The computer synthesis of expressive three-dimensional facial character animation.

Waters , K.

Abstract

This present research is concerned with the design, development and implementation of three-dimensional computer-generated facial images capable of expression gesture and speech.

A review of previous work in chapter one shows that to date the model of computer-generated faces has been one in which construction and animation were not separated and which therefore possessed only a limited expressive range. It is argued in chapter two that the physical description of the face cannot be seen as originating from a single generic mould. Chapter three therefore describes data acquisition techniques employed in the computer generation of free-form surfaces which are applicable to three-dimensional faces.

Expressions are the result of the distortion of the surface of the skin by the complex interactions of bone, muscle and skin. Chapter four demonstrates with static images and short animation sequences in video that a muscle model process algorithm can simulate the primary characteristics of the facial muscles.

Three-dimensional speech synchronization was the most complex problem to achieve effectively. Chapter five describes two successful approaches: the direct mapping of mouth shapes in two dimensions to the model in three dimensions, and geometric distortions of the mouth created by the contraction of specified muscle combinations.

Chapter six describes the implementation of software for this research and argues the case for a parametric approach. Chapter seven is concerned with the control of facial articulations and discusses a more biological approach to these. Finally chapter eight draws conclusions from the present research and suggests further extensions.

CONTENTS

Chapter 3: Modelling the Face as a Three-Dimensional Topology

Chapter 4: The Muscle Model Process

Chapter 5: Speech Synchronization

Chapter 6: The Implementation of Facial Animation

Chapter 7: The Control of Facial Expressions

Chapter 8: Conclusions and Extensions to Current Work

Appendices

List of Tables

Video Sequences

Illustrations Page

Symbols and Notation

Pi	an irrational number approximately equal to 3.1415927
$a < b$	a is less than b
$a \leq b$	a is less than or equal to b
$a > b$	a is greater than b
$a \geq b$	a is greater than or equal to b
/	division
*	multiplication
$A \cap B$	the intersection of the sets A and B
$x \propto f(A,B)$	x is the result of a function dependant on A and B
$\overline{A}$	vector
.	dot product
$\sum_{i=0}^{i=n} Ni$	is the sumation of $N_0 + N_1$ ...... Nn

Introduction

Charles Darwin, in his book The Expression of the Emotions in Man and Animals, strengthens his theories of evolution by implying that emotional expressions are, by and large, partly determined by our evolution. To explain and defend his theory, Darwin described common emotional expressions in infants and children, adults from various cultures, the mentally ill, and animals. The durability of his theories over the past hundred years illustrates the profound importance of the face as a powerful form of non-verbal communication [1].

The face has continuously demanded special attention. In practically every age and culture a concentration on the facial form can be observed. John Liggett in his comprehensive book on the face points out that all people, both sophisticated and primitive, seemed to be prepared to go through almost unbelievable suffering in pursuit of the purely local ideals of their particular society concerning the face [2]. Not only is the face painted, tattooed and marked, but it is also physically distorted and even deliberately disfigured. The reason for such activity is complex, but it re-enforces the fact that the face has always been an enduring form of non-verbal communication.

Psychologists, naturalists or scientists do not possess a monopoly of the knowledge concerning emotional expression. Painters, sculptors and photographers are profoundly knowledgeable about the emotive potency encapsulated in facial expression. They have incessantly captured static facial expressions which portray moods or narrative in their images. Daumier the remarkable French political satirist conveyed the highly charged spectacle of activity in court rooms of the nineteenth century. Within an individual illustration he competently characterized the rage of the prosecution, the horror of the defendant and the surprise of the jury [3].

Portrait artists, and, later, photographers have through the centuries studied the face in intricate detail, both in terms of the physical structure and the extremes of facial communication: from the enigmatic smile of the Mona Lisa to the haunting faces of Edvard Munch. Static portraiture can capture the diversity of human emotions, while photography makes it possible to record accurately snap-shots of expressions frozen in time. It was the advent of film before sound that re-enforced the face as a highly emotive device, and not long after the face was being used in animation as an extension to the film maker´s art. Animation studios, such as the Disney´s studios, created many endearing characters like the seven dwarfs capable of the full range of facial expressions [4]. Computer graphics now offers an exciting medium in which artists and designers can explore fresh approaches to the animation of the human face. What is evident is that the face will unquestionably remain the most challenging, mystical and communicative feature of the human form that we will incessantly desire to duplicate, manipulate and record.

Animation is the process of bringing the static image to life, and is achieved by displaying successive images rapidly enough to exploit the phenomenon of persistance of vision. Using the computer to make animation is not innovative. Algorithms can effortlessly and accurately translate an imaginary three-dimensional world into a two-dimensional image, then by successively modifying the computer´s view point, the illusion of a three-dimensional world can be established [5].

Computer animation eliminates much of the physical drudgery of ´inbetweening´ present in traditional animation techniques, but the luxury of removing this tedium does not replace the fundamental skills of the animator. Frank Thomas, the noted Disney animator, describes how competent cell animation can be attained [6]. He justifies at length his view that successful animation is specifically attributable to the long hours spent studying living animals, and is directly proportional to how accurately the animator has understood the kinematics, timing and structure of the subject. Consequently, animation is not the mimicry of live action film, nor is it the facile exaggeration, squashing and stretching of lines to present the appearance of movement. Only with a comprehensive understanding of the subject matter can the art of animation be practiced.

The medium of three-dimensional computer animation brings a new perception of space and depth beyond what could possibly be achieved by hand, and the ease with which the observer can manoeuver around objects is its hallmark. Yet these imaginary micro-worlds are saturated with sterile monolithic objects that are a far cry from the early animators´ aspirations to bring real objects and characters to life. In addition, the attempt to transpose traditional cell techniques into the third dimension does not substantially capitalize on the true nature of three-dimensional computer animation. Unlike hand-drawn animation, three-dimensional computer animation presents the animator with exact analytical data that can be used either to perform predetermined manipulation, or to serve as a form of feedback which can regulate actions. The specific nature of this control is an active area of research and only a small number of environments have been conceived with the purpose of adaptive and automatic control [7].

The ease with which it is possible to create and manipulate geometric structures within three-dimensional computer animation does not apply to sophisticated natural free-forms such as the face. While computer generated human characters have been attempted, the significance of the face as an expressive entity has, on the whole, been disregarded because of the intrinsic complexities of the structure and mobility of the face. Consequently, computer animation persists in offering inert micro-worlds comprise of dummies, and deficient in characters endowed with expression and speech. This present research endeavors to break this stereotyped mould.

The object of the work this present research is to describe a framework for the synthesis of expressive three-dimensional facial animation. The operation involves two individual domains: the modelling of facial topologies that can be represented as polygonal networks, and the incorporation of the motion characteristics of expression.

References

[1] Paul Ekman, Darwin and Facial Expression a Century of Research in Review, Academic Press, London, 1973.

[2] John Liggett, The Human Face, Constable and Company, London, 1974.

[3] K E Maison, Daumier, Sagamore Press, New York, 1960.

[4] Frank Thomas and Ollie Johnston, Disney Animation: The Illusion of Life, Abbeville Press, New York, 1984.

[5] William M Newman and Robert F Sproull, Principles of Interactive Computer Graphics, McGraw-Hill, London, 1979, pp. 333-354.

[6] Thomas, Disney Animation:, pp. 169-189.

