (O'Grady & Dobrovolsky, 1987) with some adjustments based on data from a speechreading text (O'Neill & Oyer, 1981).

/i, ɪ, e, ɛ, æ, a, ʌ, ə/

/o, ɚ, ʊ, ɔ, u/

2.6.2. Speech Reading. Considering how little hard data is conveyed by lipreading itself, it is amazing how much the hearing-impaired can understand about what is being said. But, recalling that this activity is also frequently termed *speechreading*, the information gathering process becomes a bit more clear. Speechreading involves deriving information, not only from lip and jaw positions, but from the *context* of what is being spoken, gestures, and evidence of emotion (Tatoul & Davidson, 1961). The more information provided about what is being said, the clearer the understanding, the better and more effective the communication.

Visual and auditory information can conflict with surprising results. Research by McGurk and MacDonald on visible articulatory movement perception shows that a visual /ga/ combined with an auditory /ba/ can induce the observer to perceive /da/. This is known as the McGurk effect (McGurk & MacDonald, 1976).

2.6.3. Speech Animation. A number of techniques for animating speech with computers have evolved. Some of the more prominent examples are discussed in this section including the synchronization methods Waters examined in his thesis work, and algorithmic techniques for automating the synchronization process which reduces the "grunt work" of animation. The timing constraints of synchronizing animation to the speech track are described. Combining emotion with speech in an animation is a largely unexplored area of research, and some basic ideas are touched on at the end of this section.

2.6.3.1. *Keith Waters.* Waters used several different methods of synchronizing speech to animation in the development of his thesis. He began by applying traditional animation methods, first determining phonemes by hand from a recorded soundtrack, then at the appropriate time animating the phonemes with the facial model. It was quickly discovered that the synchronization with the recorded speech was poor and that the face appeared to be "over-articulating" (Waters, 1988, p. 216). He had not taken coarticulation effects into account, and by searching for phonemic markers by hand, he was missing a lot of information. The effect of compressing the information given to his animation driver was the same as attempting to resynthesize speech that has been too ruthlessly compressed. In the end, Waters gave up on using artificial means to drive his speech and began using rotoscoping techniques similar to those employed by Apple's Advanced Technology Group.

As an aside in his Ph. D. thesis, Waters mentioned that synchronization of speech to animation (and many speech synthesis problems as well) might be solved if both speech and facial expressions were generated from the same model. This is to say that if the speech model were based on actual physical modeling of the articulators of speech, including the tongue, teeth, lips, throat, velum, etc., then the articulants that are visible would naturally be synchronous and accurate. Thus articulatory synthesis would be useful in this respect.

2.6.3.2. *Algorithmic.* Parke and Lewis (Lewis & Parke, 1987) use linear predictive coding (LPC) to analyze a recording and extract phonetic parameters which are then used to drive the facial model. Once something has been digitized, it becomes possible to manipulate it by computer. So it is that sometimes Parke and Lewis will resynthesize their audio track, instead of using the original recording, when subtle adjustments have been made, or occasionally just for "synthetic realism" (Lewis & Parke, 1987, p. 64). But, it is for precisely these reasons that a purely synthetic source of speech and image generation is highly desirable.

Hill and Wyvill take the control problem one step further by linking certain facial actions to each phoneme, and then using the input script of a speech synthesizer to drive the animation of a spoken sequence (Hill et al., 1988; Wyvill & Hill, 1989).

2.6.3.3. *Timing Constraints on Synchronization.* There are various ways to generate animation and sound so that the end product is synchronized. Traditional animators record the soundtrack first, time out events by frame, generate the images, and then add the sound to the film. For some unknown reason, perfect synchronization doesn't feel right. Animators often delay the onset of the action to occur *after* the sound, shifting the action to start as many as four to five frames after the sound. On the other hand, lack of synchronization is disturbing as well. Just as delaying exact synchronization one or two frames can make a difference to the overall impression, so can being off by that much.

Synchronizing image and sound with computers is both easier and more difficult. Synchronization markers can be generated by the computer without the need to search for them as traditional animators do. However, tape stretch or clock timing differences can change the *real* time that the soundtrack and the animation take. The discrepancies can build until, at the end of a three minute animation, the image and sound will be out of synchronization by whole seconds (Hill et al., 1988). Synthesizing both soundtrack and animation at the same time is one solution to the problem. Another solution is to record one of the two onto the final tape, time it independently, and generate the other according to the new time.

2.6.3.4. *Emotion.* Many facial models can display both speech and emotions (see (Badler & Platt, 1981; Nahas et al., 1988; Parke, 1974; Pelachaud, 1990; Waters, 1988; Hill et al., 1988)) but almost never at the same time. Some newer systems (Parke, 1989, pp. 35–40) will display emotion on the top of the face leaving the mouth area free to speak, but many emotions must be shown in all areas of the face to be properly identified (Ekman & Friesen, 1984).

Much is known about how emotions are displayed on the face, during speech, through gesture, in social situations, etc. Also, there are many automatic functions such as breathing, blinking, lip biting, etc. that should logically be automated instead of adding to the burden of the animator. Other more advanced actions — such as looking away from the speaker until a point is made, small flashes of honest emotion, the display of false emotion, etc. — are linked to emotional states and social interaction (see (Ekman & Friesen, 1984; Birdwhistell, 1970; Druckman, Rozelle, & Baxter, 1982; Kendon, 1981; Morris, 1982)). These are all possible candidates for automation.

It should also be remembered that emotion affects vocal quality (Bezooyen, 1984). Ideally, the emotional state of the speaker and its attendant physical effects would be accounted for in image and speech synthesis and linked to the controls of the vocal and animation generation software.

2.7. Other Thesis Topics

Many implementation possibilities were considered but not used for various reasons. A few are listed here to provide some insight into the alternatives.

To animate speech precisely, it would have been possible to trace the lip positions of someone saying each of the phonemes (a set of such photographs being available locally) and produce a cartoon line drawing of a face for animating. An almost three-dimensional effect could have been achieved by applying techniques similar to those used to obtain a three-dimensional model from orthogonal photographs, but such a model could never serve as the basis for something more than a talking face.

Precise techniques of animation, that more closely simulate biological processes provide more natural motion, require detailed information about the skull and fasciæ. This type of information is difficult to come by and difficult to generate at the appropriate level of detail. Faces on the other hand are much easier to generate, and many different methods of scanning (most of which don't involve radiation) can be used to obtain this data. A method by which it should be possible to derive skull data from this information is to borrow techniques from forensic science which are used to recreate the face of a victim from a skull. The skull is used as a base and skin thicknesses at specific points on the skull are marked, then the areas inbetween filled in in a systematic manner until the face has been fleshed out. With face data as a starting point, the opposite of this process could be applied, in effect reverse engineering the skull of the individual in question. This data could then be used to make finite element methods and dynamic skin simulation more flexible and general techniques since they could also be applied to any face.

2.8. Modeling Summary

Synchronization of video to audio in speech animation is of critical importance for proper effect, and to promote better communication. Visual communication alone is not nearly as effective as non-verbal and verbal communication together. Non-verbal communication is much more than flapping lips. Every facial action, from the rate of blinking to evidence of emotion, conveys useful information that can help promote better understanding of speech.

Control of a facial model for speech animation is no simple task. Synchronizing speech with animation is an exacting procedure made easier by driving the animation employing the same parameters used to generate speech.

A successful model that synchronizes realistic speech with a realistic human image would be useful in many ways. The results could be used for a more realistic and visually interesting animated human character, a training device for speechreading, a tool for speech therapy sessions, entertainment films, or even psychology experiments.

From a study of related literature, certain conclusions about various aspects of a human character animation system can be drawn.

2.8.1. Surface Model Selection. By choosing splines rather than polygons, freedom of choice with respect to renderers is somewhat restricted. But flexibility and smoothness of the surface, reduction in the amount of data that needs to be stored, and avoidance of certain artifacts of animation are gained. By choosing hierarchical B-splines, a further reduction in the amount of data along with the ability to add details or modify the structural model at will without affecting the underlying data structures can be achieved.

2.8.2. Actuation Model Selection. By choosing the muscle model process in preference to more area-specific routines for actuation, freedom from the actual geometry of the structural model can be gained. By choosing the muscle model process over a dynamic skin model or finite element model of flesh, a certain amount of realism is lost, but less structural information and less computation are required to use the model.

2.8.3. Speech Model Selection. Writing a speech synthesis system is an entire Masters' or Ph. D. thesis in itself (Schock, 1993), and creating one was never part of the intent of this work. While a system exists locally, and systems are available from the Internet[7],

[7]See the comp.speech FAQ for more details. This list of questions is maintained by Andrew Hunt who can be contacted as follows:

any speech synthesis system would have to be significantly modified to drive a speech animation system. Unless an articulatory synthesis system were available, extensive testing and work would be required to generate the viseme data needed for facial configurations and animation sequences for a new model. For this reason synchronized speech synthesis was placed outside of the scope of this thesis, but every effort was made to facilitate the addition of these functions, should it be decided to implement them at a later date.

2.8.4. Animation Model Selection. Generating an animation through algorithmic techniques, as opposed to depending on performers, or sequencing actuation parameters by hand, delivers more flexibility, freedom, and power to the user.

2.8.5. Intent. In combining the above choices, an extra dimension of power can be realized. Hierarchical B-splines provide a structural model to which details can be added and modified at will, leaving the underlying data structure untouched. Muscle model process actuation provides motion unfettered by the underlying geometry of the structural model. Algorithmic techniques of animation and proper parameterization will mean that motion can be specified in a natural, high level manner. This means that a user could, with a single script, animate virtually any face without changing a single line of code or modifying the control parameters.

Andrew Hunt

Speech Technology Research Group

Dept. of Electrical Engineering

University of Sydney, NSW, 2006, Australia

Ph: 61-2-692 4509

Fax: 61-2-692 3847

email: andrewh@ee.su.oz.au

Chapter 3

Langwidere: A Face Model with Pull

Now I must explain to you that the Princess Langwidere had thirty heads — as many as there are days in a month. But of course she could only wear one of them at a time, because she had but one neck.

L. Frank Baum
page 79 of
Ozma of Oz

There's no art
To find the mind's construction in the face.

Shakespeare
Act I, Scene 4
Macbeth

3.1. Whence the Name

Langwidere (Baum, 1980; Wang, 1992) is the name of a princess of Ev (a kingdom adjacent to Oz) who could change her head like hats. As the story goes, she was not actually an evil person, only very vain and narrow-minded.

The Langwidere modeling and animation system implements a hierarchical spline structural model with FACS-based parameters controlling a simulated muscle actuation model. A language derived from C, called Ev, that can be interfaced with a speech synthesizer, is used to generate animations.

This chapter contains a more detailed hierarchical B-spline discussion, features of hierarchical B-splines and how to build a head with them. Waters' muscle model and appropriate biological counterparts will be analyzed. The Langwidere facial animation system will be described in Section 3.6, and Ev, the animation scripting language, will be planned out.

3.2. Hierarchical B-Splines

As explained in Section 2.2.2.2, hierarchical B-splines represent a smooth, flexible surface efficiently. This section shows how that is possible, how they can be used, and what limitations exist. A description of the hierarchical B-spline editor, Drag, is included in Section 3.2.2.2 and a discourse on how hierarchical splines and the human face are particularly suited to each other is presented in Section 3.3.

3.2.1. Traditional versus Hierarchical B-Splines. A cubic B-spline patch is defined by sixteen control points (see Figure 3.1) that also delimit a convex hull which entirely contains the patch. The actual surface does not intersect any control points unless they are repeated (occupy the same location in space), and only "exists" near the interior four control points.

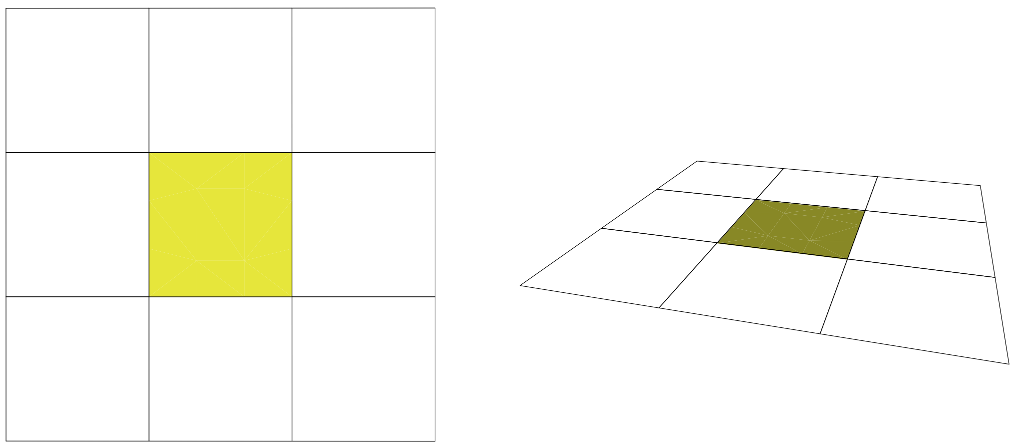


FIGURE 3.1. Example of a single patch. The shaded area shows the surface of the patch. The control points exist at the grid line intersections.

To extend the surface, more control points are added (see Figure 3.2). Each group of sixteen vertices defines a single patch.

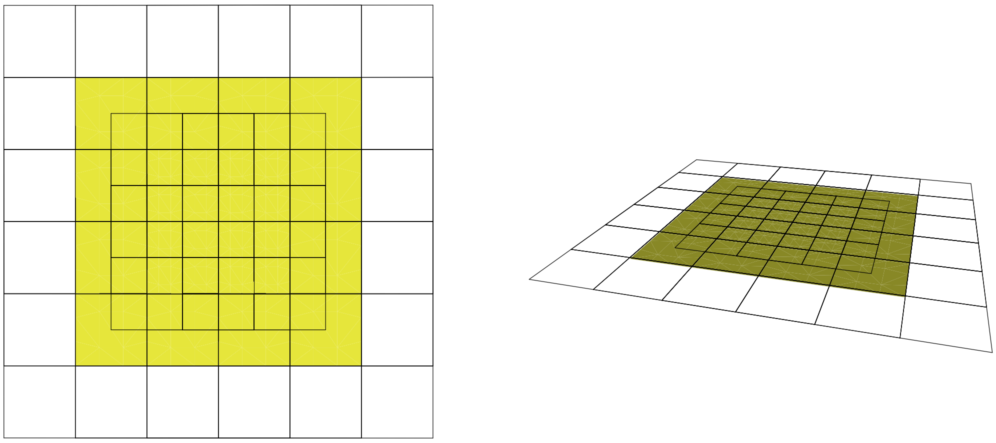
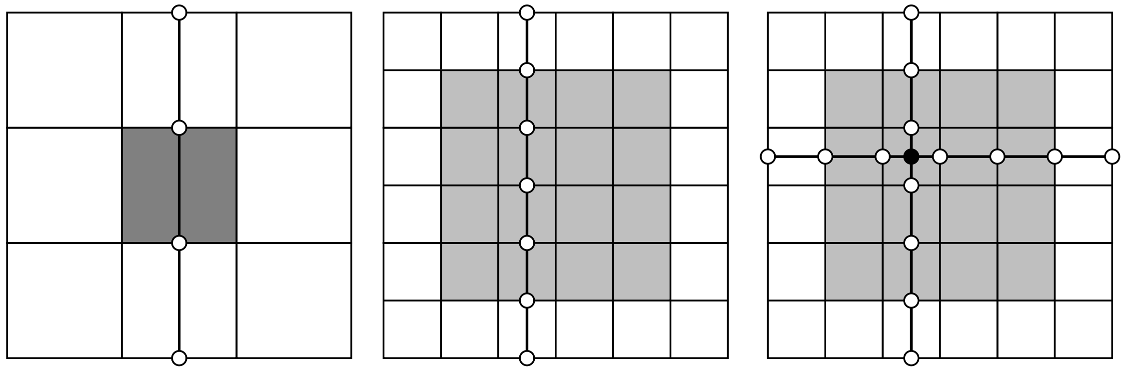


FIGURE 3.2. Example of a refined patch.

While spline patches can effectively supply an infinitely flexible surface [1], only so much detail can be defined by sixteen control points. To increase detail, more control points must be added (see Figure 3.3(a)). It is possible to add control points to an existing surface by using the Oslo algorithm which inserts new control points into the control array that defines the surface, splitting the surface without changing its shape (Cohen, Lyche, & Risenfeld, 1980). When dealing with a continuous surface, all vertically and horizontally contiguous patches must be similarly split to maintain surface integrity (see Figure 3.3(b) & (c)). This adds control points where they are not necessarily wanted.

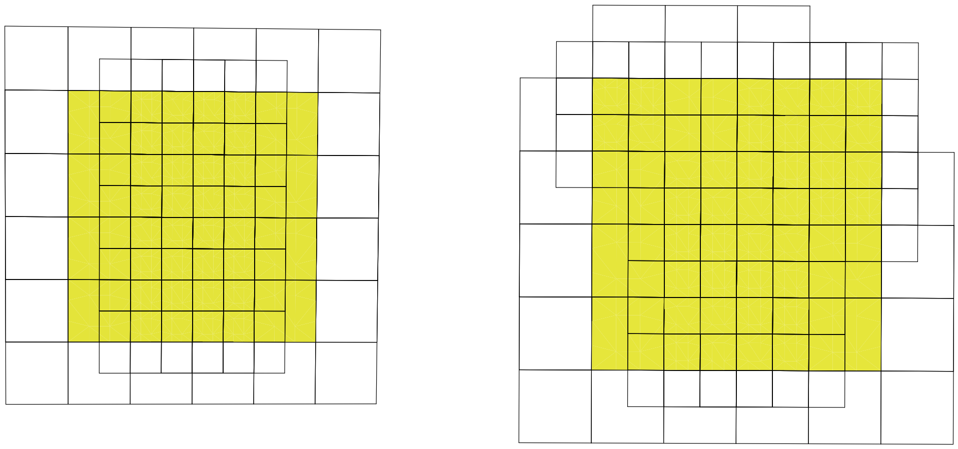
By using hierarchical splines, detail can be added only where it is desired (see Figure 3.4). This is achieved by overlays of more finely subdivided spline patches that conform to the surface to be refined. Changes to a single control point affect the surface up to two control points distant, thus changes to a single central control point will affect a grid of 5×5 control vertices. By increasing the size of this grid one line of control points on each side, defining overlays to be a grid of at least forty-nine control vertices (7×7) and restricting changes to the centre control point, the refined surface can be guaranteed to conform smoothly to

[1] A polygonal surface can only bend or flex at its vertices and edges, unlike a spline patch that can bend at any point on its surface.



(a) (b) (c)

FIGURE 3.3. Examples of split patches



(a) (b)

FIGURE 3.4. Examples of hierarchical subdivision

the underlying surface at all times, creating a single continuous surface with continuity in both first and second derivatives.

Once an overlay has been defined and moved from its original state, it may later become necessary to adjust the underlying or supporting surface. By defining the overlay control point positions in terms of the supporting surface, instead of just location in space, surface integrity can easily be maintained. This is termed "reference plus offset" representation (see

Figure 3.5).

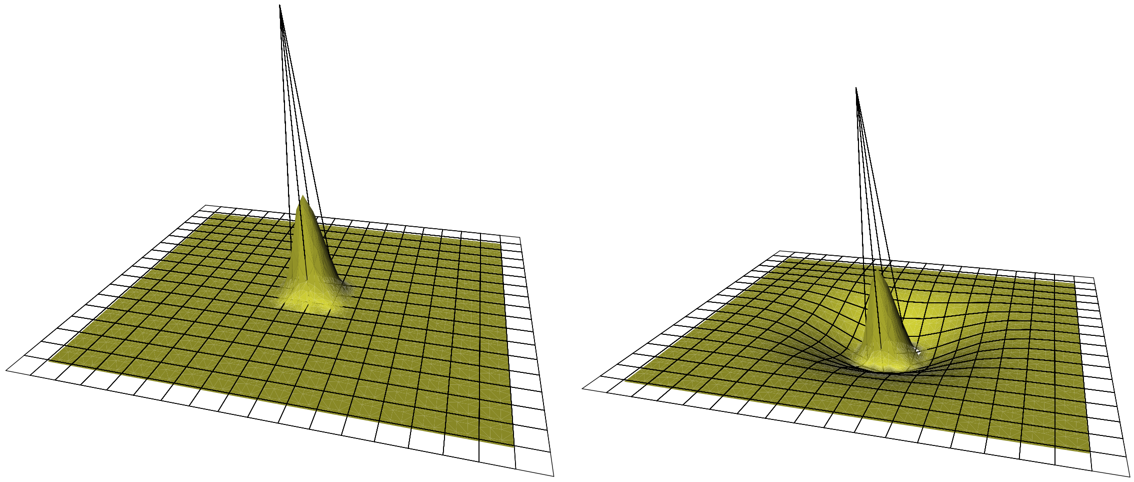


FIGURE 3.5. Example of overlay behaviour

3.2.2. Surface Generation/Form and Conformation. Since hierarchical splines are an enhancement of traditional splines, they can be used like traditional splines without difficulty. This technique, of course, would not take full advantage of the extra capabilities that hierarchical splines afford the user. Further, certain features of hierarchical splines can be abused. The following section explains how to generate a surface, specifically the human face, with hierarchical splines.

3.2.2.1. *Surface Fitting.* Routines exist for fitting splines to surfaces (Flannery, Press, Teukolosky, & Vetterling, 1992; Burden & Faires, 1985; Cheney & Kincaid, 1985) and by adding surface refinement when the divergence of the spline from the surface exceeds some tolerance, the same algorithms can be applied to hierarchical splines. There are two problems in applying splines to face modeling.

(1) While the final surface may look close to the original, the underlying levels don't. They may look like a smoother version of the target surface, but subtle variations in the surface means that when one attempts to animate it, the surface can behave in

unexpected and unacceptable ways. This problem with surface fitting has recently been solved at UBC (Forsey & Wang, 1993).

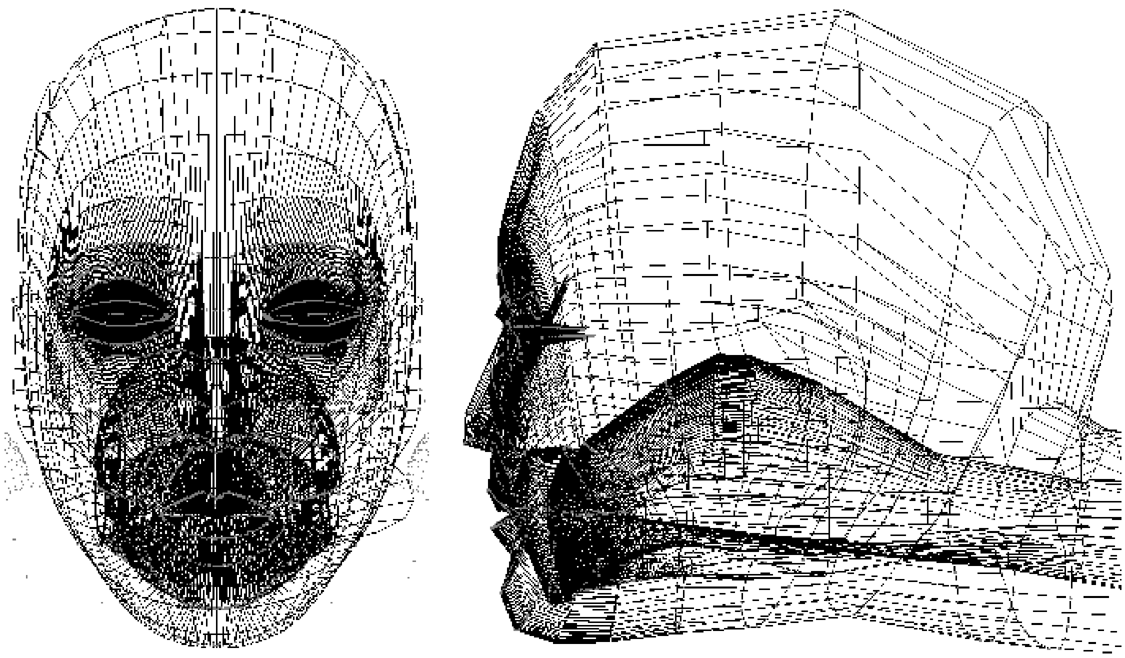
(2) The body is a closed surface. In order to use a closed surface for the model of the body and avoid the north pole problem, a torus should be used as a base form for the surface. Especially since "A human or animal figure is essentially a torus, where the hole of the torus extends from the mouth to the anus as the *alimentary canal*" (Forsey, 1990, p. 59). Ensuring that the mouth matches the opening of the torus during the fitting process is an unsolved problem.

Solving either of these problems properly would each require another Masters or even Ph. D. project. This means that to generate a face for this thesis, it must be created by manual editing.

3.2.2.2. *Editing Patches.* David Forsey originated hierarchical splines in his 1990 Ph. D. thesis (Forsey, 1990) and also wrote Drag, a system for creating and editing hierarchical splines, which he generously made available, complete with source code, for this project.

There are two control structures built into Drag. The first provides control points for spline surface definition and the second comprises "skeleton" segments, or movable frames of reference, to which control points can be bound. Once a control point is attached to a skeleton segment it will rotate or translate with that segment. In this way bone structures such as joints can be quickly simulated.

To build a spline surface in Drag a user begins with one of the basic forms (plane, cylinder or torus). Control points on the surface can be moved individually, or in groups. Groups of points can be moved as a single locked unit, or proportionally relative to a reference point. Since overlays must be of a minimal size, patch refinement or subdivision occurs *around* the selected control points. The patch to be refined is subdivided, horizontally and vertically, and new control points are generated equally in both directions. If the entire surface is refined repeatedly, the number of control points grows exponentially, care must be taken to take maximum advantage of the lower levels so that the surface need not be refined unnecessarily, or the number of control points may quickly get out of hand (see Figure 3.6).

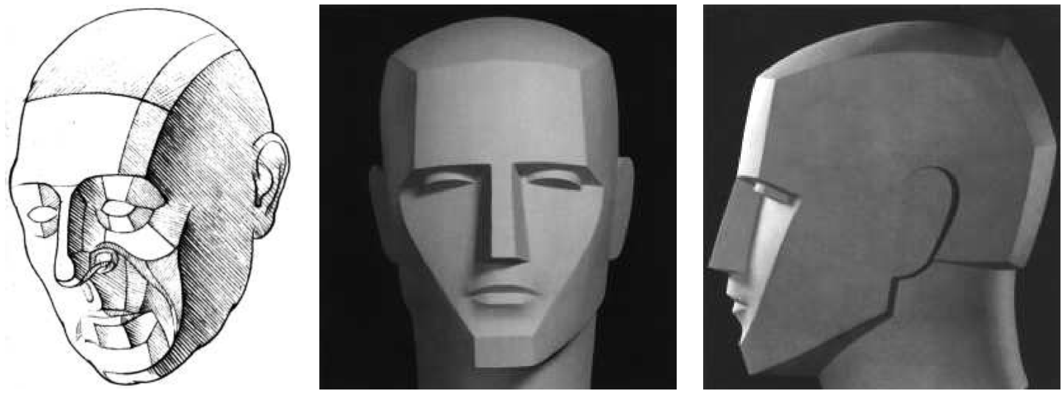


(a) (b)

FIGURE 3.6. An early attempt where refinement was handled badly. Every other intersection of lines in these diagrams is a control point.

3.3. Modeling Faces

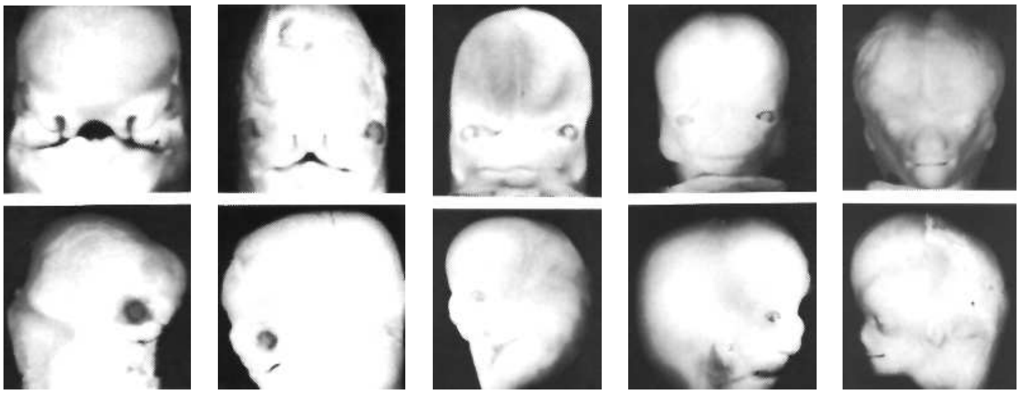
Hierarchical splines grow, fine layers over coarse layers, gross shapes under small details. Clues as to how this evolution of form might progress can be found in the developmental detail of embryos and mass conceptions of figures in art. The human head is a complex shape which may, in the beginning, bewilder the modeler. By analyzing the curves of the face and simplifying them to "planes" (see Figure 3.7) it is possible to gauge more accurately where the gross details are, and from them where to begin. Images of embryo development can also show which areas need to be constructed first. The use of knowledge related to embryo



(a) (Gordon, 1983) (b) (Goldfinger, 1991) (c)

FIGURE 3.7. Analyses of the planes of the head

development is fairly obvious, once pointed out, but it is easily overlooked and therefore worth emphasizing. Figure 3.8 shows a sample sequence of embryo development which can be compared to the development of the head model. An examination of volumetric



35 Days 37 Days 41 Days 45 Days 8 Weeks

FIGURE 3.8. A sequence of images from the development of a fetus (Kraus, Kitamura, & Latham, 1966)

abstractions of the face (see Figure 3.9) can also give hints as to how much detail will be needed to start the surface.

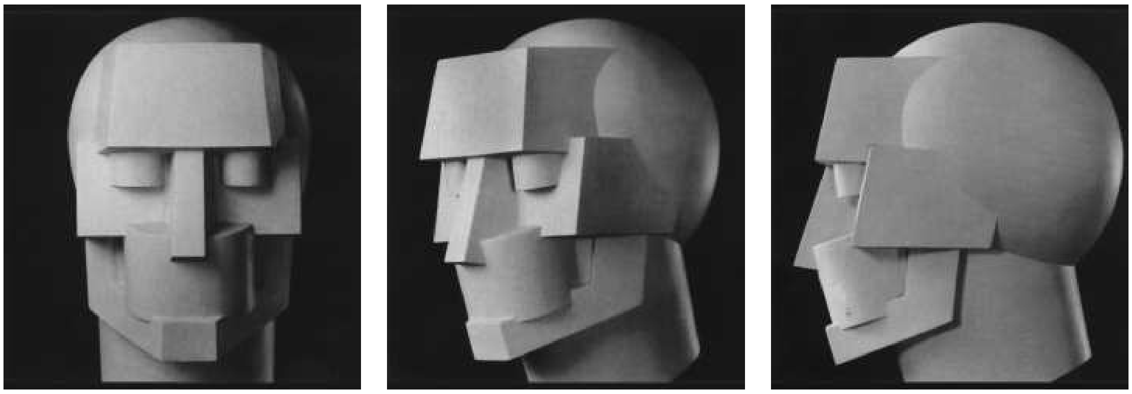


FIGURE 3.9. An analysis of the underlying volumes of the head (Goldfinger, 1991)

3.4. Modeling Muscles

While there is always the brute force technique of hand editing points to simulate movement and animation, nature's muscle actuation method is much more elegant if we can manage it. The next section will cover the relevant details of the biological aspects of muscle-skin interaction, as well as the process of simulating these interactions.

3.4.1. Anatomical Detail of Muscles. There are two different kinds of muscles: skeletal and smooth. Smooth muscles occur internally in areas over which we normally have no voluntary control such as the the bowels, veins, and arteries. Skeletal muscles are also called "striated" due to their striated appearance under the microscope (Smith, 1990). These are under voluntary control with the exception of the heart, which is composed of striated muscle but is not normally a voluntary muscle.

Muscles are bundles of muscle fibres, surrounded and connected to bone, cartilage, and skin by fasciæ and tendons. While different, both fasciæ and tendons are inelastic, collagenous, and fibrous tissues. Tendons connect muscles to bones or cartilage. Very wide tendons are called *aponeuroses*. There are two types of fasciæ: deep and superficial. Deep fasciæ form thick protective sheaths that surround each individual muscle, enveloping each muscle group, separating the muscles, holding them in place, and controlling the amount of friction between them. Superficial fasciæ can be found just under the skin, over most of the body. While deep fasciæ form a dense membrane, superficial fasciæare usually thin meshes holding globules of fat. Superficial fasciæ connect the skin to the underlying structures (the strength of the connection being determined by the thickness of the fascia layer) and serve as a conduit for nerves and blood vessels. It is by means of these structures that muscles are connected to skin, bone, and cartilage. Deep fasciæ may be indistinguishable from tendons.

In general, striated muscles, such as those on the face, will be attached to bone or cartilage at one end, and connected to bone, cartilage, or skin at the other. The end attached to bone or cartilage is formally called the attachment point or origin point. The other end is called the insertion point. When both ends are attached to bone this terminology can be confusing[2], but since this is irrelevant to the muscles of the face other than those associated with moving the jaw (*Masseter*, *Temporal*, and *External* and *Internal Pterygoids*), we can clearly define these terms as follows:

Attachment Point: The end of the muscle affixed to bone.

Insertion Point: The end of the muscle connected to skin.

There are various types of muscle, defined by the arrangement of their muscle fibres, and in many cases the attachment or insertion points will not be a *single point*. Nevertheless, it is convenient to use this naming convention since, in program code, points in space will serve as reference for both areas of attachment and areas of insertion.

The types of muscles that can be found in the face are as follows:

Sheet: A quadrilateral muscle with wide areas of attachment and insertion where the muscle fibres run parallel from one end to the other. A thin sheet, with a small area of attachment, is a strap.

Triangular: A muscle whose fibres do not run parallel, but tend to converge to a relatively narrow point of insertion or attachment, but not both (which would be a *fusiform*

[2]When both ends of the muscle are attached to bone, the attachment point is the "fixed" point, and the insertion point is the end that moves when the muscle is contracted. With an action such as a chin-up, many of these relationships are reversed, but again this is not something that often happens with facial muscles, so the point is moot.

muscle like the bicep, something not found on the face).