[7] David Zeltzer, ´Representation and Control of Three-Dimensional Computer-Animated Figures´, Ph.D Thesis, Ohio State University, 1984.
Michael Girard, ´Interactive Design of 3D Computer-Animated Legged Animal Motion´, IEEE Computer Graphics and Applications, June 1987, pp. 39-51.

CHAPTER ONE

Chapter One

A Review of Previous Work

1.1 Introduction.

Three-dimensional synthetic facial animation relies upon general techniques of computer graphics such as geometric transforms and rendering algorithms. In recent years there have been substantial innovations in many areas of computer graphics. This review cannot therefore be entirely comprehensive and is limited to those areas directly influencing facial animation.

Cell animation is traditionally two-dimensional, and rudimentary procedures have been evolved to produce believable characters. This review scrutinizes methods such as squeeze, squash and stretch, which can endow the character with a considerable repertoire of facial expressions by the distortion and contortion of the face. In addition, many fundamental techniques such as inbetweening in traditional animation are examined in relation to computer animation.

Undoubtedly the most significant investigation into three-dimensional computer-generated facial animation was carried out some ten years ago by F Parke for his PhD at the University of Utah, where a great many crucial contributions to computer graphics were made in the 1960´s [1]. This review critically appraises the approach and the techniques involved.

Perhaps the most notable example of a computer animation short film involving a character was Tonie De Peltrie, produced by a group of students from the University of Montreal over the period of two years from 1983. Funded by a grant from the National Film Board of Canada and using the facilities at the Centre de Calcul de l´Universite de Montreal, the work involved the development of the software packages called DADS and TRAANA which facilitate three-dimensional character development for designers and artists with no computer background [2]. The techniques are reviewed and examined.

Only one documented attempt has been made to create an approximation of muscle activity and the resultant skin displacements. Platt used a tension net approach to the modelling of the skin motion and this approach is also examined [3].

1.2 Traditional Cell Animation and Computer Animation.

While the computer synthesis of the face relies on strategies of computer-generated imagery, it is also influenced by traditional animation techniques. Animation is the process of bringing static images to life, and is achieved by successively displaying images fast enough to exploit the phenomenon of persistence of vision [4]. Typically in film twenty-four frames a second are recorded, while in video twenty-five frames a second are employed, although it is possible to use as little as twelve or even eight frames a second to create the illusion of motion. Traditional cell animation is a two-dimensional process such that single frames are painted sequentially by the artist. Often many preliminary pencil tests are recorded before the final motion is painted on cell, as it cannot be modified once it has been committed to cell.

Animation is one of the most challenging and demanding disciplines in art. Frank Thomas at the Disney Studios points out that there were never twenty animators who fully mastered the craft [5]. Computer animation will not provide an alternative to the skills of an individual animator. It may remove the drudgery, such as repetitive inbetween drawing, from animation, but it will still be up to the artist to search for ways to use the capabilities of the computer to the best advantage. Manipulation techniques and faster repose times may help animators, but they do not of themselves make creative animation any easier. There is a general misconception that anyone can be an instant animator provided he or she has the right computer tools, but this is far from being the case. Bad animators will always produce poor animation, and computer technology, however wonderful, will never make great animators out of averagely gifted animators.

Computer animation has been criticized for its appearance of complexity, its inflexibility and lack of expression, and attempts at human action in three-dimensions have, in the most part, resulted in wooden, lifeless characters. This is not the problem of naive computer programming (indeed the sophistication of some of the algorithms is astonishing) but of poor animation and direction. Consequently, the use of computer does not make the art of animation any easier, rather it adds further to the already complex array of animation tools.

1.2.1 Computer Interpolation as Inbetweening.

The early work in computer animation developed two-dimensional techniques based upon traditional animation [6]. Two of the key methods that are fundamental tools of every animation system are interpolation/inbetweening and keyframe animation.

Traditionally, the process of inbetweening was carried out by the assistant animator once the main poses (the extremes) of the sequence had been created. Drawings would then be made for the inbetween frames as specified by the chief animator. This was not as straight-forward as it may at first appear, as very few living organisms are capable of dynamics that have mechanical straight-line motion. This can clearly be seen in the process of a walking figure [7], where the activity is full of arcs with no straight line-motions. Therefore, presenting animation that has equal steps can completely kill the essence of the action.

For computer animation the process of inbetweening can be described as interpolation and can be explained in simple terms as the linear calculation of a displacement activity between two extremes. Typically, considering a line as two points $(x1, y1)$, $(x2, y2)$, a general point $(x3, y3)$ on that line is given as:

$$((1 - t)\, x1 + t * x2,\ (1 - t)\, y1 + t * y2)$$

If t is a parameter value in the range $0 <= 1$ then the new point x3, y3 lies on the line between x1, y1 and x2, y2 [8]. Consequently, an interpolant can be specified as a percentage displacement from point to point. Thus if t = 0.5 then this is the midpoint between the extremes. Considering a two-dimensional drawn shape on cell as a group of x, y points, it is possible to apply the above principle to compute the transitions from one shape to another. This depends upon the two shapes having the same number of points and these being specified in the same order. Generally, in two-dimensions, drawings are not created with the same number points, so an alternative solution is required.

One approach is to calculate additional points for the image with fewer indices until the two are equal, and then to interpolate the shapes [9]. However, as in traditional animation, straight-line motion can completely kill the essence of the move, especially if it concerns the animation of characters. Some of the reasons for the lack of realism concern motion inertia. As objects move, they are under the influence of gravity and other reacting forces [10]. Newtonian mechanics, such as F = ma, where m is the mass, a is the acceleration and F is the resultant force, indicate how generalized bodies move, but with an articulated body the calculations become mathematically complex and computationally expensive. Traditional methods of animation applied techniques to overcome this problem which were learned by trial and error. Results were obtained by placing inbetweens close to each extreme. This process was named slow in and slow out, since this was how the frames were timed as they were in essence acceleration and deceleration [11].

The interpolation process of ease-in and ease-out can be translated into a computer algorithm (Appendix A.1). This allows the selection of an ease-in or ease-out activity as described by the traditional animators, but this is only a first approximation, and there are a number of associated problems. Firstly, the process will always interpolate the motion at a fixed rate, which is uncharacteristic of many activities. Secondly, sequences with more than two keyframes result in motional discontinuity, as no consideration is given to other key positions during interpolation. The first problem can be simply rectified by modifying the existing interpolation from a cosine activity to a non-linear coefficient between 0 and 1. The second problem of discontinuity requires more sophisticated computing, using some of the wide variety of techniques which allow the animation to achieve smooth transitions across key-frames. Fundamentally, it can be seen as a process of spline fitting to existing points that can have additional properties such as tension, continuity and bias control [12].

The two attributes of keyframing and interpolation mentioned above are techniques fundamental to both traditional and computer animation, but they are achieved by dramatically different means. Extensions of these properties to three-dimensional computer animation have been commonplace, and have been developed as the most elementary mode of control. The principles of traditional two-dimensional cell animation differ radically from those of three-dimensional computer animation, because the first employs two-dimensional drawings while the other three-dimensional objects. Nonetheless, many of the integral principles of animation have been utilized with computer animation [13].

It is still too early in the development of computer animation to define distinct principles for animation: only the artists/designers who use the medium will discover these. The Disney animators were concerned with entertainment and this focused their attention on emotions and theatrical studies. They were also acutely aware of movement and timing, and realized that they had to understand many of the physical properties of the various motions before they could mimic and characterize them. The Disney animators established some fundamental principles and focused their attention on the key issues of bringing characters to life. Squash and stretch, timing, anticipation, staging, follow through, pose-to-pose, slow in and out, arcs, exaggeration, secondary action, appeal and personality are described by Thomas and Johnston as the fundamental principles to animation [14]. These principles have been suggested by John Lasseter, an ex-Disney animator, to be valuable tools for the computer animator [15].