Sphincter: An elliptical muscle whose fibres are arranged in curved bundles.

A muscle generates motion when the muscle fibres contract and thicken, shortening their length, sliding the fibres over underlying bony structures (*e.g.* skull, cheekbones, jaw, etc.) if appropriate, pulling the muscle's two ends together and causing the middle, or belly, to bulge. The area most affected by this action is the insertion point. To limit the range of movement of a muscle, we use the fact that "Typical vertebrate locomotor muscles are permitted by skeletal constraints to shorten through about 25 per cent (maximum) of their extended lengths" (Pennycuick, 1992, p. 22-23), although for the purposes of animation it is often desirable to exceed natural limits to allow exaggerated or dramatic presentations, as the artist may require.

Simulating the effects of muscle on skin is complicated by the fact that skin is a highly complex organ with many layers.

3.4.2. Anatomical Detail of Skin. On the surface, skin or integument is a layer of dead, stratified cells that protects the more delicate tissues underneath. Below is an intricate structure of hair and hair follicles, sweat glands, sebaceous glands, nerves, capillaries, and collagen and elastin fibres over, and intermingled with, fatty tissue (Waters, 1990; Gray, F.R.S., 1977; Pieper, 1991). At this level skin is a hopelessly involved, nonhomogenous object to model. At a slightly larger scale it can be perceived as a smooth, somewhat elastic surface. The elasticity and capacity for stress is affected by age and the thickness of the underlying fat layer. Some areas have different qualities due to frequent subjection to stress (Langer's lines). Force transmission from the muscles, as well as the visual effects of their actuation, are all propagated through the skin and subcutaneous layer of fat.

3.4.3. Muscle Model Process. Waters' Muscle Model Process models the effect of muscle contraction on skin. The surface model he uses is a polygonal mesh. In this model, muscles are simulated by approximating the shape of the muscle when flexed, from the point of insertion to the point of attachment. Around the sides[3] of the muscle, there is a natural tapering from the bulge of the muscle's belly which easily blends into the default shape of the surface, but at the point of insertion, where maximum displacement has occurred, the distortion of the surface is too great. Attached to the end of the muscle by superficial fascia, real skin would be pulled towards the attachment point by the muscle contraction, and pushed upwards by the thickened belly. Where forces generated by the muscle are less than normal skin elasticity, effects of the contraction will no longer be seen. Natural force propagation through the flesh must be replaced by a "fall off" function.

Functions define the shape of the muscle based on type. In Waters' system there are three types:

Linear: A triangular muscle where the convergence occurs at the point of attachment.

Sheet: A quadrilateral or strap muscle.

Sphincter: An elliptically shaped muscle that pulls towards the centre.

The area of influence for each muscle is defined by its shape plus a fall off zone located at the edge of the affected area near the insertion point. The effect of a muscle is to deform all parts of the surface within its area of influence which is defined by its attachment and insertion points. These are points on or near the structural model's surface in three space. The deformation effect is scaled by a parameter that determines how far the muscle will contract. Since most muscles are of different lengths, a maximum parameter value must be determined for each of these. At this point there are no forces involved. The amount that a muscle will contract, and from that how the surface of the skin will be deformed, are determined by this single parameter. Waters' Ph. D. thesis calls for combining forces — from interacting muscles and from the skin. Other natural forces are contributed by the underlying bones, atmospheric pressure, and gravity, but how to determine what these forces are is not clear. In his 1987 article for *Computer Graphics*, Waters states that there

[3]When viewing a muscle from the "outside", let us consider the "sides" of the muscle to be the border of the two dimensional silhouette impressed upon the skin.

is a danger of displacing points on the surface to "impossible" degrees, and recommends "preprocessing the structure" to ensure that this does not occur (Waters, 1987).

By creating a small number of functions that simulate muscles and that actuate any part of the structural model, the muscle model process frees the programmer from the need to create a specific routine to generate each movement that may be desired. With each new structural model, the user need only change the attachment and insertion point locations for each muscle instead of, in code, retuning each of the routines that generate movement for each new head (see Section 5.2 for more discussion).

3.5. Parameterizing Control

Actuators give the face motion, but do not truly animate the model. As with actuation, a brute force technique of animation would be to specify muscle motion parameters for each new task, but this can be unwieldy at best. A hierarchy to control the surface at various levels of abstraction is needed. This section describes the Facial Action Coding System (or FACS) in some detail as well as the possible design of a hierarchy of parameterized controls.

3.5.1. Facial Action Coding System. Psychologists Paul Ekman and Wallace Friesen made a detailed study of the human face. Their purpose was to create a system for measuring and distinguishing visible facial expressions. They chose to give their system an anatomical basis "Since every facial movement is the result of muscular action, a comprehensive system could be obtained by discovering how each muscle of the face acts to change visible appearance." (Ekman & Friesen, 1978b, p. 3)

FACS is designed to record visibly-differentiable facial actions. Actions that produce no visible effect are not in the system, and different muscles that produce the same visual result are lumped into one Action Unit (AU). Conversely, groups of muscles that combine to create a single visibly differentiable result are lumped into a single action unit. Further, a muscle that produces more than one effect will be given more than one action unit (*e.g.*

Orbicularis Oculi). As many as half the AUs are based on a single muscle; the rest rely on groups of muscles to produce the action.

Thirty-three muscles (counting the Orbicularis Oris as one) serve as the basis for 32 AUs. Fifty-eight AUs if you want to include some of the more esoteric groupings, but the muscles involved in those are not listed in FACS (see Table A).

Some of the possible uses for FACS predicted by Ekman and Friesen are:

- Facial signals as punctuation to speech.
- Patterns of facial action as an indicator of clinical conditions.
- Differentiating honest and mendacious behaviour.
- Measuring cultural differences in facial display.
- Determining or inferring emotional states.

3.5.2. Control Hierarchy. It is strange how after hearing revolutionary ideas one often says "Well, of course, that only makes sense ... " So it is with the important idea of parameterizing control (Parke, 1974, 1982). Pick an object that you want to manipulate, decide a range of values which you want to be able to set it to, scale these values between 0 and 1 so that you no longer need to remember what range goes with what object, group objects that frequently work together and can logically be set by a single range. Repeat as necessary.

In effect muscles parameterize the changing positions of the data in the structural model. Muscles, themselves, can also be parameterized. Waters does not scale the controls for his muscles between 0 and 1, but it was done in Langwidere by basing the amount of shortening on a maximum percentage of the length instead of the length itself (see Section 3.4.1). The muscles can be grouped or separated into AUs which lend themselves to specificating emotion. Each of the six emotions acknowledged by FACS: happiness, sadness, fear, surprise, anger, and disgust have multiple hallmark elements, traces of the emotion that appear in the upper and lower parts of the face. Not everyone expresses emotion in exactly the same way, and not every situation calls for a maximum expression of emotion, but genuine emotion generally involves the whole face. One of the easiest ways to tell if someone is concealing or faking an emotion is to see if all parts of the face agree. For instance, someone may have angry eyes while smiling broadly as if happy. Fear expressions used as punctuation in discussing horror films (an *emblem* of fear) may involve only the eyes and brows. These elements of emotion can be constructed from AUs and grouped into parameterized units creating a language that would allow the face to perform real or artificial emotional gestures.

Visemes are another class of actions that could be built from muscle positions.

3.6. Langwidere

The name Langwidere was chosen for this system because it was hoped that, just as Langwidere could choose to wear different heads without impairment of function, so this system might be able to change between different structural head models and control them as easily as she. A short description of the elements used in Langwidere follows.

Hierarchical splines efficiently provide a smooth, flexible surface. The reference plus offset representation allows control points to be referred to without knowing their location in space. This allows the structural data to be used without needing to know the specifics that make generalized manipulation techniques difficult to implement. The muscle model generates movement based only on the placement of the muscle and is independent of underlying geometry. By combining the muscle model and hierarchical splines we can fix muscles to control vertices, then change the structure (while retaining the same overall morphology) without being required to relocate the muscles. Even adding extra details such as warts, antennae, scars or dimples would not require changes as long as they are added as refinements to the surface on levels of the hierarchy above those controlled by the muscles.

Each layer in the construction of Langwidere depends only on the data structures of the previous layer and not the data itself. This means an unprecedented degree of freedom to create and animate new heads without the need to revise the muscle model or any other control systems.

3.7. Ev

As Langwidere rules the Kingdom of Ev, so the animation system and scripting language will collectively be called Ev. Ev is a language derived from C that retains the power of C and adds extra routines to aid in the timing and synchronization of diverse elements. Programmers will be able to use Ev as a scripting language naturally, although the same can probably not be said of non-technical users. Control and behaviour of grouped muscles can be described in functions. Frequently performed actions become subroutines. Short repeated sequences can be made into a library of routines for future use.

Ev does not exist except as a rudimentary text interface to Langwidere. As a scripting language, Ev will be developed below in conjunction with various categorizations of time through the use of example scripts.

3.7.1. Time and Ev.

Time is a critical element. Individual actions must be sequenced and coordinated. Actions and gestures must be given duration. A simple script is presented below that will be used to illustrate time in relation to animation in Ev.

> Hi! Ummmm ... My name is Langwidere.
> You may know Carol, who created me to avoid public speaking.

The above script is a first sketch of the dialog only, completely lacking in context. To provide an emotional context and some stage direction we expand the script below.

> [looks around nervously, waiting to begin.]
> "Hi!"
> [widens eyes and raises eyebrows, simulating enthusiasm.]
> "Ummmm ... "
> [small moment of consternation]
> "My name is Langwidere. You may know Carol, who created me to avoid public speaking."
> [attempts to be serious, but is smiling openly by the end]

By now, the programmer begins to see how facial gestures, traces of emotion, and dialog must be coordinated in parallel. In this script, dialog is the major factor with which the other actions must be synchronized. On the other hand, some actions need to be performed in sequence such as the pieces of dialog. For instance, the scene begins with the character nervous and a pause of two seconds so that it can be seen that Langwidere is nervous. It then becomes happy and says "Hi!", at the same time it eyes open wide for a moment and then return to a more normal state. After this, it becomes serious and says "Ummm ... " at the same time that it draws it brows together slightly and compresses it lips. Simultaneously, it looks first down, and then up to the viewer, and so on.

In the script below, "{ }"bracket actions that occur in sequence, and "[ ]" bracket actions that occur in parallel. Fractions that precede "::" are proportions of time which the function following the "::" should occur. Functions and parameters have been given descriptive names that in a real program would likely be a great deal more cryptic to save typing or increase flexibility.

Pause durations are absolute, and the "Say" function has inherent timing that would be generated by a speech synthesizer and used as absolute durations.

```
main()
{
   SetEmotion(NERVOUS);
   Pause(2s);
   [  SetEmotion(HAPPY);
      Say("Hi!");
      {  2/3 :: OpenEyes(1.0);
         fill;
         1/4 :: OpenEyes(0.7);
      }
   ]
   [  SetEmotion(SERIOUS);
      {  Say("Ummmm...");
         DrawBrowsTogether(0.35);
         CompressLips(0.8);
      }
      {  3/5 :: gaze(down);
         fill;
         1/5 :: gaze(camera);
      }
   ]
   [  Say("My name is Langwidere.");
      {  1/2 :: TiltHead(left);
         fill;
         1/2 :: TiltHead(up);
      }
   ]
```

```
    [   Say("You may know Carol");
        {  fill;
           1/6 :: RaiseEyebrows(0.7);
        }
        {  1/5 :: nod(down);
           fill;
        }
    ]
    [   Say("who created me to");
        LeftBrow(0.5);
        RightBrow(0.0);
        Smile(0.3);
    ]
    [   Say("avoid public speaking.");
        SmileRight(0.55);
    ]
    [   Pause(2s);
        SmileRight(1.0);
        OpenJaw(0.2);
    ]
}
```

The concept of absolute time is straightforward. If you want something to occur for two seconds, such as a pause, then simply specify a 2 second duration. If actions should proceed in sequence, such as pieces of dialog, then use "{ }" to bracket the list of functions in the desired order. If something needs to occur at the same time that something else is taking place (occurring in parallel), but not for the same duration — such as raising eyebrows at the end of a phrase for punctuation — bracket these actions with "[ ]" to indicate that they occur in parallel, but also make the eyebrow raising occur in sequence with filler. The effect of parallel execution, is to ensure that the eyebrow raising will occur at some time during the speaking of the phrase "You may know Carol." The "1/6 :: ... " specifies that the raising of the eyebrows should take one-sixth of the time (relative time) needed to say the phrase. The time filler forces the raising of the eyebrows to the end of the phrase.

3.8. Summary of the System

This chapter related the theoretical basis for the Langwidere and Ev implementations. Some of the pros and cons of hierarchical splines were described. Idiosyncrasies of interaction with the hierarchical spline editor Drag were discussed. The anatomical details of muscles and skin gave clues as to how they might be simulated by computer. Parameterizing muscles that produce visible actions with FACS was presented along with a hierarchical control structure. Finally, the elements of Langwidere in combination were discussed and Ev's relationship to time was explained.

Chapter 4

Langwidere: Making Her Work

The shape changes but not the form;
The more it moves, the more it yields.
More words count less
Hold fast to the center.

Gia-Fu Feng and Jane English (trans)
Chapter 5, page 7 of
Lao Tsu Tao Te Ching

"I wish you could, indeed, free my aunt and her ten royal children," said she, eagerly. "For if they were restored to their proper forms and station they could rule the Kingdom of Ev themselves, and that would save me a lot of worry and trouble. At present there are at least ten minutes every day that I must devote to affairs of state, and I would like to be able to spend my whole time in admiring my beautiful heads."

L. Frank Baum
page 100 of
Ozma of Oz

4.1. The Shape of Things to Come

Before construction on Langwidere began, evaluations of Parke's public domain face model and Waters' Ph. D. implementation were conducted to see if either could serve as the basis of the facial animation system that we wished to construct. When it was determined that the neither was adequate, work on a hierarchical B-spline model was started. The structural model went through three generations starting from scratch each time. Tools for control point manipulation were added when the need arose, and eventually muscles were built into Drag. Muscles based on Waters' Muscle Model Process with a few variations were all implemented in Drag for comparison purposes. The theory and results are presented in

Section 4.5. Creating new heads by modifying an existing model is described in Section 4.4. The state of Ev and the details of generating the animation are given in Section 4.6.

4.2. Pre-Implementation Testing

Since there is no point in reimplementing something that already works well and is available, Parke's polygonal model and Waters' muscle model were analyzed for suitability.

4.2.1. Parke. An earlier, modified version of Fascia was already in use for speech animation at the University of Calgary. The version used was changed to allow bilaterally independent actions. The speech lab's speech synthesizer at that time was already configured to output control information to the Parke model for animating speech. Users of that system were queried as to their opinions of the visual results, and the general response was one of dissatisfaction. The lips of the model could not be properly rounded and were generally not as flexible as necessary.

See Section 2.3.1.1 for an evaluation of the surface and structural model.

4.2.2. Waters. Discussions with Keith Waters and Len Breen lead to acquiring the head model from Waters' Ph. D. thesis implementation. The original implementation was for a VAX/VMS system and rendered with *Prism*. The code needed to be modified for Silicon Graphics workstations using the *GL* graphics library. There were seventeen controls for sixteen muscles. As it turned out, the sphincter muscle was not implemented according to the formulae on pages 155–156 of Waters' thesis (Waters, 1988). Since the existing code was unsatisfactory, these formulae were added to the program by myself and Vince DeMarco. This system was then ported to the NeXT by Vince who also conducted some tests with the new face.

In the end, the lip rounding was still judged inadequate for speech, and the model abandoned.

4.3. Langwidere's First Head

This section describes the development of Langwidere and the creation of a head with the system.

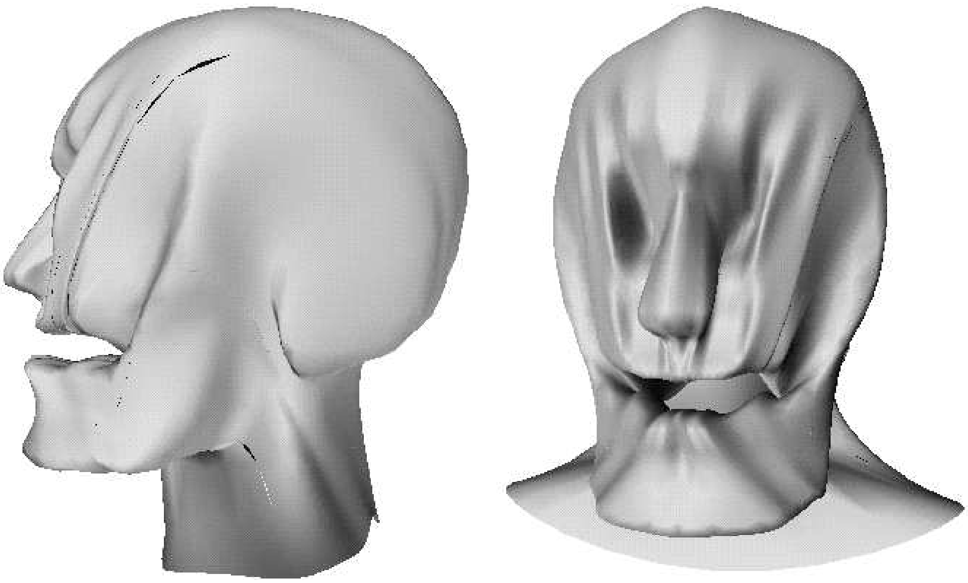


FIGURE 4.1. Al the First, suffering from an unknown degenerative bone disease

4.3.1. Al the First. Since most speech synthesizers only produce adult male voices it was deemed prudent to create the first model as an adult male, although the figure is deliberately somewhat androgynous. The name "Al" was chosen, inspired by the lyrics from the song "You Can Call Me Al" from the album *Graceland*. The second head is female and is called "Betty".

In consultation with David Forsey, it was decided that creating a head model by fitting a spline to scanned data from a real head would not work (see Section 3.2.2.1). Therefore, the head was constructed by hand with Drag. The first decision was selecting a base topology from which to begin. The only closed surface available was the torus, and since the human body is more or less a torus, topologically, this shape was used.

Al the First was extracted from top of a partially generated body created by Dr. Forsey. The head was pulled from the neck stump. Drag, at that time, had no method to move more than one control point at a time, this resulted in significant human error, in part due to the need to pull the opening in the torus down at an angle for the mouth. Thus Al suffered from horrible wrinkling problems (see Figure 4.1). Modeling Al with its head tilted back so that the centre of the torus remained straight seemed to be a solution, requiring only minimal distortion to form the mouth and lips. The head could later be tilted back to a more appropriate position with the use of skeletal segments.

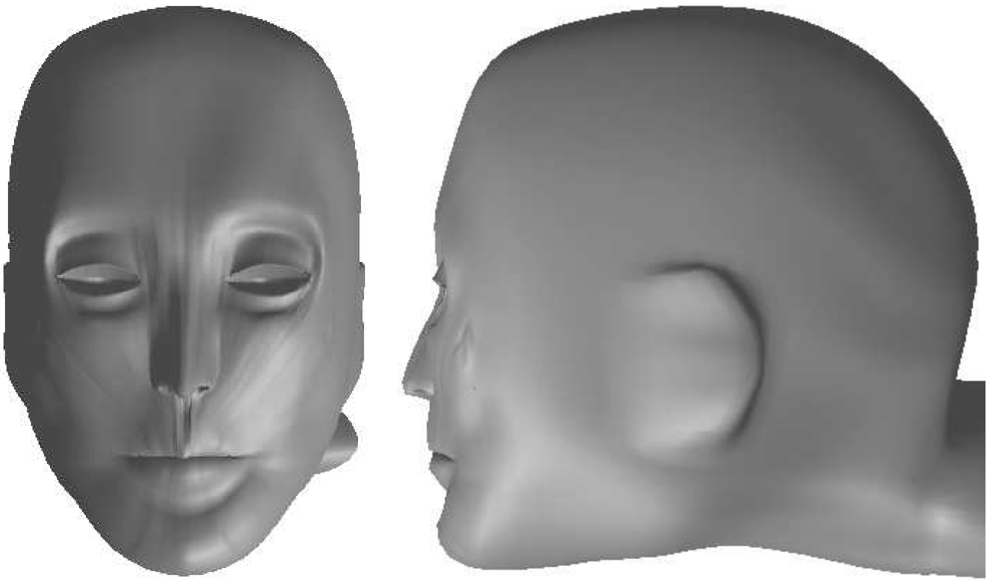


FIGURE 4.2. Al the Second, grew up in a rough neighbourhood

4.3.2. Al the Second. The new Al was begun from scratch, with a pristine torus, taking what was learned from creating Al I to decide what would make editing easier. Eventually, Al the Second was also discarded in favour of a fresh model, but the idea of deriving new models from the second Al, as a base, was tested before it was completely abandoned.

4.3.2.1. *Editing Tips.* Some things that were learned from Al I can only be gained from experience, but this advice may be useful.

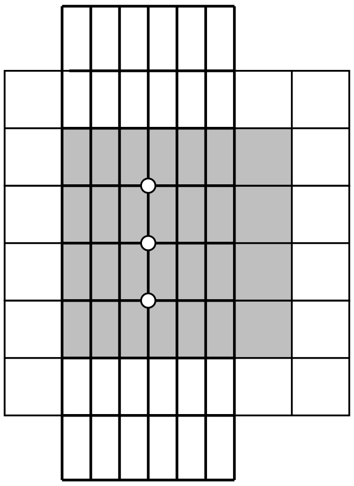
- It is *much* easier to move a lower level surface into a general approximation of form and then fill in (or hollow out) the shape at a higher level of refinement than it is to create the entire shape at a higher level of refinement. There is also a better chance of avoiding wrinkles that way.
- Smooth control lines are a good indicator of a smooth surface. Keep the wireframe grid as regular as possible to prevent unexpected wrinkles or creases.
- Wrinkles are not always the result of "outwardly" displaced control points. Often spacing of the control points can cause problems which evening out the spaces between control points will solve.
- Small wrinkles at lower levels become deep wrinkles after refinement. Make the lower levels as perfect as possible first.
- Rotate the object frequently while editing. Views from odd angles and varied distances can sometimes provide the sense of the complete form that is lacking in a flat display.

4.3.2.2. *Editing Tools.* Two of the greatest problems that plagued the development of Al I were the inability to move groups of control points in a regular, constrained manner and difficulty with relating the point being moved to the surface under construction.

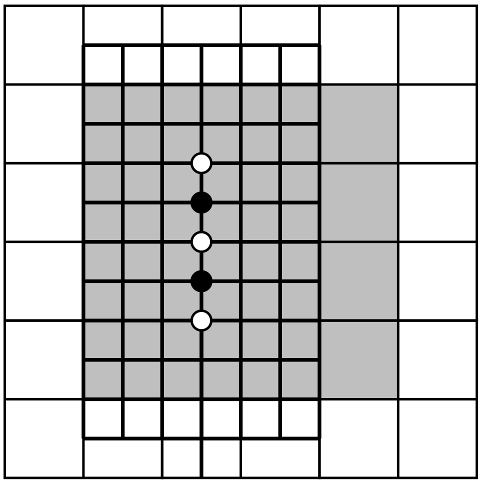
Moving things by hand allows human error; for this reason, automatic bilateral symmetry was already an option of Drag. Dr. Forsey soon added the ability to translate groups of control vertices, all points in the group subject to the same translation. To facilitate the creation of curves and more organic forms, I added the ability to shape a control line into an arc, and fit the control points of an area of the surface to an ellipsoid. The rationale for fitting the surface to an ellipsoid was to quickly create a general skull shape and indentations and protrusions such as the eye sockets and the nose. This proved less useful than originally envisaged since at the level of refinement that these shapes need to be created, the number of control points is small and consequently the shape is easy to form by hand.

Since workstations are not yet fast enough to show more than a simple surface at interactive speeds, the area being modified in Drag is shown as a wireframe. A wireframe is a collection of lines and points suggesting the surface. There are no shadows or gradated shading to help indicate the curve of the surface; no fading or other perspective indicators, except size, show distance. The width of lines is one pixel, the same regardless of distance. No shading obscures what is behind a surface; backfacing surfaces are not removed from the image. It is often difficult to differentiate the lines that compose the front of an object from the back. There is a danger of moving a control point in an entirely unintended direction in this type of environment. For instance, if the desired effect is to translate a point in the y direction (which on the current display is down) and you accidentally move the point in the z direction (which is currently into the screen and is affected by perspective such that the further back a point goes, the further down it moves), as you are moving the point to the visually correct position, you are in fact wildly distorting the surface. In order to prevent this and speed up the modeling process, I set up three orthogonal views based on the current main window view, since looking at the surface from several different views is sometimes required to properly align a control vertex. These windows would update the wireframe view as the surface was edited.

Dr. Forsey continued to develop Drag and regularly provided updates. Since these updates tended to be global changes with desirable characteristics, and did not incorporate any of my modifications, reimplementing them eventually became tiresome. Of the surface editing aids, only the arc conformation tool has survived and has evolved into four types. Two types affect only one of the three x, y, z coordinates at a time, and two types affect all three and make the selected points co-planar. Of each pair of these types, one leaves the end points fixed, and the other allows one of the end points to move.



(a) directional refinement

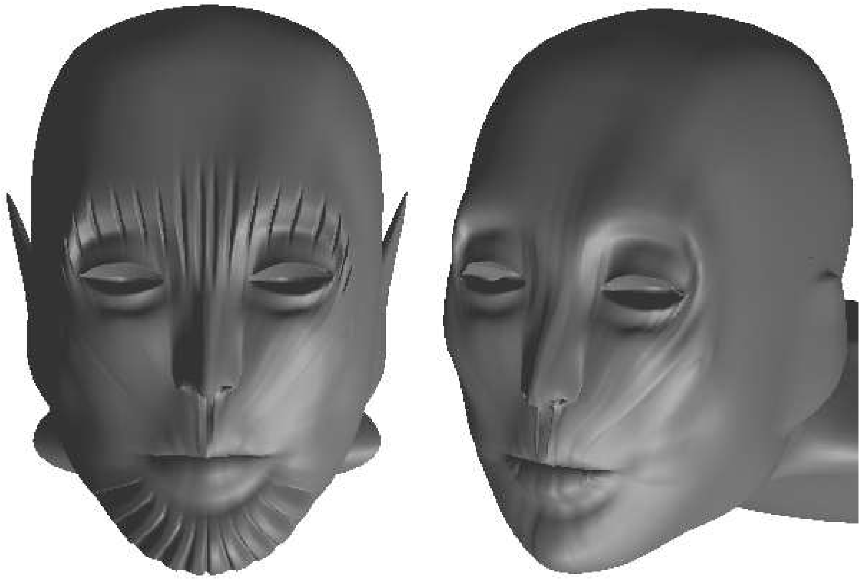


(b) refine around

FIGURE 4.3. Directional refinement versus refine around

4.3.2.3. *Why Al II Had To Die.* Although Al II was a more successful model than Al I, it suffered from a more subtle, insidious kind of wrinkling. Al I's wrinkles were moderately large distortions of shape, Al II's wrinkles were narrow, almost scar-like seams (see Figure 4.2). Further, refinement levels in the model, especially in the eye area, were far too high making it large and slow to use. It seemed that directional refinement was needed. Directional refinement would be the ability to insert new control points between two existing control points only, instead of all around each control point to be refined (see Figure 4.3). This would allow patch subdivision to cause a less drastic effect. But, since this was not possible, an alternate solution was required (see Section 4.3.3).

4.3.2.4. *Al Begins Proliferating.* Before Al II was completely abandoned, an experiment with the model was conducted to see how quickly new heads could be created from an existing one. In the space of approximately 0.5 hrs, an alien variation on Al II (see Figure 4.4 (a)) was made. In another 0.75 hrs, Betty I was born (see Figure 4.4 (b)). The method clearly showed promise.



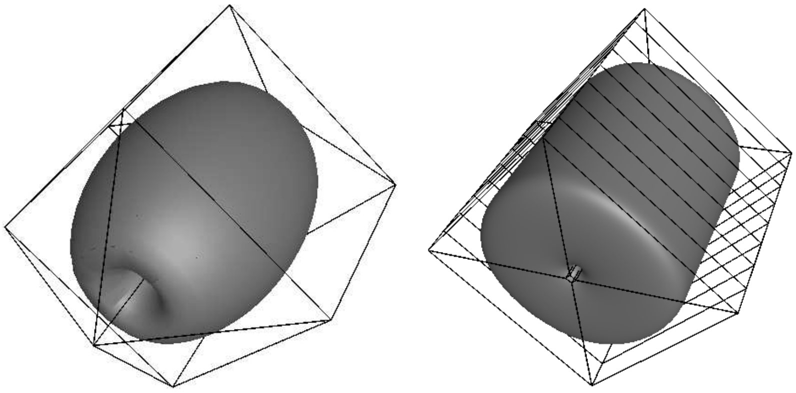
(a) (b)

FIGURE 4.4. Al II's alien offspring

4.3.3. Al the Third. From observations of Al II's development, it seemed that more horizontal than vertical detail existed, and the surface refinement necessary to obtain the control vertices for the horizontal detail produced far too many extra control points. Adding directional refinement was not possible due to the error checking that permeates Drag which ensures that patch levels are consistent. Implementing it would be tantamount to rewriting Drag which contains more than 43000 lines of code.

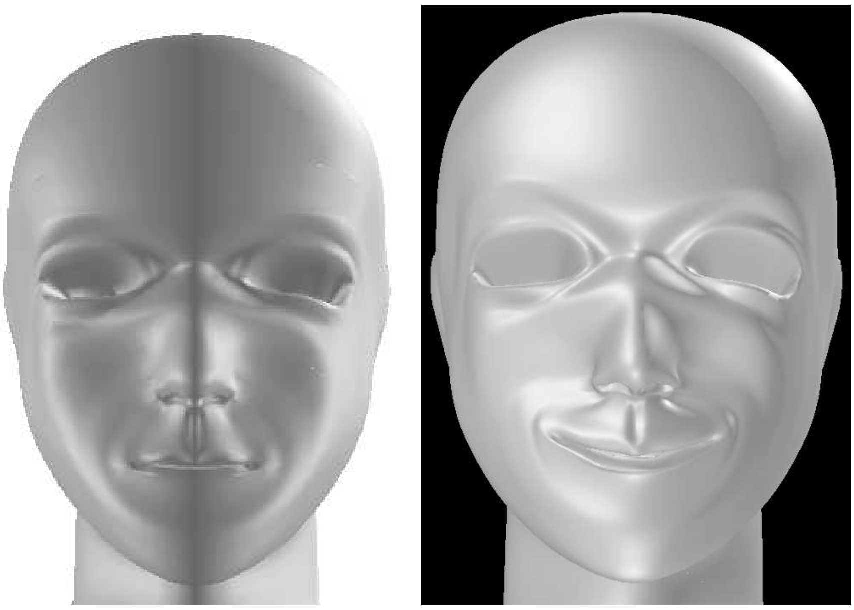
By sketching a body and dividing it where curve segments reached a certain level of complexity, an initial size estimate was generated. The original 4×4 torus (see Figure 4.5 (a)) was replaced with a new one that was 4×18 (see Figure 4.5 (b)).

With fewer control points to misalign, less wrinkling was experienced. Al III was much more successful (see Figure 4.6).



(a) (b)

FIGURE 4.5. Basic tori



(a) (b)

FIGURE 4.6. Al the Third, happy at last

4.4. Betty: Giving Al a Sex Change Operation

The morphology of normal human faces is identical from face to face. Even the proportions vary within only a very small range, so what characteristics allow people to identify sex, age, and race so easily based only on the face? This section outlines some identifying characteristics of age, sex and race, then describes some derivations from Al.

4.4.1. Classic Characteristics. In children, the upper and lower jaws are less developed in relation to the cranium which is growing more rapidly. A child's skin is smooth with a thick layer of baby fat, its orbital cavities more round. As we age, the orbital cavities become more angular, the skin first loses its fat layer and then its elasticity as gravity has its way with the soft tissues, causing the bones to appear more prominent. Lines and wrinkles begin to etch the face and, for some, liver spots appear.

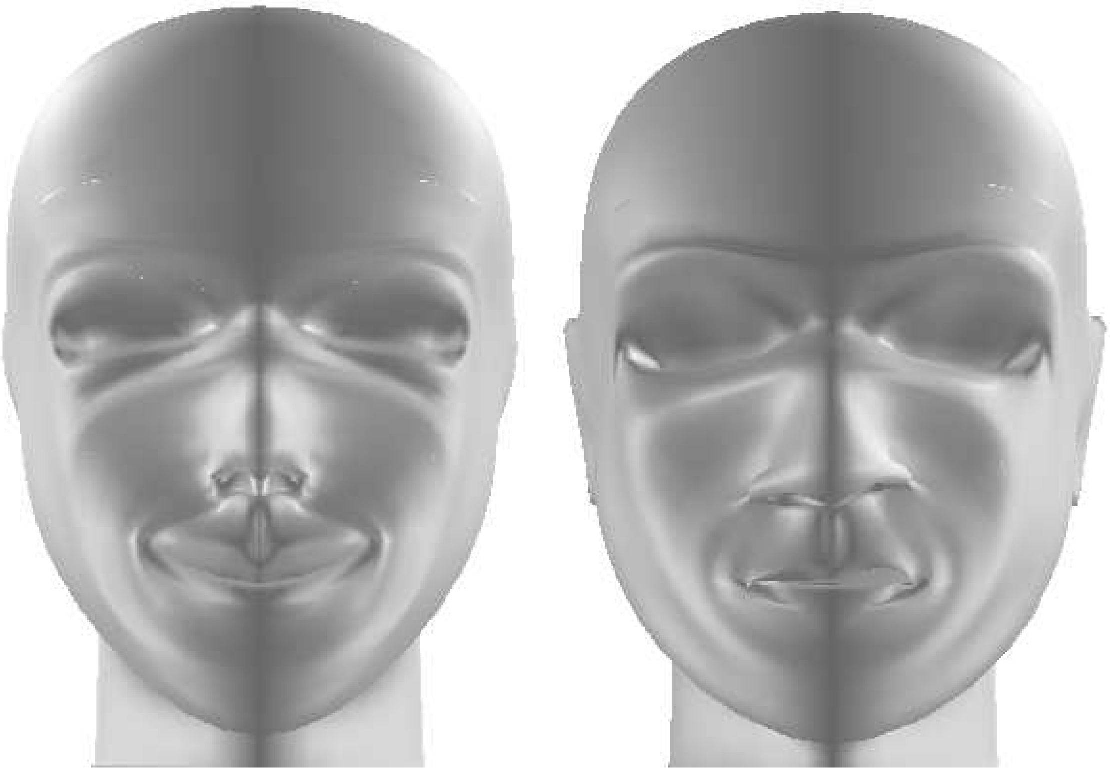
In general, the female tends to have a greater body fat ratio. This means that less muscle or bone definition shows through the skin. The orbital cavity tends to be more round, the jaw smaller and the brow ridge less prominent.