It has already been shown that the most basic principles of key framing and slow in and slow out, have direct parallels in computer interpolation techniques. Likewise squash and stretch has a computer equivalent in solid shape deformation, which is of particular relevance to computer facial animation. In addition to the fundamental animation principles, rotoscoping is a technique used by cell animators that is also a valuable tool for the computer animator.

1.2.2 Squash and Stretch as Solid Deformation.

When living objects move, there is a considerable deformation of the shape during the action, whilst the volume remains approximately constant. Squash and stretch one of the most important facilities of traditional character animation [16]. Flesh is supple and stretches, bulges or sags in relation to forces acting upon it. For the face this is particular relevant during such actions as eating, talking and smiling.

Computing these characteristics can be achieved using shape interpolation. The principles of interpolation described above can easily be expanded into three-dimensions by applying the procedures to each dimension. Intermediate forms of the surface are achieved by interpolating the vertex between its extreme positions. For this procedure to work, the mesh describing the surface of the object must be the same for both extremes and the facet interconnection and definition must be identical. Figure 1 illustrates the interpolation of a polygonal surface of a wedge of clay transforming into the shape of a pot over twenty frames. In this example the extremes were created and a cosine interpolation used to calculate the intermediate vertex positions over time. This technique has been used extensively within computer animation because of its simplicity and precision of use [17]. While it is a powerful form of motion specification, it does have severe limitations that are discussed in section 1.3.

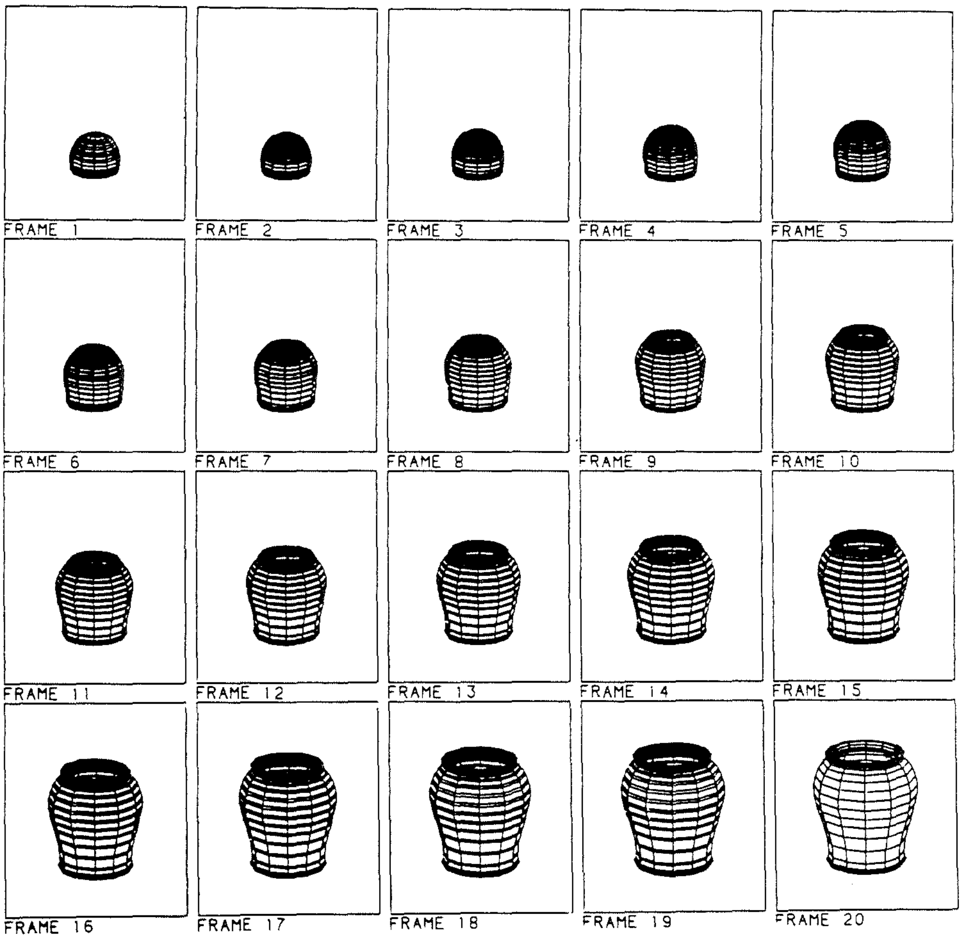


Figure 1

A three-dimensional wedge of clay transforming into a pot.

1.2.3 Rotoscoping.

Rotoscoping is an important technique used by cell animators to extract complex motion for analysis or direct copying. The process conventionally involves tracing live-action footage frame-by-frame onto cell. Traditional animators used rotoscoping primarily as a guide to help them achieve realistic animal motion and was used by the Disney animators in films such as Snow White and Sleeping Beauty because Disney believed that his films should be able to compete with live-action features of the era [18]. Film sequences were usually filmed from a single specified view to capture the essence of the motion, but it was sometimes necessary to record from a second viewpoint set up at ninety degrees or one hundred and eighty degrees to the first. Muybridge, over a brief period between 1884-1887, had captured over twenty thousand stereo-photographic images of a wide variety of animal locomotion, including the human activity, using two simultaneous stop-frame cameras [19]. Many typical locomotion characteristics were recorded and have been used as direct references in traditional animation, and in computer animation.

The principles of rotoscoping in three-dimensions are slightly different and have been applied in computer graphics [20]. The process involves recording live-action motion with three cameras to capture the x, y and z displacement activity of locations on the object´s surface. The individual views can then be rotoscoped and digitized, thus capturing the motion of the points on the surface in each dimension over time. As a result, every location will have data storing the motion in x, y and z space, which is used to displace objects in the computer. A similar process for capturing three-dimensional data of articulating bodies the use of potentiometers located at the joints of the figure [21]. In this way the displacement data is recorded automatically thereby removing the need for film rotoscoping. For the face, rotoscoping has been used primarily for lip synchronization. Here the lip shapes are traced off and either exaggerated or used directly [22].

1.2.4 Cell Animation of the Face.

For traditional animators the face was dealt with only after all the expression of ideas though the body had been exhausted. Only certain relationships help the expression to be read correctly. Frank Thomas and Ollie Johnston point out:

> We have found over the years that certain relationships make an expression read. No one part of the expression, the eye, or any other part by itself, is going to communicate if the expression does not work as a unit. To get this unity, there must be a close relationship between the crucial parts [23].

The view that no particular singular element of the face is paramount in the perception of an expression has been borne out in psychological research and by studies by other artists [24].

For traditional animators, the face was a flexible unit, in which the features of the face would act together to form expressions. The exaggerations of the eyes including the eyelids and brows in the characters meant that they had to be drawn with extreme care. A great deal of attention was given to the pupils, as any false move would destroy the communication and believability. Speech synchronization was an important characteristic for the mouth. Lip animation would be created from the phonetic breakdown of speech, that could be reduced to approximately seven or eight key structures [25]. Cartoon characters are primarily exaggerations of both human and animal forms, consequently a large latitude could be employed in the formation of mouth shapes, as we do not expect rabbits, mice or elephants to speak human languages.