The supraorbital margin tends from round to rectangular, progressing in this respect from Beige[1] to Pink to Brown peoples. Brown people are characterized by wider, flatter noses, full lips, and curly black hair (Gordon, 1980). Beige people are characterized by flatter facial features, eyelids with epicanthic folds, and straight black hair. Pink people are characterized by thin lips, prominent noses, and wavy, varicoloured hair.

4.4.2. Al's Offspring. Betty the First was quickly modified from Al the Second by making its brow ridge less pronounced and cheek bones higher (see Figure 4.4 (b)). Betty the Second comes from better stock, namely Al the Third. It has fuller lips, higher cheek bones, reduced brow ridge, and a smaller jaw (see Figure 4.7 (a)).

Since Al III was deliberately androgynous, a more obviously masculine model, Chuck, was created. It has a larger jaw, more prominent brow ridge, recessed eyes, and a wider mouth (see Figure 4.7 (b)).

[1] Racial terms are very sensitive issues, and the legal and scientific terms for races (Caucasoid, Mongoloid, and Negroid (Flexner, 1988, p. 214, p. 862, p. 891)) please very few. The more common and somewhat less inflammatory terms: White, Yellow, and Black have also fallen into disfavour. For this reason I will instead use terms entirely of my own devising (Pink, Beige, and Brown).



Betty Chuck

FIGURE 4.7. Al proliferates

4.5. Muscular Al

Generating a good structural model is but part of the battle, actuation is the next step. Muscles need to be affixed to the structural model and given an area of influence. The various types of muscles that were implemented in Langwidere, and how they interact with the structural model are described in this section.

4.5.1. Simulated Flesh. In the original Waters system, all the points in the entire surface were checked to see if they were in the area of influence of the actuated muscle and then moved accordingly. This is very inefficient since each muscle has a small area of influence compared to the entire surface. Langwidere generates a list of control points that are in the area of influence as well as a large buffer region, but still much smaller than the entire surface, to allow for points moved out of and into the area of influence by the actions of other muscles.

4.5.1.1. *Muscle Pull and Displacement.* As stated earlier, a fully contracted muscle can shorten to 75% of its original rest length (Pennycuick, 1992, pp. 22–23). When the muscle shortens, its volume remains constant, so its cross-section thickens as its length decreases. The effect of the muscle on the surface of the skin is to pull the surface from the area of the insertion point towards the attachment point. The contraction towards the attachment point is accompanied by a corresponding outward displacement of the surface caused by the thickening of the muscle as well as the flesh (fat and skin) above it.

4.5.1.2. *The Underlying Structure of Flesh.* Bones and cartilage give structure to the face. Flesh and muscles slide over bone on a layer of fat. In many cases, muscles are short and a simple linear model will do, but for the longer muscles such as the frontalis, which slides over the forehead and top of the skull, a simple model will cause unpleasant results like collapsing the model's frontal lobes.

Simulating flesh more accurately including nice effects such as proper wrinkling and furrowing is possible using a physically based, dynamic, finite element model. But such an implementation would be almost useless without a skull model and is beyond the scope of this project.

4.5.2. Muscle Types. *Gray's Anatomy* describes a wide range of muscle types including: rectangular, triangular, pyramidal, fan-shaped, trapezoidal, rhomboidal, and sphincter. Only a subset of these appear in the face and Waters (1988) chose to implement three generalizations that were named linear, sheet and sphincter.

Linear: The linear muscle is fan-shaped. It is defined by its attachment and insertion points and an angle of influence. An example would be the *depressor anguli oris*, a triangular muscle which pulls the corner of the mouth down towards the jaw (see Figure 4.8 (a)).

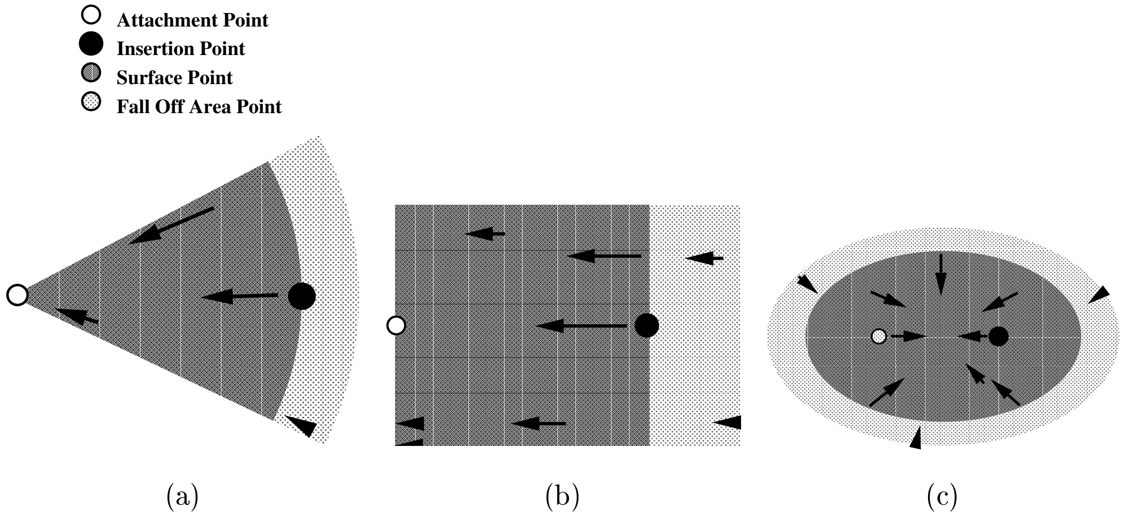


FIGURE 4.8. Muscle types, areas of influence, and pull direction

Sheet: The sheet muscle's area of influence is rectangular. It is defined by its attachment and insertion points and a width. The most obvious example is the *Occipito-frontalis*, the forehead muscle that raises the eyebrows (see Figure 4.8 (b)).

Sphincter: There are only two instances of sphincter muscles in a human face, the *Orbicularis Palpebrarum* or *Orbicularis Oculi* around each eye. These are the blink muscles. A sphincter muscle is elliptical and is specified by two foci and a length that defines its area of influence (see Figure 4.8 (c)).

For a list of the muscles in Langwidere see Appendix B for more details.

4.5.3. Displacement Types. In order to more closely simulate actual physical processes without creating a fully fledged physical model, three deformation types were implemented in this project: skin thickness; proportional; and gradient-based.

4.5.3.1. *Skin Thickness.* Each muscle pulls generally towards its attachment point. The point of greatest displacement is at the insertion point, the effects then taper off to zero at the borders of the area of influence. In the case of a sphincter, the pull is towards the midpoint between the two foci. In the case of a sheet muscle, it is more towards the

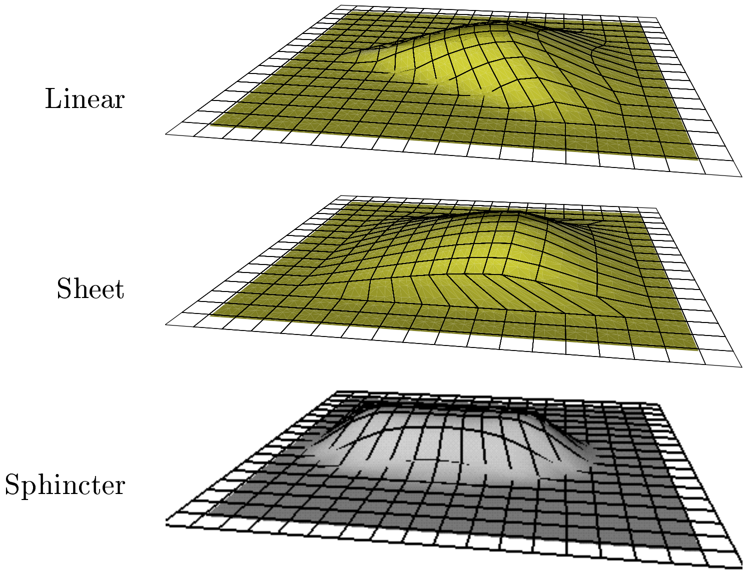


FIGURE 4.9. Muscle contraction with skin thickness displacement

"back" of muscle than towards the attachment point. But, these are more or less planar displacements that do not take into account the muscle bulge.

Areas with larger fat deposits (or greater skin thickness) force more tissue to shorten and occupy the same base area, thus requiring greater outward displacement. Therefore, a "skin thickness" factor was introduced which is used to scale outward displacement along the vector normal to the surface at the point being displaced. The same tapering techniques which are used for the linear displacement are applied to the outward displacement (see Figure 4.9.

4.5.3.2. *Proportional.* Since the skin thickness variable was chosen arbitrarily, another method for determining the outward displacement factor was developed, this time based on the size of the linear displacement. This method was called "proportional" (see Figure 4.10). Selecting a skin thickness factor carefully seems to produce results very similar to the proportional method, but both methods fail miserably in areas of high curvature.

4.5.3.3. *Gradient-Based.* The problem in areas of high curvature is that, while surface points should be sliding along the underlying skeletal structures towards the attachment

Linear

Sheet

Sphincter

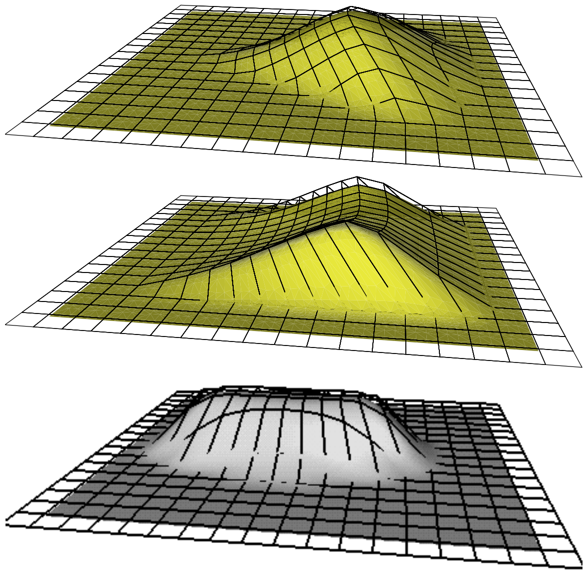


Figure 4.10. Muscle contraction with proportional displacement

Linear

Sheet

Sphincter

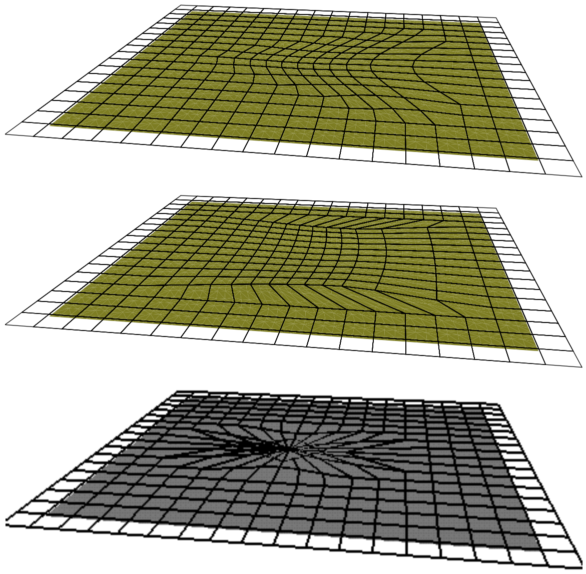
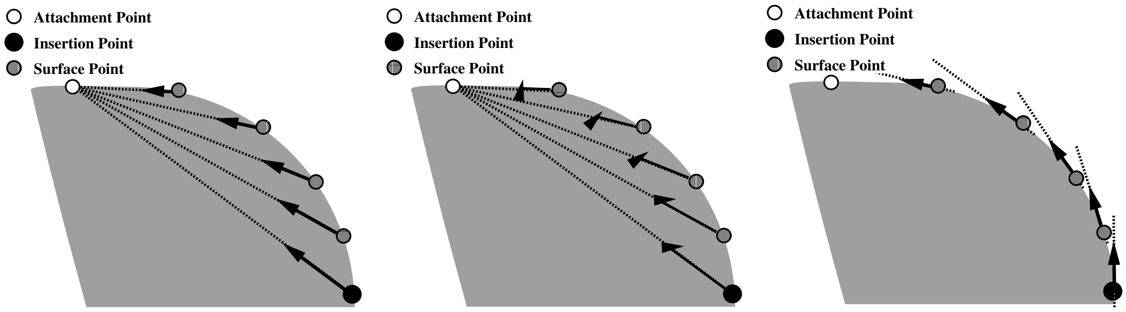


Figure 4.11. Muscle contraction with directional derivative-based displacement (gradient)

point, they are instead displacing in a straight line from the point being displaced to the attachment point, collapsing the surface (see Figure 4.12 (a) and Figure 4.13 (b)).



(a) Standard flat pull (b) Skin thickness pull (c) Gradient pull

FIGURE 4.12. Direction of pull

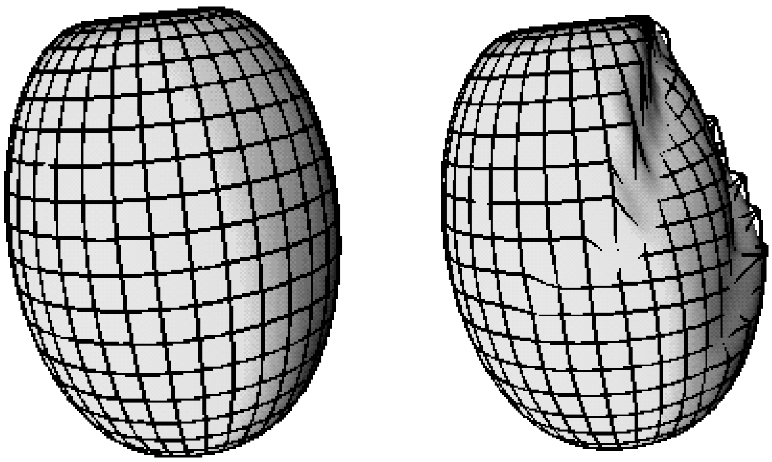
A new method calculates the directional derivative of the surface in the direction defined by the vector between the point being displaced and the attachment point projected to the tangent plane at that point. The point displacement is then scaled by the directional derivative as well as the relative strain rate along the projected vector to produce the final result (see Figure 4.12 (b)).

This method, as might be expected, works well on curved surfaces (see Figure 4.13 (d)), but on a flat surface produces no bulge (see Figure 4.11).

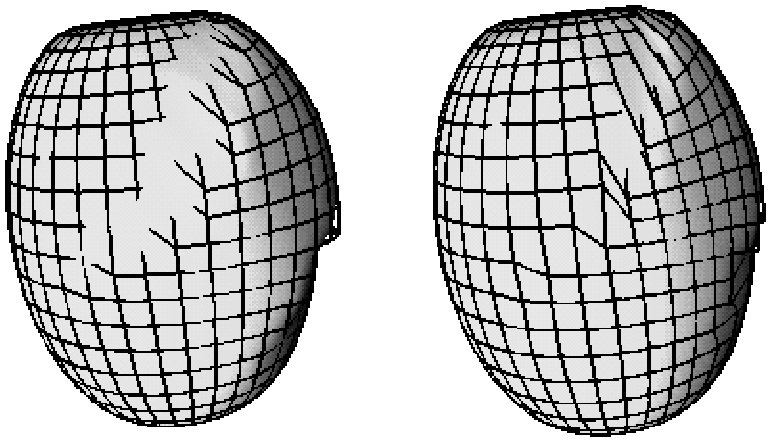
4.5.4. Interaction. More problems occur when muscles interact. If there is no special case handling code at all, the results are not pretty (see Figure 4.14).

4.5.4.1. *Forces.* In Waters' Ph. D. thesis, muscles interact by adding the forces applied by each muscle on the surface point together and then applying the resultant force to the surface point to calculate the final displacement. The problem is that in his surface model there are no forces to add. All movements are purely displacements. It might be possible to cobble together a system that represents skin elasticity and muscular activities as forces, but without a full tissue model this would be as arbitrary as any other model.

4.5.4.2. *The Non-Interaction Method.* In later works, Waters suggests simply recalculating the area of influence for each muscle being activated and accumulating displacements



(a) Initial surface (b) Skin thickness displacement

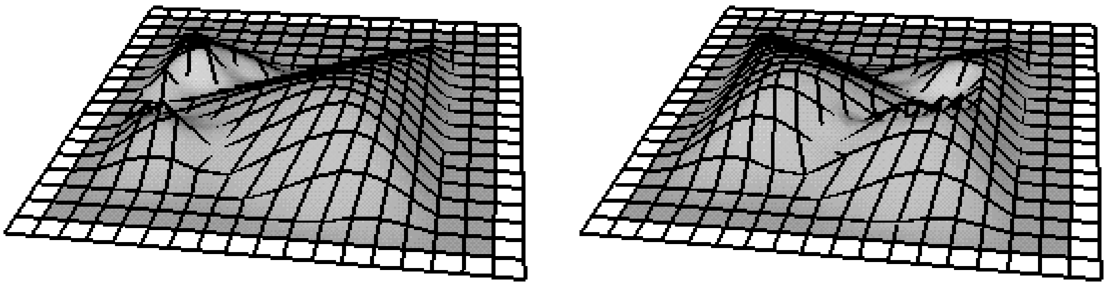


(c) Proportional displacement (d) Gradient displacement

FIGURE 4.13. The effect a curved surface of the various contraction displacement methods. The breaks in the mesh show where control points, at the intersection of the lines, have been pulled beneath the surface they define and help delineate the shape of the muscle.

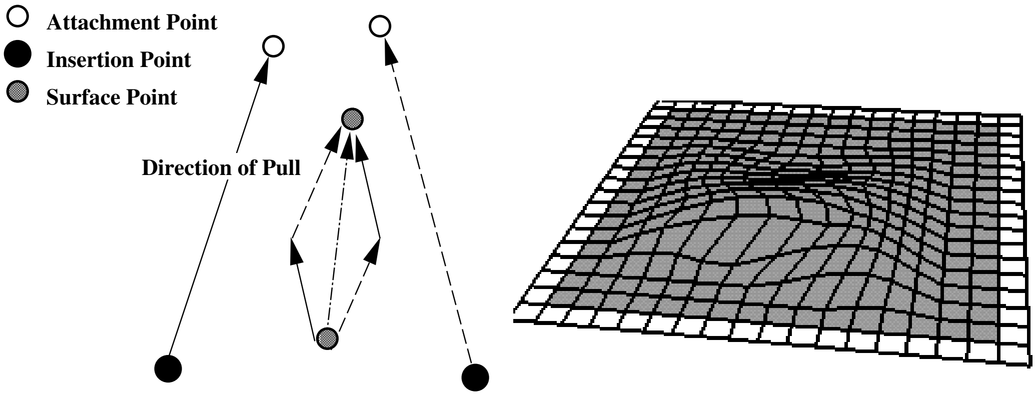
without any regard for interaction. People who have used this system report that it "produces usable results for the major facial muscles[2]" (Franks, 1993). But this is not a very

[2] Recalling that Waters' publicly available Ph. D. thesis model implements 16 muscles in total.



(a) (b)

FIGURE 4.14. Why muscle interaction code is needed



(a) (b)

FIGURE 4.15. Non-interactive muscle pull addition

intellectually satisfying solution. Imagine two identical muscles, close together in roughly the same orientation, pulling at their maximum. When the first muscle pulls, it moves point x to a new location x' by a certain amount d. When the second muscle pulls, point x' is moved to x'' by d' resulting in the point being moved d' further than one might expect (see Figure 4.15 (a)). Also, since Langwidere adds a normal displacement factor, the silhouettes of the muscles remain clearly visible (see Figure 4.15 (b)) when using this method, due to the second muscle displacing outward on top of the first.

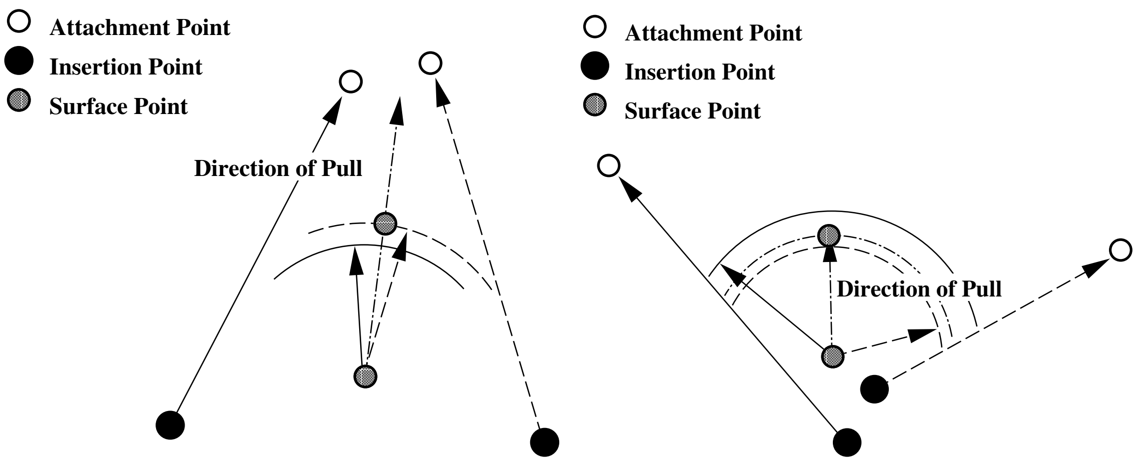


FIGURE 4.16. Adding muscle pull vectors

4.5.4.3. *The Quasi-Interaction Method.* Considering the interaction between two muscles with overlapping areas of influence, it seems intuitively obvious that the direction of displacement for a point in the area of intersection can be found by adding the separate displacements together ($v_1 + v_2$). The problem is with the length of the final displacement vector. Even at maximum displacement, each muscle can only shorten through a certain amount of distance (let us call these d_1 and d_2). It then makes no sense if $|v_1 + v_2| > max(d_1, d_2)$ since their maximum shortening distance is based on the muscle fibres themselves and does not change. Accordingly, Langwidere blends muscle actions by truncating the length of the added result to the length of the maximum individual distance if it is smaller than the added distance (see Figure 4.16 and Figure 4.17 (d)).

4.5.4.4. *Miscellaneous Interaction Methods.* Other ways of dealing with muscle interaction were attempted for the sake of comparison. All tests used the same two linear muscles (see Figures 4.17 (a) and (b)). Sequential displacement of points as advocated by Waters produces Figure 4.15 (b). Trusting to the "natural" spline drop off effects produces Figure 4.17 (c).

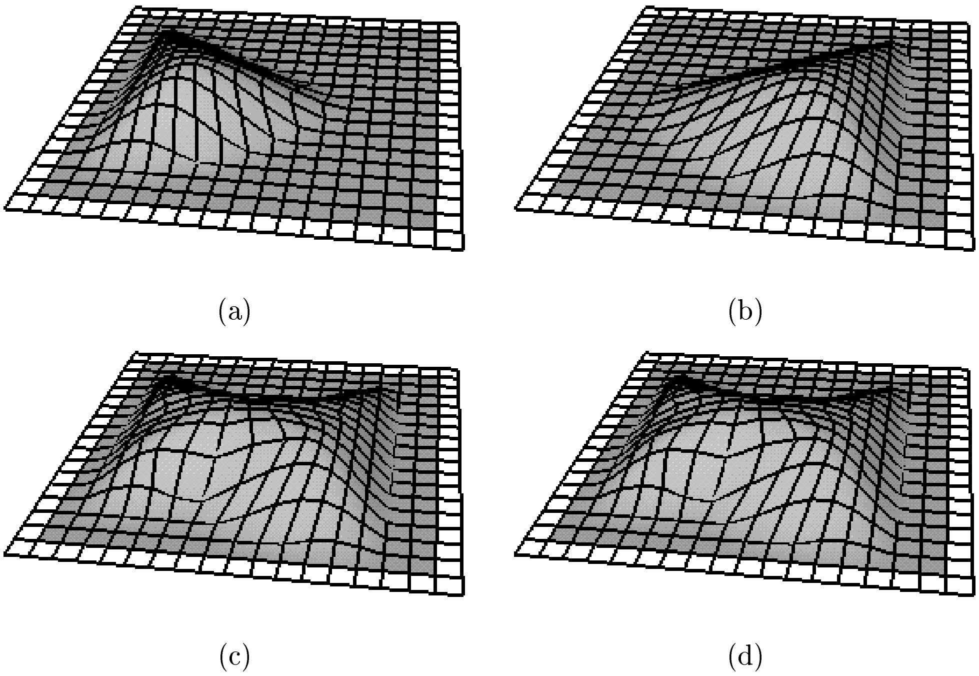


FIGURE 4.17. More muscle pull interaction

4.6. Animation

Animation produced with Langwidere[3] is of two different forms. The unique evolution of form that comes from creating a structural model with Drag was animated. The finished self-actuated muscular models can be animated with changing expressions as well.

4.6.1. Hold Fast to the Centre. Modeling with Drag resembles sculpting clay in certain aspects. While there is no tactile feedback with Drag and only rarely can you see through a clay model to the other side, the development is remarkably similar, except that Drag can reverse mistakes much more easily and is less messy to work with. On the other hand, clay models can be generated anywhere and don't need to be plugged in.

From a starting shape, the surface is pulled out to create forms or pushed in to hollow out shapes. First, an overall structure is assembled, then the surface refined to add more details.

[3]Please write to Carol Wang C/O the University of Calgary, Department of Computer Science, Calgary, AB, T2N 1N4, for information on how to obtain copies.

The shape emerges from amorphousness gaining structure and form. Al's development from torus to more developed states can be seen in the video sequences *Marshmallow Surprise*, *Shape & Form*, and *Inside Shape & Form*, created by placing frames representing successive stages of the modeling process, in sequence, on videotape.

4.6.1.1. *Marshmallow Surprise.* This video sequence clearly shows the original torus and Al's evolution to the point before the cheekbones have been finalized and before any eye detailing had been done.

4.6.1.2. *Shape & Form.* This video sequence shows Al from a different angle focusing more closely on the face which had been further refined from that in *Marshmallow Surprise*.

4.6.1.3. *Inside Shape & Form.* This video sequence shows Al in profile, but cut in half so that the interior structures (oral cavity, tongue, teeth, nostrils, eyes, and ears) can be seen. Unfortunately, patch trimming had not yet[4] been implemented in Drag, and so the surface could not be cut at the lowest level of refinement where the oral cavity was shaped from the inner surface of the torus.

4.6.1.4. *Marshmallow Surprise 2.0.* This video sequence exactly repeats the content from the original sequence, but with a moving viewpoint that shows more of the development of the entire body before focusing on the face which is in a more refined state than when the original video sequence was made.

4.6.1.5. *Al Expresses.* This video sequence shows the simulated muscles and jaw motionn in action. Al is shown smiling, frowning, and evincing surprise.

4.6.1.6. *Chuck Expresses.* This video sequence used the script for *Al Expresses*. One word was changed to save the frames by a different name from that used for Al's animation. To generate the different animation, Langwidere was loaded with Chuck's surface file before the script was run.

[4]Dr. Forsey has stated an intention to implement patch trimming in Drag.

4.6.1.7. *Betty Expresses.* Again, this video sequence used the same script as *Al Expresses* and *Chuck Expresses* with a one word change. Since this script does not include view information, the animation could be produced from a different view point by loading the betty model with colour and changing view in Langwidere before animation script was run.

4.6.1.8. *Blithering Aliens.* This video sequence demonstrates jaw movement and shows some of the oral cavity detail. The forehead muscles move up and down slightly during the cycle causing the antennae to wave.

4.6.2. A Kingdom to Rule. The Ev language is presently a C preprocessor, written with Perl, to allow maximum flexibility. Absolute, relative, sequential and parallel execution times are assigned to actions, and Ev produces an intermediate file with absolutely timed actions. The actions are functions in a larger program that will eventually sequence the timed actions, resolving conflicts where needed, blending actions where appropriate, and producing a list of command events for input to Langwidere.

4.7. Implementation Summary

The Langwidere system is a synthesis of original research, reimplementations of previous work, and modified borrowed source, the breakdown of these is as follows. Langwidere was built on top of Drag, the hierarchical spline editor, which was written by Dr. David Forsey. The local area surface deformation muscle simulation code was reimplemented and modified from the original research by Dr. Keith Waters. Original source code was available but not used in the final system. All contraction types were original ideas implemented for Langwidere. Orthogonal views in Drag was an original implementation for Langwidere but abandoned in the final system. Multiple control vertex spline curve conformation tool was an original implementation for Langwidere. All head models were created with Langwidere by hand, without benefit of scanned data. Surface growth and muscle animation code were original implementations for Langwidere. Animation parameters were generated without scanned data. Ev's syntax has been worked out to the extent that a parser exists, but it is essentially non-functional at this point in time.

Chapter 5

Discussion: Results and Problems

> They found the Princess Langwidere in her mirrored chamber, where she was admiring one of her handsomest heads — one with rich chestnut hair, dreamy walnut eyes and a shapely hickorynut nose. She was very glad to be relieved of her duties to the people of Ev ...
>
> L. Frank Baum
> page 240 of
> *Ozma of Oz*

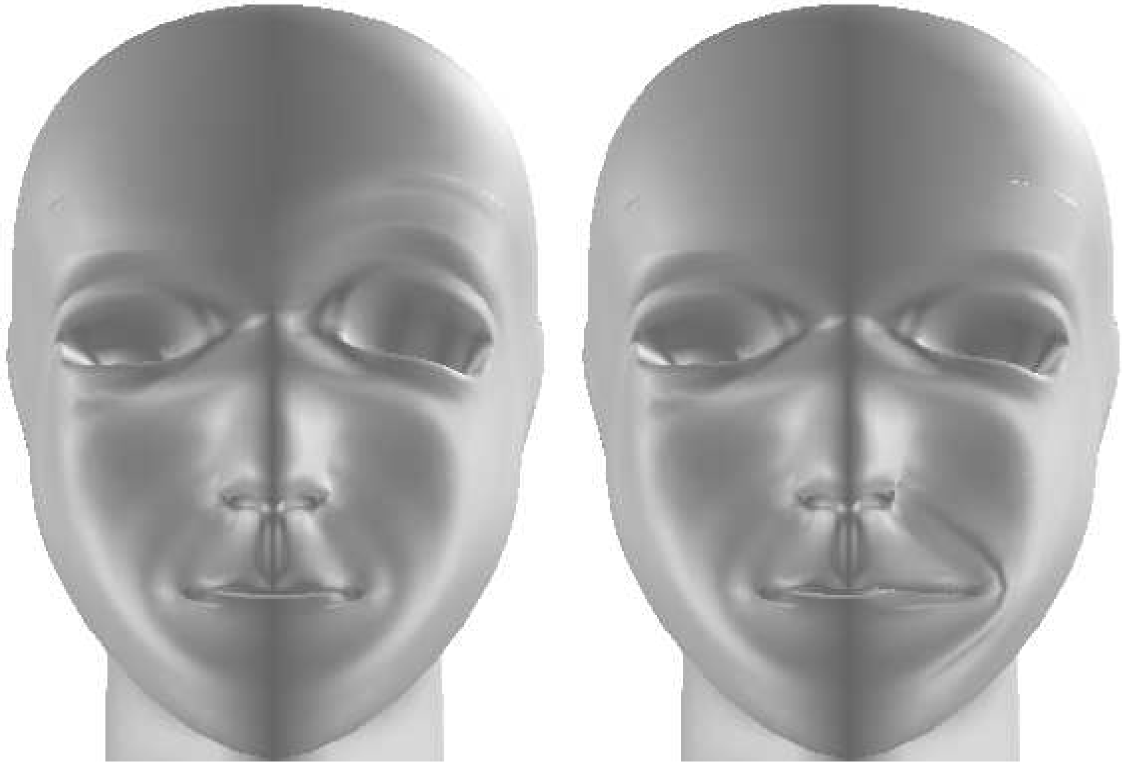
This chapter discusses the results and contributions of Langwidere's implementation, and outstanding problems. The success of hierarchical splines as a basis for modeling is addressed, followed by a discussion of muscles as applied to computer modeled heads. The chapter ends with a short discourse on the animations created with Langwidere, and a summary of the results.

5.1. Hierarchical Splines as Sculpting Tools

Hierarchical splines lend themselves to modeling complex organic shapes, often faster and more efficiently than most other existing modeling systems. If the whole body is to be modeled, a torus — or a cylinder with the ends tucked into the interior of the cylinder — should be used together with hierarchical bicubic B-splines. This both avoids singularities where the spline surface is closed, and more accurately models the topology of the body.

5.1.1. Planning Surface Subdivision. Careful pre-planning of the surface is necessary if the model is to be as intricate as a head. If the surface is poorly designed, the complexity of the model can become unmanageable. What is accomplished by generating a starting surface explicitly for each model can be more easily and flexibly performed by directional refinement. The problem is that directional refinement requires an almost complete rewrite of Drag. However, generating a more appropriate starting surface would then not require further software revisions as were necessary for this project.

5.2. Muscles for Each Task



(a) (b)

FIGURE 5.1. Al raises a brow and grimaces

Waters' Muscle Model Process is a good beginning, but cannot handle all the muscles on the face properly since it does not take into account the underlying skeletal structure supporting the flesh. Publicly available skull data is difficult to obtain, or create, in the variety and volume that superficial face data can be found. As a compromise between "hack" solutions to this problem and full, dynamic finite element models; Waters' Muscle Model Process can be used for short muscles that do not encounter areas with a high degree of curvature. Gradient-based displacement and contraction can be used for longer muscles over areas of high curvature. A muscle type that was missed by Waters, which does appear on the face, is the triangular muscle, a fan shaped muscle with a wide point of attachment and small point of insertion.