The comprehensive skills of the traditional animators cannot be ignored. Again according to Thomas and Johnston, they were able to draw up rules for facial expression. To illustrate their understanding of their medium some of their basic rules were as follows: Ensure that you the right expression for what your character is thinking. Make sure that all the parts of the head and face relate to this one idea. Do not change shapes too much all over the face. At times, hold down the activity on the face so that just the mouth is moving [26].

1.3 Two-Dimensional Synthetic Faces.

Two-dimensional faces were the first type of facial images to be generated on a computer. One of the first was a system called WHATSISFACE, a non-animated system developed by Gillenson, which allowed a non-artist to create, on a graphic display, any male Caucasian facial image [27]. The development of the system was initiated as a replacement for the traditional photo-identikit and the artistic skills of a police artist. The system presented the user with an average face from which some seventeen paired features could be individually modified until the desired facial image was created in wire-frame on the screen.

Computer-generated two-dimensional facial images were also created by Summerfield for the analysis of the perception of visible, facial, articulatory movement in lip-reading [28]. Here a form of video rotoscoping was employed that captured the articulation of the lips during vowel sounds for the frontal and lateral views. Some five hundred low resolution frames of real lip articulation could be stored, sufficient to produce ten seconds of speech. These could then be digitized for analysis and played back in real-time.

1.4 Three-Dimensional Synthetic Faces.

The first attempts at three-dimensional facial animation were carried out by Fred Parke at the University of Utah, where a comprehensive system was developed that handled the majority of the problems involved in modelling and animation [29]. The resultant three-dimensional parametric model produced shaded facial images. These synthetic faces were constructed from polygonal surfaces and were manipulated through the use of parameters which control interpolation, translation, rotation and scaling of the various facial features. Parke suggested that fewer than ten parameters are required to produce reasonable facial speech synchronization and facial animation [30]. Further development by Pat Hanrahan into the scripting of the animation, allowed faces to be generated on a calligraphic system in real-time, such that parameters could control facial attributes [31].

The polygon mesh that described the facial form was input using a photographic technique, and the eyes, eyelids and teeth were modelled independently [32]. Interpolation, scaling and translation parameters described the conformation characteristics for the face. Interpolation was used to construct the facial shape for the forehead, cheekbone and chin, scaling was used for proportional distortion such as the chin-to-mouth and eye-to-forehead relationships and, finally, translation was used to offset the chin, nose and eyebrows. The parameters for jaw rotation, upper lip position, mouth width, eyebrow arch, eyebrow separation, eyelid opening, pupil size and eye tracking were described as being the most effective for facial expression and lip animation [33].

The strategy developed by Parke in his earlier work, involved initiating a detailed and specific topological mesh for the human face [34]. This has a two-fold advantage:

i. Firstly, it facilitates shape interpolation of faces with identical topologies. This was the earliest example of facial animation described by Parke, and Figure 2 an animated head by the present author illustrates the same mechanism applied to the transformation of facial expressions (this is the same process as that depicted in Figure 1 of a three-dimensional wedge of clay transforming into a pot) [35].

While interpolation produces very accurate results, the major draw-back concerns the storage of key-frame positions, as each key-frame must be explicitly defined and stored. A slight improvement can be achieved by storing just the differences between key positions [36]. However, the face may well articulate practically every vertex in each frame, thereby negating the advantage of storing just the differences. Another major deficiency of key-framing is that the system is only capable of generating distortions defined by the key positions. This too can be improved with minor modifications: additional key positions can be created from the interpolation of two existing extremes [37], but unfortunately, this still limits the system to the motion between the original extreme key positions.

ii. Secondly, the strategy enables specified vertices to be identified, grouped and manipulated in pre-defined

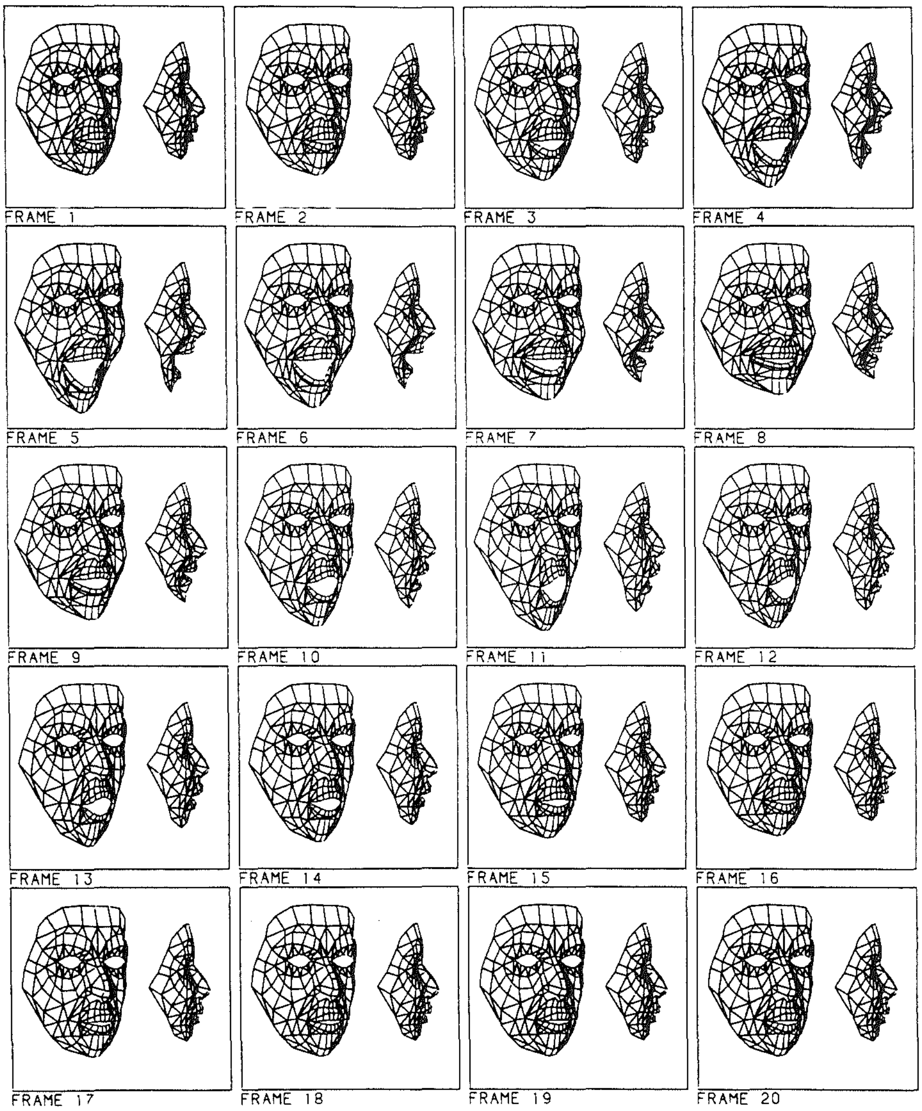


Figure 2

Three-dimensional face interpolation between static poses over twenty frames. The key frames are at one, five, eight, twelve and twenty.

procedures. For example, the motion of the corner of the mouth during a smile is a complex activity. This can be defined by the animator as a precise movement of the vertices at the corner of the mouth, while the surrounding designated vertices are moved by a pre-defined proportion of the original displacement to form the chosen articulation [38].