The skin-thickness variant of the Muscle Model Process does not behave properly with Waters' suggested non-interaction method of dealing with the confluence of two (or more) muscles. If skin-thickness contraction is to be used, then quasi-interaction of muscles should also be used.

Basing the amount of contraction solely on length (see Section 3.4.1 and Section 3.5.2) does not always have the expected results due to the imprecise nature of the simulation and the inability of ordinary humans to move precisely one muscle[1] at a time. Therefore, some experimentation may be needed for best effect.

5.2.1. Can We Talk? By implementing twenty muscles[2] in the mouth area (or maxillary region) instead of four[3], lip positions formerly unattainable with the Muscle Model Process method have been achieved (see Figure 5.1, Figure 5.2 and Figure 5.3).

5.3. Animation Update

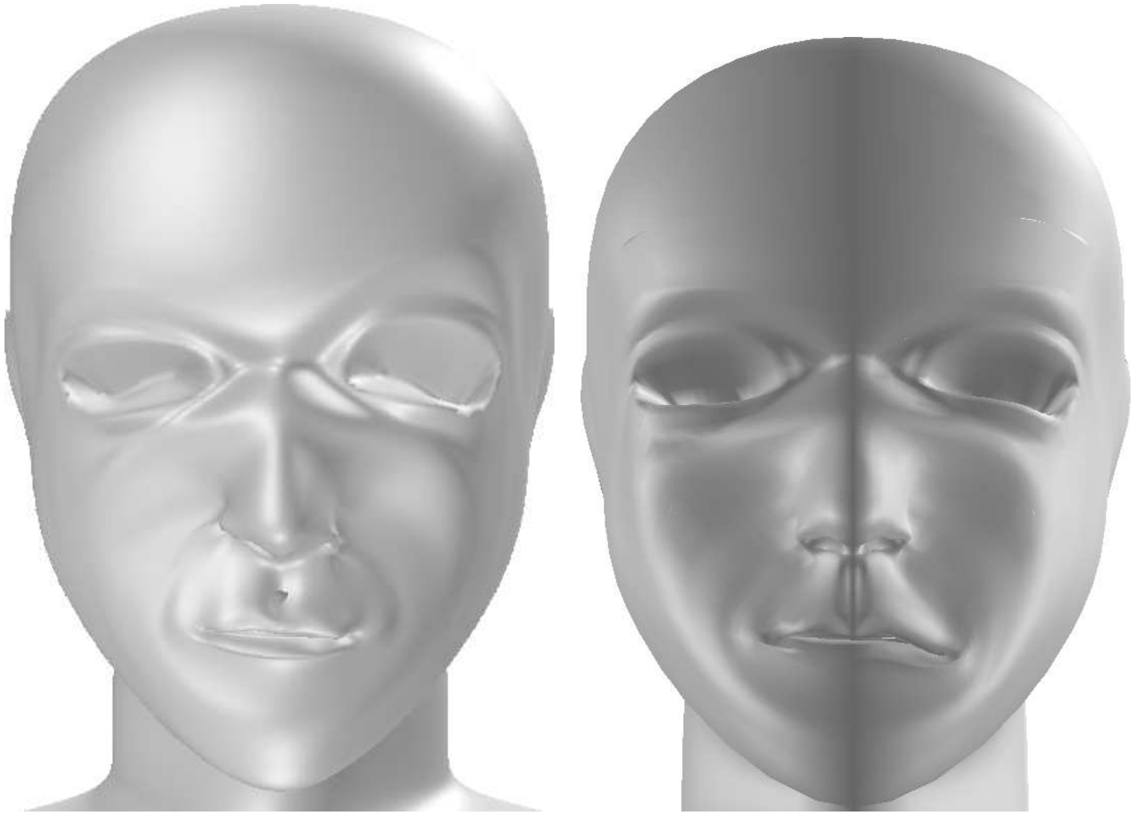
The animations generated with Langwidere so far give insight into the modeling process. The evolution of the face and vestigial body from a simple torus is useful in the study of forms and for those who intend to use Drag for similar purposes in the future.

Animated muscle and jaw movements partially demonstrate the range and flexibility of Langwidere as a facial animation system.

[1]Although Ekman and Friesen both claim to have learned this skill.

[2]With bilateral symmetry there would be forty muscles in this area, not including muscles relating solely to the jaw.

[3]Waters implemented one set of Zygomatics, the Depressor Anguli Oris, "the labii", and, as a sphincter muscle, the Orbicularis Oris.



(a) (b)

FIGURE 5.2. Al sneers and gives a lop-sided frown

5.4. Summary of Results

Hierarchical splines are well suited to modeling humans, but can use some improvement in the form of directional refinement of the surface. A muscle model variant with new methods of contraction simulation, and careful application of the muscles, produced a mouth better able to illustrate the visual aspects of speech. A few animated sequences were generated to show the results.

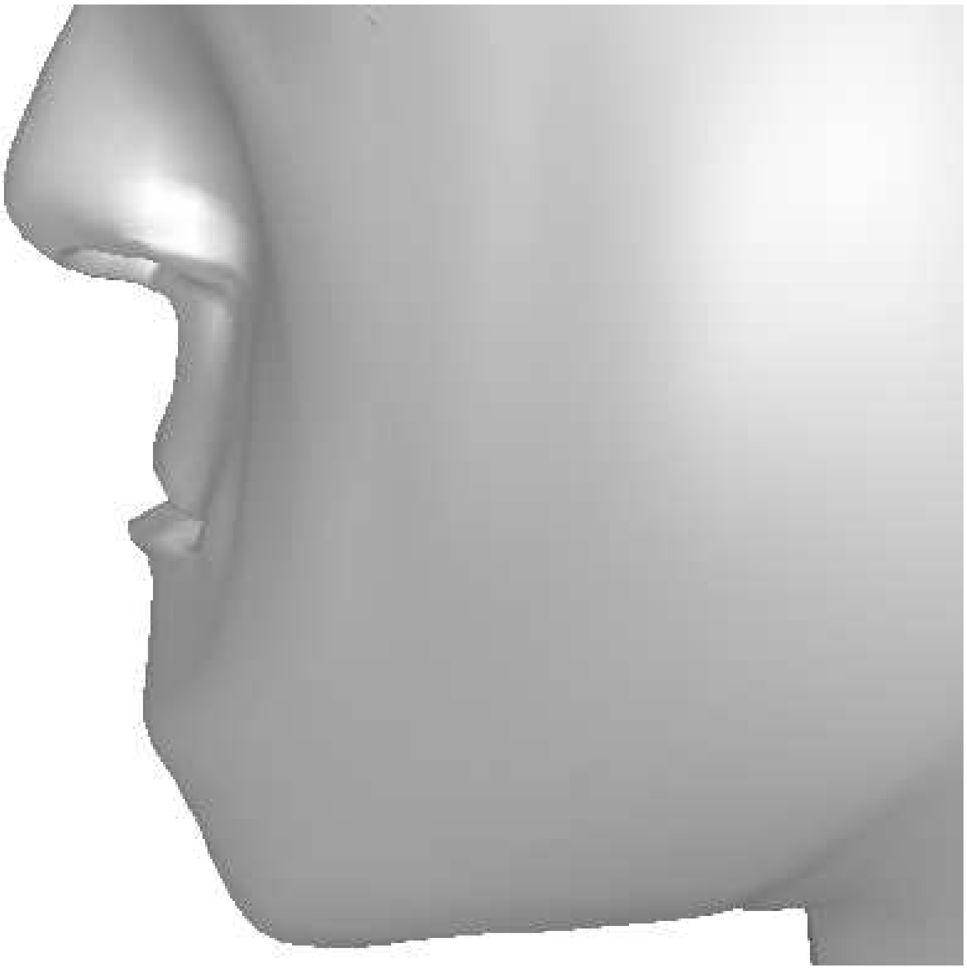


FIGURE 5.3. Al pouts

Chapter 6

Conclusions and Future Work

> "It's a very wonderful story," said Dorothy, "and it proves that the Land of Ev is really a fairy land, as I thought it was."
>
> L. Frank Baum
> page 48 of
> *Ozma of Oz*

This chapter states what conclusions may be drawn from this attempt to build the foundation of a human character animation system and what directions future research in this area might take.

6.1. The Modeling Gestalt

Hierarchical splines and simulated muscles combine to create a system of greater power and flexibility. A model constructed in this way can be refined and changed without affecting the underlying data structure. Muscles simulated by local surface deformation actuate the structural model regardless of the underlying geometry. Algorithmic techniques of animation and proper parameterization will mean that motion can be specified in a natural, high level manner. As was done with Betty and Chuck, the structural model was edited on all levels of the model, in all areas of the head, to produce different faces. Then without any recoding, resetting of the muscles, or activity other than structural editing, muscles that were initially installed in Al were used to produce the expressions shown in Figure 4.7. This means that a if Ev were a finished system, a single script that had been used to generate an animated sequence featuring Al could immediately be used to create an animation featuring

Chuck or Betty without changing a single line of code or modifying any control parameters but the once specifying whic head to use.

6.2. Teaching the Dumb to Speak

Mouth positions for speech are achievable, but a speech synthesizer needs to be modified to output mouth positions and appropriate synch and timing signals.

6.2.1. Combining Emotion and Speech. In combining emotion and speech, the greatest question arises when both want control of the mouth. How actions might be blended and in what proportion is an area for future analysis. Creating moods and giving quirks can probably be done by using personality profiles to record what actions the character is likely to generate periodically based on the current emotional state.

6.2.2. Dialog. Catherine Pelachaud is working on the problem of dialog for a post-doctorate at the University of Pennsylvania. Rules of interaction that govern the synchronization of gestures to speech are further complicated by the fact that people conversing with each other also synchronize to the speech and gestures of the other participant.

6.3. More Things To Do

To make Langwidere a complete facial animation system, there are still elements that need to be incorporated. Some of these elements are listed below.

The triangular muscle type the fan muscle, but converges towards the insertion point. An example would be the *Depressor Anguli Oris*, a triangular muscle which pulls the corner of the mouth down towards the jaw. Currently triangular muscles are simulated by sheet muscles.

Two muscle pull types pinned and floating are needed. There is only one instance of a pinned muscle in the expressive muscles of the face, the *Compressor Naris*. Both ends of this muscle are fixed to immovable bone. The muscle flattens and pulls towards both the attachment and insertion points. This is currently being simulated by two gradient fan muscles that meet at the bridge of the nose.

The various segments that comprise the *Orbicularis Oris* are not attached to bone or cartilage like most other muscles in the face, instead they are attached to skin and muscle via fasciæ. Since neither attachment nor insertion points are fixed, contraction occurs around the midpoint between the two; this would be a floating pull type. Floating muscles are currently being simulated by fan muscles.

Animating speech is a complex issue, but using a speech synthesizer to drive the speech animation parameters is an important beginning that duplicates the least amount of work.

Currently animation parameters for Langwidere are generated by hand or by short specific programs. A general purpose scripting language, most probably like Ev, is needed.

Existing Langwidere models have detailed faces, but only vestigal bodies and ears. Even if Langwidere's actuation system is not extended to the rest of the body, more detailed ears and a complete body should be sculpted for completeness.

The lack of a skull model has been a great detriment to general facial modeling due to the affect on structure and motion bone have on flesh. One way to make skull data as easy to generate as face data is to construct skull data from face data using forensic flesh thickness statistics and facial reconstruction techniques to interpolate it.

There are many more tools to make animation and model generation easier for the animator, but first things first ...

6.4. Summary of Conclusions and Future Work

Many and unending are the possibilities for extending and improving Langwidere. The purpose of the project to create a solid foundation for a realistic character animation system was fulfilled.

FIGURE 6.1. Al the Andorian

References

Abel (1988). The Making of BRILLIANCE. SIGGRAPH '85 Film & Video Show.

Badler, N., & Platt, S. (1981). Animating facial expression. *Computer Graphics*, *13*(3), 245–252.

Baum, L. F. (1980). *Ozma of Oz*. Coles Publishing Company Limited, Toronto. Illustrated by John R. Neill.

Belleville, Associate Editor, L. (1993). Sculpting with Computer Clay. *Computer Graphics World*, *16*(4), 21. Product spotlight on Thomson Digital Image (TDI) Blob Modeler.

Bergeron, P. (1989). Controlling Facial Expressions and Body Movements in the Computer-Generated Animated Short "Tony de Peltrie". In Parke, F. (Ed.), *ACM SIGGRAPH '89 Course Notes: State of the Art in Facial Animation*, pp. 202–220 Boston, Massachusetts. Siggraph.

Bezooyen, R. v. (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Netherlands Phonetic Archives. Foris Publications, Dordrecht.

Birdwhistell, R. (1970). *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Publications in Conduct and Communication. University of Pennsylvania Press, Philadelphia.

Bloomenthal, J., & Wyvill, B. (1990). Interactive Techniques for Implicit Modeling. In Riesenfeld, R., & Sequin, C. (Eds.), *Computer Graphics (1990 Symposium on Interactive 3D Graphics)*, pp. 109–116.

Breen, D. E., House, D. H., & Getto, P. H. (1991). A particle-based computational model of cloth draping behavior. *Scientific Visualization of Physical Phenomena (Proceedings of CG International '91)*, 113–134.

Bruce, V. (1988). *Recognising Faces.* Essays in Cognitive Psychology. Lawrence Erlbaum Associates Ltd.

Burden, R., & Faires, J. (1985). *Numerical Analysis* (3 edition). Prindel, Webber and Schmidt, Boston.

Chen, D. T., & Zeltzer, D. (1992). Pump It Up: Computer Animation of a Biomechanically Based Model of Muscle using the Finite Element Method. *Computer Graphics*, *26*(2), 89–98.

Chen, D. T.-W. (1991). *Pump It Up: Computer Animation of a Biomechanically Based Model of Muscle using the Finite Element Method.* Ph.D. thesis, Media Arts & Sciences, Massechusetts Institute of Technology.

Cheney, W., & Kincaid, D. (1985). *Numerical Mathematics and Computing* (Second edition). Brooks/Code Publishing Company, Monterey.

Cohen, E., Lyche, T., & Risenfeld, R. F. (1980). Discrete B-Splines and Subdivision Techniques in Computer Aided Geometric Design and Computer Graphics. *Computer Graphics and Image Processing*, *14*(2), 87–111.

Culhane, S. (1988). *Animation from Script to Screen.* St. Martin's Press, New York.

DiPaola, S. (1991). Extending the Range of Facial Types. *The Journal of Visualization and Computer Animation*, *2*(4), 129–131.

Druckman, D., Rozelle, R., & Baxter, J. (1982). *Nonverbal Communication: Survey, Theory, and Research.* Sage Publications, Beverly Hills.

Duncan, J. (1991). A Once and Future War. *Cinefex*, *47*.

Ekman, P., & Friesen, W. (1984). *Unmasking the Face.* Consulting Psychologists Press, Palo Alto, CA.

Ekman, P., & Friesen, W. V. (1978a). *Facial Action Coding System.* Consulting Psychologists Press, Palo Alto, CA.

Ekman, P., & Friesen, W. V. (1978b). *Facial Action Coding System Investigator's Guide.* Consulting Psychologists Press, Palo Alto, CA.

Elson, M. (1990). "Displacement" Facial Animation Techniques. In *State of the Art in Facial Animation: SIGGRAPH 1990 Course Notes #26*, pp. 21–42. 17th International Conference on Computer Graphics and Interactive Techniques, Dallas Convention Center.

Flannery, B. P., Press, W. H., Teukolosky, S. A., & Vetterling, William, T. (1992). *Numerical Recipes in C* (2 edition). Press Syndicate of the University of Cambridge, New York.

Flexner, S. B. E. (Ed.). (1988). *The Random House College Dictionary Revised Edition.* Random House.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer Graphics: Principles and Practice* (2 edition). Addison-Wesley Publishing Company, Don Mills, Ontario.

Forsey, D., & Wang, L. (1993). Multi-resolution Surface Approximation for Animation. *Graphics Interface*, 192–199.

Forsey, D. R. (1991). Hierarchical Free-Form Surface Modeling. In Bartels, R. (Ed.), *Course Notes 25: Topics in the Construction, Manipulation, and Assessment of Spline Surfaces*, pp. 5–0 : 5–56 Las Vegas Convention Center. ACM SIGGRAPH.

Forsey, D. R. (1990). *Motion Control and Surface Modeling of Articulated Figures in Computer Animation.* Ph.D. thesis, University of Waterloo, Computer Science Department.

Forsey, D., & Bartels, R. (1988a). Hierarchical B-spline refinement. *Computer Graphics*, 205–212.

Forsey, D., & Bartels, R. (1988b). Local refinement editing of B-spline surfaces. In *Proceedings Graphics Interface '88.* Graphics Interface.

Franks, S. (1991–1993) Various personal communications.

Fromkin, V. (1964). Lip positions in American English vowels. *Languagae and Speech*, *7*, 215–225.

Glassner, A. S. (1991). *Maintaining Winged-Edge Models*, Vol. II of *Graphics Gems*, chap. IV.6, pp. 191–201. Academic Press, Inc., Ithica, New York.

Goldfinger, E. (1991). *Human Anatomy for Artists: The Elements of Form.* Oxford University Press, New York.

Gordon, L. (1980). *How to Draw the Human Figure: An Anatomical Approach.* Penguin Books, Ltd., Markham.

Gordon, L. (1983). *How to Draw the Human Head: Techniques and Anatomy.* Penguin Books, Ltd., Markham.

Gray, F.R.S., H. (1977). *Gray's Anatomy: The Classic Collector's Edition.* Bounty Books, New York.

Gross, D. (1991). Man and Machine. *Computer Graphics World*, 47–50.

Haman, A. C. (1992). Deformation Based Modelling. Master's thesis, University of Calgary, Department of Computer Science.

Hill, D. R., Pearce, A., & Wyvill, B. (1988). Automating speech: an automated approach using speech synthesised by rules. *The Visual Computer*, *3*, 277–289.

Iversen, W. R. (1991). Animation Takes the Stand. *Computer Graphics World*, *14*(11), 48–54.

Kajiya, J. T., & Kay, T. L. (1989). Rendering Fur with Threee Dimensional Textures. *Computer Graphics*, *23*(3), 271–280.

Kendon, A. (Ed.). (1981). *Nonverbal Communication, Interaction, and Gesture: Selections from Semiotica.* Mouton Publishers, New York.

Komatsu, K. (1988). Human skin model capable of natural shape variation. *The Visual Computer*, *3*(5), 265–271.

Kraus, B. S., Kitamura, H., & Latham, R. A. (1966). *Atlas of Developmental Anatomy of the Face: With special reference to normal and cleft lip and palate.* Harper & Row, Hoeber Medical Division, New York.

Krogman, W. M. (1962). *The Human Skeleton in Forensic Medicine.* Charles C. Thomas, Springfield, Illinois.

Lewis, J. (1991). Automated Lip-sync: Background and Techniques. *The Journal of Visualization and Computer Animation*, *2*(4), 118–122.

Lewis, J. (1989). Algorithms for Solid Noise Synthesis. *Computer Graphics*, *23*(3), 263–270.

Lewis, J., & Parke, F. (1987). Automated lipsynch and speech synthesis for character animation. In *Proceedings Human Factors in Computing Systems and Graphics Interface '87*, pp. 143–147.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Morris, D. (1982). *The Pocket Guide to Manwatching.* Triad Grafton, Glasgow.

Nahas, M., Huitric, H., Rious, M., & Domey, J. (1990). Facial image synthesis using skin texture recording. *The Visual Computer*, *6*(6), 337–343.

Nahas, M., Huitric, H., & Saintourens, M. (1988). Animation of a B-spline figure. *The Visual Computer*, *3*(5), 272–276.

O'Grady, W., & Dobrovolsky, M. (1987). *Contemporary Linguistic Analysis.* Copp Clark Pitman Ltd., Toronto.

O'Neill, J., & Oyer, H. J. (1981). *Visual Communication for the Hard of Hearing* (Second edition). Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 07632.

Orth, T. (1993). Graphical Modeling Using L-Systems and Particle Systems: A Comparison. Master's thesis, University of Calgary, Department of Computer Science.

Parke, F. (1974). *A Parametric Model for Human Faces.* Ph.D. thesis, University of Utah, Salt Lake City, Utah.

Parke, F. (1982). A parameterized model for facial animation. *IEEE Computer Graphics and Applications*, *Second*(9), 61–70.

Parke, F. I. (Ed.). (1989). *Course Notes 22: State of the Art in Facial Animation.* ACM SIGGRAPH.

Pelachaud, C. (1990). *Facial and vocal expressions of emotion for an animation system (Draft).* Ph.D. thesis, University of Pensylvania, Philadelphia, Pennsylvania.

Pelachaud, C. (1991). *Communication and Coarticulation In Facial Animation.* Ph.D. thesis, Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 19104–6389.

Pennycuick, C. J. (1992). *Newton Rules Biology: A physical approach to biological problems.* Oxford University Press, Oxford.

Perlin, K., & Hoffert, E. M. (1989). Hypertexture. *Computer Graphics*, *23*(3), 253–262.

Pieper, S. D. (1991). *CAPS: Computer-Aided Plastic Surgery.* Ph.D. thesis, Media Arts and Sciences, Massachusetts Institute of Technology.

Reeves, W. T. (1983). Particle Systems — A Technique for Modeling a Class of Fuzzy Objects. *Computer Graphics*, *17*(3), 359–376.

Reeves, W. T. (1990). Simple and Complex Facial Animation: Case Studies. In *State of the Art in Facial Animation: SIGGRAPH 1990 Course Notes #26*, pp. 88–106. 17th International Conference on Computer Graphics and Interactive Techniques, Dallas Convention Center.

Ritchard with Kathleen Moloney, D. (1987). *Ventriloquism for the Total Dummy.* Villard Books, New York. Illustrations by Sam Berland.

Robertson, B. (1991). The Electronic Theatre. *Computer Graphics World*, *14*(7), 46–54. Caption: *Virtually Yours*, Matt Elson, Symbolics, Los Angeles.

Rogers, D. F., & Adams, J. A. (1990). *Mathematical Elements for Computer Graphics* (Second edition). McGraw-Hill.

Schock, C.-R. (1993). Synthesizing Intonation for Computer Speech Output. Master's thesis, University of Calgary, Department of Computer Science.

Sims, K. (1988). Particle Dreams. SIGGRAPH '88 Film & Video Show - *part 2.* Optomystic.

Smith, G. (1990). Modeling the motion of human muscle systems with computer graphics. Master's thesis, University of Utah, Salt Lake City, Utah.

Tatoul, C. M., & Davidson, G. D. (1961). Lipreading and Letter Prediction. *Journal of Speech and Hearing Research*, *4*(2), 178–181.

Terzopoulos, D., & Waters, K. (1990). Physically-based facial modelling, analysis, and animation. *The Journal of Visualization and Computer Animation*, *1*, 73–80.

Thomas, F., & Johnston, O. (1981). *Disney Animation: The Illusion of Life*. Abbeville Press, New York.

van Overveld, C., & Wyvill, B. (1993). Potentials, Polygons and Penguins: An efficient adaptive algorithm for triangulating an equi-potential surface. In *Proceedings of the Fifth Annual Western Computer Graphics Symposium*, pp. 1–34 Silver Star Mountain, Vernon, BC.

Videosystem (1991). The Making of "Don Quichotte". SIGGRAPH '91 Film & Video Show - *part 1.* Contact Alain Guiot.

Waite, C. T. (1989). The Facial Action Control Editor, FACE: A Parametric Facial Expression Editor for Computer Generated Animation. Master's thesis, Massachusetts Institute of Technology. spline model with muscles.

Wang, C. L.-Y. (1992). Current Trends in Facial Animation, or Langwidere: Not Just Another Witch. In *Proceedings of the Fourth Annual Western Computer Graphics Symposium*, pp. 103–108 Sunshine Village, Banff, AB.

Waters, K. (1987). A muscle model for animating three-dimensional facial expression. *Computer Graphics*, *21*(4).

Waters, K. (1988). *The Computer Synthesis of Expressive Three-Dimensional Facial Character Animation*. Ph.D. thesis, Middlesex Polytechnic, Queensway, Enfield, Middlesex.

Waters, K. (1990). Part 1: Modeling 3D Facial Expressions. In Parke, F. I. (Ed.), *Course Nodes 26: State of the Art in Facial Animation*, pp. 109–164 Dallas Convention Center. ACM SIGGRAPH.

Waters, K. (1992) Personal Communication.

Waters, K., & Terzopoulos, D. (1991). Modeling and Animating Faces using Scanned Data. *The Journal of Visualization and Computer Animation*, *2*(4), 123–128.

Williams, L. (1990). Electronic Mask Technology. In Parke, F. I. (Ed.), *Course Nodes 26: State of the Art in Facial Animation*, pp. 164–184 Dallas Convention Center. ACM SIGGRAPH.

Woodward, M. F., & Barber, C. G. (1960). Phoneme Perception in Lipreading. *Journal of Speech and Hearing Research*, *3*(3), 212–222.

Wyvill, B., & Hill, D. R. (1989). Expression control using synthetic speech. In Parke, F. (Ed.), *State of the Art in Facial Animation*, pp. 163–175. ACM SIGGRAPH.

Yoshimoto, S. (1992). Ballerinas Generated by a Personal Computer. *The Journal of Visualization and Computer Animation*, *3*(2), 85–90.

Glossary

To aid the reader distinguish terms, general terminology is denoted by the sans serif font, acronyms are denoted by ALL CAPS, and system names are denoted by *italics*.

Actuation Model the basic method by which the structural model is modified to simulate motion.

Animation Model groups and modifies the actuation parameters for animation.

AU Action Unit from FACS.

C a popular programming language.

Drag the hierarchical spline system created by David Forsey.

Ev the animation language.

fasciæ fibrous, inelastic, collagenous connective tissue.

face Clea Waite's Facial Action Control Editor.

Facial Animation System see FAS.

FACS the Facial Action Coding System.

FAS Steve DiPaola's Facial Animation System.

Fascia Frederick Parke's public domain face model.

Langwidere A hierarchical spline based facial animation system with simulated muscles.

Muscle Model a simulation of the physical processes that affect the skin when a muscle is contracted.

north pole wrinkled seam problem with bicubic splines terminating at a rounded point.

Speech Model generator of speech synchronization data for animation.

Structural Model collection of data representing the physical shape of the head.

Surface Model the method by which the surface of the object is represented.

waldo a device that, in real-time, remotely operates a similarly configured device that exactly performs the actions of the manipulator. Probably the most well known example would be waldo hands used to handle radioactive or biologically virulent material.

Appendix A
Facial Action Coding System Action Units

There are eight missing AUs in the sequence from 1 to 66: 3, 40, 47, 48, 49, 50, 59, 60. Why these particular numbers are missing is unknown. A statement made in the *Investigator's Guide Part I* may help: "The numbers are arbitrary and do not have any significance except that 1 through 7 refer to brows, forehead or eyelids." (Ekman & Friesen, 1978b, pp. 7–8)

Table A.1. Facial Action Coding System Action Unit definitions.

Au #	Facs Name	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae, Depressor Supercilli, Corrugator
5	Upper Lid Raiser	Levator Palpebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palpebralis
8	Lips Toward Each Other	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

continued on next page

continued from previous page

AU #	FACS Name	Muscular Basis
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Caninus
14	Dimpler	Buccinnator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris, Incisivii Labii Inferioris
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	Lips Part	Depressor Labii or Relaxation of Mentalis or Orbicularis Oris
26	Jaw Drop	Masetter, Temporal and Internal Pterygoid Relaxed
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	

continued on next page

continued from previous page

AU #	FACS Name	Muscular Basis
35	Cheek Suck	
36	Tongue Bulge	
37	Lip Wipe	
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
41	Lid Droop	Relaxation of Levator Palpebrae Superioris
42	Slit	Orbicularis Oris
43	Eyes Closed	Relaxation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relaxation of Levator Palpebrae and Contraction of Orbicularis Oculi, 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	

Appendix B
Anatomical Implementation Detail

Many anatomy, general medical, and fine arts texts were used to locate muscles for placement and describe the functions of muscles, but the most invaluable of these with respect the construction of Langwidere was *Human Anatomy for Artists: The Elements of Form* by Eliot Goldfinger which has detailed anatomical information and profuse illustrations combined with photographs to demonstrate both form and function for even the smallest muscle.

Since anatomy texts tend use as much medical jargon as possible in their descriptions, this chapter contains a layman's guide to muscle location, placement, and function for the facial muscles of expression, not including those associated with the jaw. These are the muscles that have been implemented for Langwidere. Also included for each muscle are the short codes by which the muscle is referred to in Langwidere. Where more than one name for a muscle was encountered in reference texts, the additional names are included in parentheses.

- Medial Frontalis
  - middle forehead muscle
  - raise inner portion of eyebrow
  - lmf
  - rmf
- Lateral Frontalis

  - outer forehead muscle
  - raise middle portion of eyebrow
  - llf
  - rlf
- Depressor Supercilli
  - inner corner of eye to eyebrow
  - pulls inner part of eyebrow down
  - lds
  - rds
- Corrugator Supercilli
  - between the inner corner of the eye socket and the nose to the middle of top margin of the eye socket (deep)
  - pulls the eyebrows together
  - lcs
  - rcs
- Procerus (Pyramidalis Nasi)
  - the bone-cartilage interface of the nose to mid-forehead
  - pulls down the middle of the forehead and inner corners of the eyebrows
  - pro
- Compressor Naris (Transverse Nasalis)
  - above the canines but below the nostrils to the middle of the cartilaginous part of the nose
  - flattens the sides of the nose
  - lcn
  - rcn
- Depressor Alae Septi (Alar Nasalis)
  - above the lateral incisor but below the nostrils to the nostrils and septum

(deep)

- pulls the tip of the nose down and pushes the upper lip down
- ldas
- rdas

- Dilator Naris (Alaris)
  - lower edge of the bony nasal cavity to the lower edge of the side of the nostrils
  - flares the nostrils
  - ldn
  - rdn
- Orbicularis Oculi (Orbicularis Palpebrarum)
  - around the eye, above the eyebrow, down to the cheekbone, around the side of the head
  - closes the eye, can raise the cheek independently
  - loo
  - roo
- Levator Palpebrae Superioris
  - from the back of the eyesocket to the front of the top eyelid
  - pulls back the top eyelid
  - llps
  - rlps
- Incisivus Labii Superioris
  - from midway between the side of the nostrils and the vermilion border of the top lip to the top of the muscle node at the side of the lips
  - pulls up the corner of the mouth, works with the Levator Palpebrae Inferioris to round the lips
  - lils
  - rils

- Incisivus Labii Inferioris
  - from midway between the corner of the chin and the vermillion border of the bottom lip to the bottom of the muscle node at the side of the mouth
  - pulls down the corner of the mouth, works with the Levator Palpebrae Superioris to round the lips
  - lili
  - rili
- Levator Labii Superioris Alaeque Nasi (Quadratus Labii Superioris — Angular Head)
  - alar portion (from the bridge of the nose to just behind the wing of the nose), furrow portion (from the bridge of the nose to the top of the furrow next to the nose), lip portion (from the bridge of the nose to the top of the vermilion border, just outside the philtrum)
  - the sneer muscle
  - lllsan{afl}
  - rllsan{afl}
- Levator Labii Superioris (Quadratus Labii Superioris — Infraorbital Head)
  - just below the middle of the bottom of the eyesocket to the top of the furrow next to the nose (superficial), down to the border of the moist and dry parts of the lip (middle), "inserts at numerous points into the mucous membrane on the inside of the mouth" (deep)
  - sneer muscle, also pushes lips out slightly and pulls lips open at point of insertion
  - llls{smd}
  - rlls{smd}
- Zygomaticus Minor (Quadratus Labii Superioris — Zygomatic Head)

  - the front corner of the cheekbone to the middle of the nasolabial furrow, continuting on to the edge of the vermilion border just above the canines
  - deepens the middle section of the nasolabial furrow and pulls the top of the lip diagonally outwards (crying expression)
  - lzm
  - rzm
- Orbicularis Oris
  - the orbicularis oris consists of eight triangular sections of muscle (left or right, superior or inferior, and marginal or peripheral) each of which attach to the muscle node at the sides of the mouth and insert into the middle of the lips, right and left sides slightly overlaping. The marginal area is the area contained by the vermilion border of the lips. The peripheral area is outside the marginal area, the superior area stops at the nose, and the inferior area stops at the crease under the lower lip (between where the lip area ends and the bulge of the chin begins).
  - compresses the lips and press them against the teeth (peripheral portion), inverts the red-lip area pulling them inwards (marginal portion).
  - l{si}oo{mp}
  - r{si}oo{mp}
- Levator Anguli Oris (Caninus)
  - somewhat out from the top of the wing of the nostrils to top of the muscle node at the side of the mouth
  - pulls the sides of the lips straight up
  - llao
  - rlao
- Zygomatic Major
  - from the front corner of the cheekbone to the top fo the muscle node at the

side of the mouth
  - pulls the corner of the lips back and up (smile)
  - lzms
  - rzms
- Risorus
  - from the fascia of the cheek just over the back of the teeth to the muscle node
  - pulls the corner of the lips back
  - lr
  - rr
- Buccinator
  - from the border of the back of the jaws with the teeth and a band of tendon that bridges this area to the muscle node
  - compresses the cheeks, pulls the corner of the mouth straight back
  - lb
  - rb
- Depressor Anguli Oris (Triangularis; Triangularis Menti)
  - from the corner of the chin to the muscle node
  - pulls the corner of the mouth down
  - ldao
  - rdao
- Depressor Labii Inferioris (Quadratus Labii Inferioris; Quadratus Menti)
  - from the corner of the chin to the middle of the indentation below the lower lip
  - pull down the lower lip and push the chin down
  - ldli
  - rdli
- Mentalis (Levator Menti; Levator Labii Inferioris)

  - from a line one side of the middle of the chin just below the bottom of the teeth to the bottom of the chin (point of attachment is deep) to the skin (point of insertion is superficial) above the origin of the Mentalis on the other side of the face
  - pushes the middle of the mouth up, flattens out the chin
  - lm
  - rm
- Platysma
  - from just below the clavicle to (nodular portion) the muscle node at the corner of the mouth, (labial portion) vermilion border of the bottom lip above the canines, (mandibular portion) lower edge of the corner of the chin
  - pulls the lower lip out and down (nodular and labial) pulls the lower jaw down (mandibular portion)
  - lp{nm}
  - rp{nm}