The major disadvantage of this approach is that it requires the complete definition of all the pre-defined procedures, such as lip raising, brow curving or jaw opening. This ´hard-wires´ the vertices to perform specified tasks that are limited to the particular definition.

Parke also considered that human faces could be encapsulated within a single generic and geometric facial topology [39]. The benefit of formulating such a system is that it facilitates the interpolation of the entire face between selected alternatives, but Parke´s reasoning is unconvincing. Gillenson had restricted his investigations to the heuristic development of faces in two-dimensions, and to prevent a profuse accumulation of variable parameters, the system had been further limited to male Caucasian faces. Consequently, while a total of seventeen variable parameters were used in Gillenson´s two-dimensional system, a hundred or so supplementary parameters would be required to produce a three-dimensional equivalent.

Additional problems are encountered when the animator attempts to manipulate the finer details of three-dimensional facial topologies, as facets cannot be created or destroyed. Figure 3 and Figure 4 the Fluck and Law puppets, Mrs Thatcher and the Queen, are exaggerations of public faces and they exemplify the dilemma: the two noses of the characters are radically different, both in shape and size. Consequently, describing a polygonal mesh for the dissimilar noses requires more detail on one than the other. Therefore, shape interpolation between the noses cannot function. Furthermore, the facet locations on one nose and those on the other do not correspond, and can therefore only be created by a tedious and personalized vertex twining [40].

An alternative approach was investigated by Platt and Badler for the simulation of the human facial movements [41]. Where Parke employed the movement of vertices by pre specified procedures, Platt implemented a more biologically accurate model for the activity of the skin as a network of interconnected arcs. The investigation was part of a larger effort towards an American Sign Language (ASL) recognizer

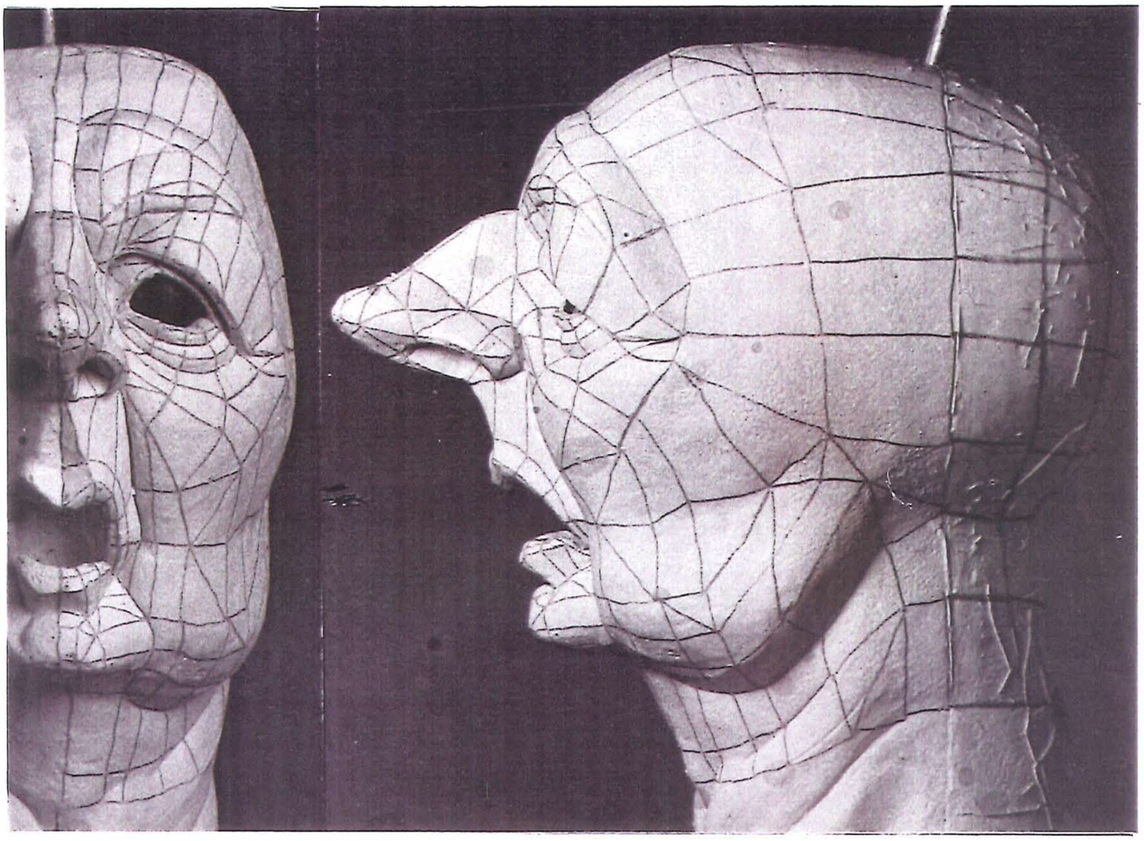


Figure 3(a)

The topological description of the Fluck and Law puppet Mrs Thatcher.

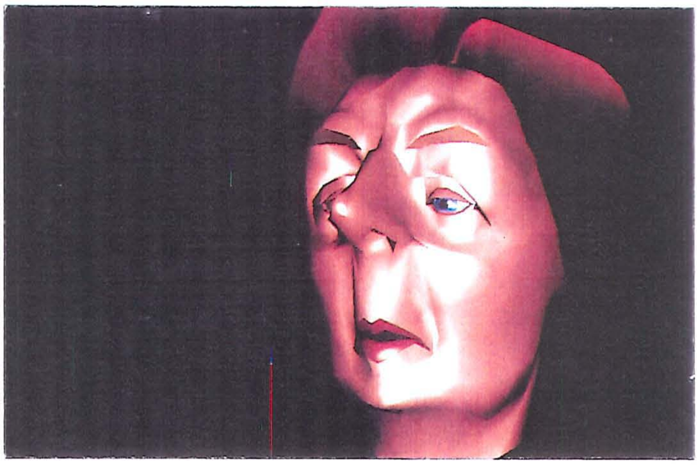


Figure 3(b)

A rendered image of the three-dimensional computer generated puppet.

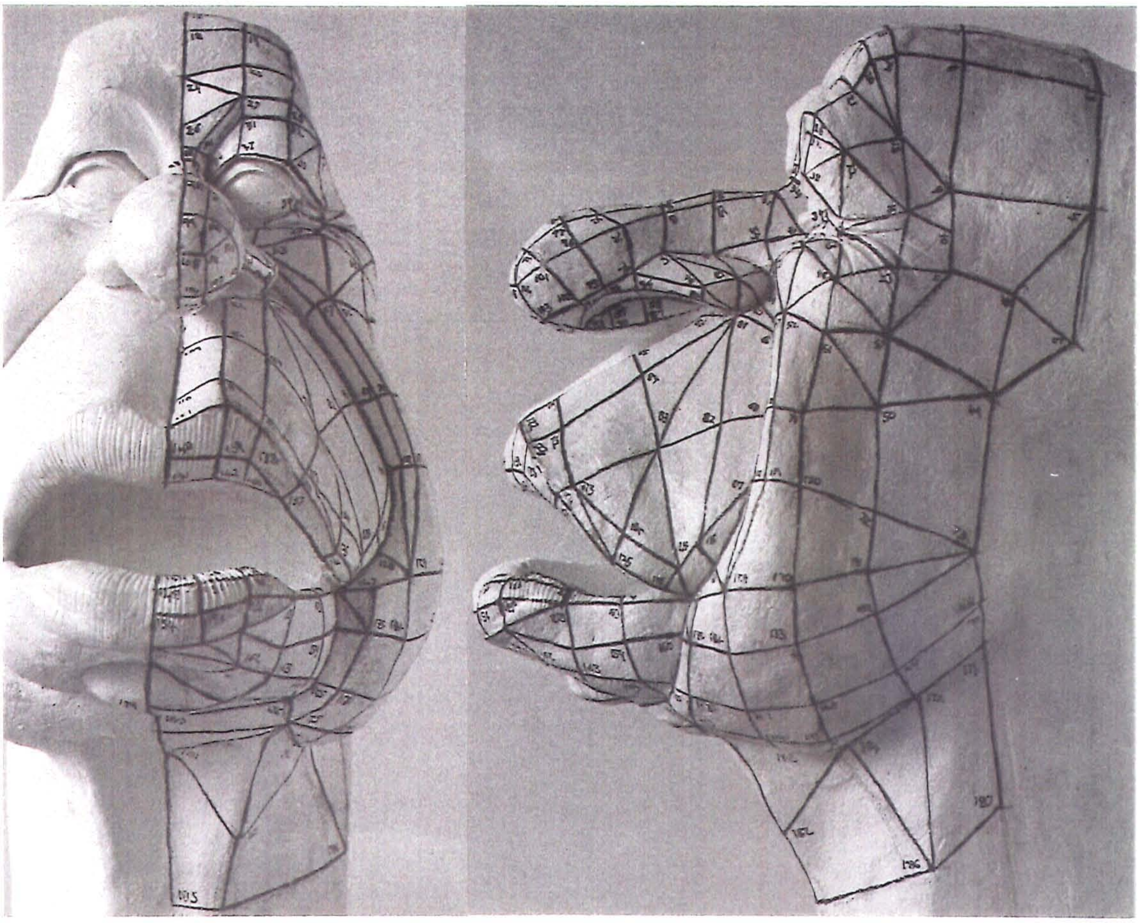


Figure 4(a)

The topological description of the Fluck and Law puppet the Queen.

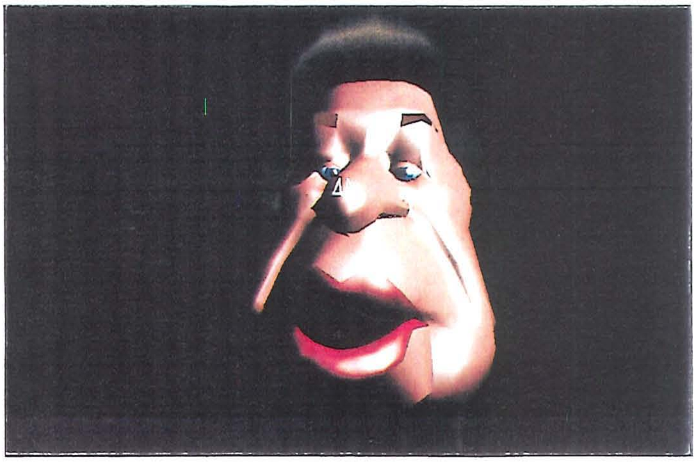


Figure 4(b)

A rendered image of the three-dimensional computer generated puppet.

and simulator, of which the face is an important feature. The facial action recognizer-simulator model consisted of three major sections: a camera processor which determined which facial actions were being performed, the Internal Model Manipulator to which the data was passed and where the motion representation converted into the structures required by the simulator, and finally, the Simulator which used this data to manipulate the model of the face to produce a suitable output. The implementation determined the physical surface of the skin as a warped plane of skin points connected to their neigbours by arcs. Each point in three-dimensions had associated information indicating its type - bone, muscle or skin - and a list of adjacent arcs. The principle of the algorithm determined the skin as a linked mesh of arcs such that when a force is applied at a node, the point is displaced as a function of the elasticities at that point. In this way the skin is treated as a rubber mesh and the following expression applied.

$$dl = f/k$$

where k is the elasticity constant, f is the force at that point, dl is the displacement from equilibrium. This expression determines the length change as a function of the applied force. These forces are then propagated outward until the iteration is completed across the face.

The approach by Platt and Badler concerns the motivators of the actions in terms of muscle actions and interactions. It is an improvement over the technique used by Parke in that it is more generalized and simulates muscle action on flesh. This allows new structures to be tested with a multitude of possible muscle combinations, whereas Parke's approach required knowledge of the displacement activity before features were animated.

An alternative approach to those of both Parke and Platt and Badler, was designed at the University of Montreal in the system called DADS and TRAANA. These systems were developed specifically for the non-programming artist to allow the creation of synthetic computer characters, and they were implemented for the computer animated short, Tonie De Peltrie [42]. Here expressions for the computer character were created from a Character Bank of Standard Expressions (CBSE). The CBSE was created from twenty standard expressions captured from a human face that was constructed as a three-dimensional polygonal mesh. The face of the computer character comprised a greater number of polygons than the human face, therefore instead of a one to one coincidence between vertices, a one to N correspondence was required that allowed the expression to be replicated from the human face onto the computer character. Further expressions and more subtle nuances could be created by blending two or more complete expressions together. This technique has the same inherent explicit definition problem as that associated with three-dimensional key framing.

1.5 Conclusions.

Many techniques of traditional cell animation can be employed in three-dimensional computer animation, and while direct duplication of techniques is futile, many characteristics can be implemented as basic tools for the computer animator.

Traditional animation techniques considered expression of the face only after all the other expressive possibilities of the form had been exhausted. When the face was considered, the basic principles of mass, squash and stretch, follow-through were used as described above.

It is evident from the description of the preceding systems that there are three fundamental approaches to three-dimensional facial animation:

i. The earliest approach of key frame interpolation between pre-specified positions. These keys are created from a real person and the animation is limited to the expressions created by the actor.

ii. As with the above, expressions can be recorded from a real person. These can then be mapped onto a synthetic model and further expressions created from blends between those keys recorded.

iii. To avoid the limitations of the key frame approach, parameterization can be employed to group characteristics of the face together to perform pre-specified tasks.

This present research does not use the key framing approach but instead employs parameterization, based upon a underlying understanding of the physical attributes of the face.

1.6 References.

[1] Fred Parke, ´A Parametric Model for the Human Face´, Technical Report UTECH-CSc-72-120, Department of Computer Science, University of Utah, 1974.

[2] Philippe Bergeron, ´Techniques for Animating Characters´, Advanced Computer Animation Course Notes, ACM SIGGRAPH, 1985, pp. 61-70.

[3] Stephen Platt, ´A System for Computer Simulation of the Human Face´, MSc Thesis, University of Pennsylvania, 1980.

[4] Frank Thomas and Ollie Johnston, Disney: The Illusion of Life, Abbeville Press, New York, 1984. p. 8.
Roy Madsen, Animated Film: Concepts, Methods, Uses, Interland Publishing, New York, 1969, p. 4.

[5] Frank Thomas, ´Can Classic Disney Animation be Duplicated on the Computer?´ Computer Pictures, Vol. 4, No. 3, July/August, 1984, p. 24.

[6] Edwin Catmull, ´The Problems of Computer-Assisted Animation´, Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 12, No. 3, 1978, pp. 348-353.

[7] Gunnar Johansson, ´Visual Motion Perception´, Scientific American, Vol. 232, No. 6, 1974, pp. 76-88.

[8] Ian Angell, A Practical Introduction to Computer Graphics, The Macmillan Press, London, 1981, p. 11.

[9] Nadia Magnenat-Thalmann and Daniel Thalmann, Computer Animation: Theory and Practice, Springer-Verlag, Tokyo, 1985, p. 44.

[10] Jane Wilhelms, ´Using Dynamic Analysis to Animate Articulated Bodies such as Humans and Robots´, Graphics Interface, Vol. 20, No. 3, 1985, pp. 107-114.

[11] Thomas and Johnston, Disney, pp. 30-31.

[12] Doris Kockanek, ´Interpolating Splines with Local Tension Continuity and Bias´, Computer Graphics Proceedings, ACM SIGGRAPH, Vol. 18, No. 3, 1984, pp. 34-41.

[13] John Lasseter, ´Principles of Traditional Animation Applied to Three-Dimensional Computer Animation´, Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 35-44.

[14] Thomas, *Disney*, pp. 15-41.

[15] Lasseter, ´Principles´, pp. 35-44.

[16] Thomas, *Disney*, p. 15.

[17] Susan Van Baerle, ´Computer-Generated Three-Dimensional Character Animation´, MA Thesis, Ohio State University, 1985.

[18] Tim McGovern, ´The Use of Live-Action Footage´, Course Notes, ACM SIGGRAPH, No. 5, 1987, p. 1-26.

[19] Eadweard Muybridge, *The Human Figure in Motion*, Dover Publications Inc., New York. 1955.

[20] McGovern, ´The Use of Live-Action´, pp. 3-8. gives as an example: Abel Image Research, *Sexy Robot*, C/G film advert for Council of Canned Food, 953 N. Highland Avenue, Los Angeles CA 90038-2481, 1984.

[21] Thomas Calvert and John Chapman, ´Aspects of the Kinematic Simulation of Human Movement´, *IEEE Computer Graphics and Applications*, Vol. 2, No. 9, November 1982, pp. 41-49.

[22] Keith Waters, ´Three-Dimensional Facial Animation´, Proceedings Online Wembley, 1986, pp. 49-56.

[23] Thomas, *Disney*, p. 301.

[24] Vicki Bruce and Tim Valentine, ´When a Nod´s as Good as a Wink: The Role of Dynamic Information in Facial Recognition´, *Practical Aspects of Memory: Current Research and Issues*, Vol. 1, Wiley, Chichester, 1988, pp. 169-174.
Stephen Rogers Peck, *Atlas of Facial Expression*, Oxford University Press, New York, 1987.

[25] Harold Whitaker and John Halas, *Timing for Animation*, Focal Press, London, 1982, pp. 122-128.
Bob Godfrey and Anna Jackson, *The Do-it-yourself Animation*, British Broadcasting Corporation Publications, London, 1974, pp. 73-74.

[26] Thomas, Disney, p. 321.

[27] Mark Gillenson, ´The Interactive Generation of Facial Images on A CRT using a Heuristic Strategy´, PhD Thesis, Ohio State University, 1974.

[28] Quentin Summerfield, ´Analysis, Perception and Analysis of Visible Articulatory Movements´, Journal of Phonetics, Vol. 11, 1982, pp. 63-76.

[29] Parke, ´A Parametric Model´, pp. 21-41.

[30] Fred Parke, ´A Parameterized Model for Facial Animation´, IEEE Computer Graphics and Applications, November, Vol. 2, No. 9, 1982, pp. 61-68.

[31] Hanrahan, ´Interactive Animation of Parametric Models´, Course Notes, Introduction to Computer Animation, ACM SIGGRAPH, 1985.

[32] Fred Parke, ´Measuring Three-Dimensional Surfaces with a Two-Dimensional Data Tablet, Computer Graphics, Vol. 1, 1975, pp. 5-7.

[33] Fred Parke, ´A Model for Human Faces that allows Speech Synchronized Animation´, Computer Graphics, Vol. 1, 1975, pp. 1-7.

[34] Fred Parke, ´Computer Generated Animated Faces´, Proceedings ACM National Conference, Boston, Vol. 1, 1972, pp. 451-457.

[35] Waters, ´Three-Dimensional´, p. 51.

[36] John Moore and Vincent O´Connor, ´Towards an Integrated Computer Package for Speech Therapy Training Talking Heads´, MicroTech Solutions Report, Bradford College of Art and Design, 1986.

[37] Bergeron, ´Techniques´, p. 65.

[38] McQueen, ´Applying Classical Techniques to Computer Animation´, Course Notes, ACM SIGGRAPH, Vol. 10, 1987, pp. 217-230.

[39] Fred Parke, ´Parameterized Models for Facial Animation´, Course Notes, ACM SIGGRAPH, 1975, pp. 38-51.

[40] Bergeron, ´Techniques´, p. 65.

[41] Stephen Platt and Norman Badler, ´Animating Facial Expression´, Proceedings Computer Graphics, ACM SIGGRAPH Vol. 15, No. 3, 1981, pp. 245-252. Platt, ´A System for Computer Simulation´, 1980.

[42] Bergeron, ´Techniques´, pp. 61-70.

CHAPTER TWO

Chapter Two

The Physical and Motion Attributes of the Face in Relation to Computer Synthesis

2.1 The Physical Attributes of the Face.

Previous attempts at the modelling and animation of the human face have only been concerned with the surface displacement characteristics [1]. While investigations by Platt suggested a more biologically accurate model by simulating skin as tension nets, his method could not handle the primary motions of the jaw and mouth and is therefore of limited use, especially when dealing with a model demanding lip animation [2]. One of the goals of this present research is to produce accurate modelling of the physical attributes and a more biomechanical approach to expressions. Therefore it is necessary to describe those universal physical attributes of the face required for a computer model and to group as many motion characteristics as possible so that they can be used in the animation synthesis. To achieve this, the physical attributes of the face and the cause motivators of expressions have to be examined.

This chapter considers two broad areas: the description of the static physical attributes and the motion characteristics that can be notated and graded. The primary concern is to delineate features and motions in terms of modular components that can be used for computer synthesis.

Therefore this chapter also reviews some techniques employed in face recognition and research in psychology, where expressions and emotions are analysed.

The human face consists of; bone, cartilage, muscle, epidermis and eyes, which possess their own unique characteristics that in turn determine the facial form and motional eccentricities. These are the fundamental building blocks from which the face is formed [3].

The face, considered by this present research, is the frontal view of the head from the base of chin to the hair line, and the front half of the head from the lateral view. These zones define the mobile areas of the head where facial expressions are formed. The ears, hair, lower neck and the back of the head are not modelled and animated by this present research for two reasons firstly, they do not form facial expression, and secondly, the hair and ears are an intricate and independent modelling problem.

2.1.1 Bone.

The skull consists of fourteen major bones of which the mandible is the only freely jointed structure. Despite the skull's immobility, the facial skeleton comprises very discrete sections:

i. The forehead consists of two large bones called the frontal bone, which forms the eyebrow ridge and the upper parts of the eyeball sockets.

ii. The eye socket, known as the orbit, is made up of a number of bones: the ethmoid, lacrimal and maxillary bones contribute to the generally conical space and the squarish outlets for the eyes, and they also determine the separation of the eyes themselves.

iii. The zygomatic bone is a prominent structure that makes up the cheek bones. They originate in the squamous part above the ear and progress round the face to form the base of the orbit.

iv. The maxillary bone structure forms the roof of the mouth, the floor of the nasal cavity and the floor of the orbit. It is the lower part into which the upper teeth are located.

v. The mandible, is a separate bone structure. It has a curved, horizontal body, which is convex forwards, and two broad rami, which project upwards from the posterior ends of the body. The lower teeth are located in the upper surface of the mandible. From the frontal view the teeth are visible and their importance should not be underestimated when animated speech segments are modelled (see 5.8.1).

The relative sizes, shapes and distances between the mandible, maxilla, nasal, zygomatic, orbital and frontal structures are almost infinitely variable, and it is this variation that makes up the unique character of each face. While the muscle and soft tissue of the skin radically change throughout life, it is the skull's structure that determines the general shape of the face that we recognize. For example the height of the maxilla alters the roundness of the face, but it does not otherwise change the individual's appearance. Liggett, in his study of the human face points out that the face is, without question, man's most diverse and most variable attribute. Not only are there great contrasts from infancy to adulthood, but also between races and within racial groups [4].

2.1.2 Muscle.

The muscles of facial expression are subcutaneous voluntary muscles. In general muscle arises from the bone and facia of the skull and insert into a deep facia layer which in turn is bounded to a layer of subcutaneous fat and the skin (Figure 5). When the muscles are relaxed, the subcutaneous fatty tissue and musculature fills the hollows and smooths the angular transitions so as to allow the general shape of the skull to be seen.

The muscles of the face vary according to the orientation of the fasciculi (the individual fibres of the muscle) and they can be broadly be divided into the upper and lower face.

i. Encircling the mouth is an elliptical muscle, the orbicularis oris, which has no bony point of attachment. Additionally the buccinator muscle originates at the rear of the mandible and wraps round this structure and separates at the modiolus of the mouth into the upper and lower lips. There are ten small muscles that are aimed at the mouth and modiolus, each having a variety of attachment points, for example the deeper levator anguli oris muscle.

ii. One of the most powerful muscles of the face is the masseter that originates from the underside of the

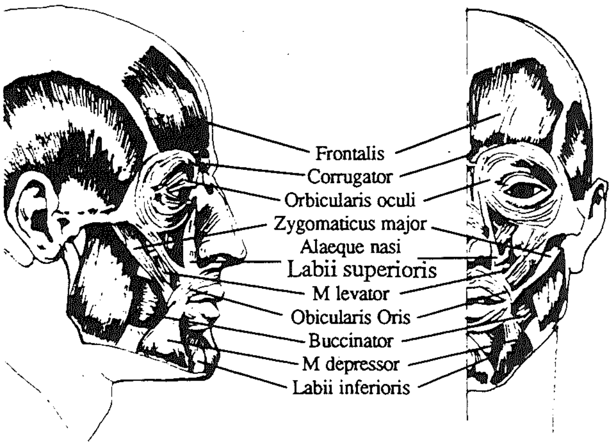


Figure 5 (a)

The muscles of the face from a lateral and anterior view.

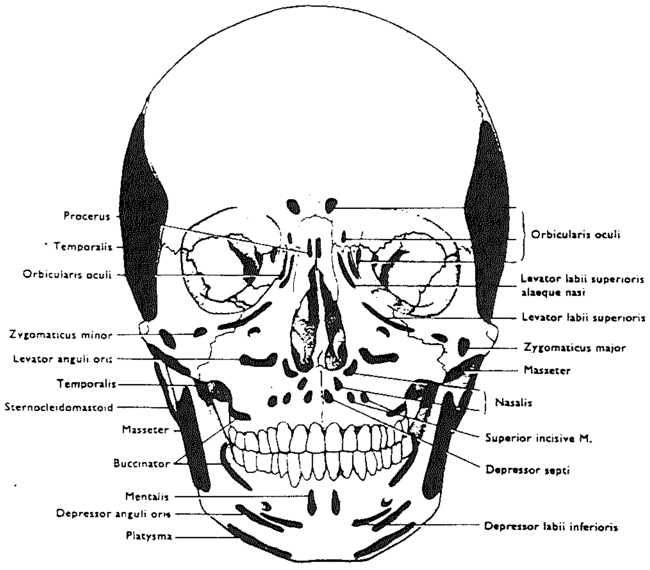


Figure 5 (b)

The anterior view of skull illustrating the muscle attachments.
From: G J Romanes, Cunningham's Manual of Practical Anatomy, Vol 3: Head, Neck and Brain. Oxford Medical Publications, 1967, p.11.

zygomatic arch and then inserts into the mandible. The masseter elevates the mandible to aid the teeth in mastication.

iii. The other elliptical muscle is the orbicularis oculi that fits into the eye socket to contract and screw up the eye.

iv. The last prominent muscle of the face is the frontalis muscle, this is a broad sheet muscle that is divided vertically and curves from the high frontal zone and is firmly attached into the frontal brow structure.

2.1.3 Skin.

The skin is composed of two layers - the dermis covered by the epidermis. Beneath the dermis is the subcutaneous fat which in turn is bound to underlying deep fibrous fascia that is connected to muscle or cartilaginous surface [6].

The skin can be subjected to considerable mechanical stress from internal and external forces, as a result the epidermis interface with the dermis is marked by ridge/groove interdigitations. This arrangement distributes forces over large areas and also prevents the epidermis from being stripped off by shearing forces.

The dermis is tough, flexible and highly elastic, deforming over and around the underlying structures. With the progression of age the epidermis and dermis lose their elasticity resulting in the deepening of flexure lines that are either folds in the dermis associated with habitual joint movement, or lines of attachment to the underlying deep fascia.

The appearance of the facial skin can be seen to be dependent upon several factors:

i. The texture of the skin that is determined to a large extent by the glands contained within the skin.

ii. The depth of the dermis that varies from a delicate thin covering for the eyelids, to a thick layer beneath the mandible that is further supported by a dense layer of subcutaneous fatty tissue. These characteristics not only affect the appearance of the skin but also the motion characteristics.

iii. The colour of the skin is determined by the blood circulation and the presence of pigments. This varies enormously from infancy to adulthood and between males and females, who tend to have larger fat deposit than males.

iv. The tension in the skin that is caused by the action of muscles. This can also affect the colour, with the lack of blood reaching the surface of the skin giving, for example, the characteristic white knuckles when the hand is clenched and, as the complete opposite, blood forced to the surface of the skin causing blushing. In extremes, such as fear, blood can be removed from the skin to give the appearance of 'white as a sheet'. Health and disease also affect skin colour and, dependent upon the severity of the infirmity, physical degradation [7].

2.1.4 The Eyes.

The eyes are considered, by this present research, to be an important and independent structures of the face. The eyeball itself consists of three distinct elements: the sclera or white of the eye, the iris and the pupil. The primary motion attributes are pupil dilation and eye convergence.

More subtle characteristics of the eyes are colour and pupil dilation [8]. Fundamentally the colour of the eyes are consistent. However, during excitement or drowsiness the eyes can sparkle or glaze over [9] and subsequently these characteristics can be governed by the size of the highlight. More dramatic movements occur with the dilation of the pupil: the Oriental jewellery traders veiled their eyes to conceal their real enthusiasm as their eyes involuntary dilated in excitement at a particular item [10]. These dilations could be clearly seen and understood by their fellow traders who would raise the price for that particular article.

The eyes play an important role in maintaining attention during conversation, consequently the following parameters are calculated as the minimum set of required parameters:

i. Pupil dilation
ii. Eye convergence
iii. Highlight positioning
iv. Colour

Of these characteristics, eye convergence is the most significant, while the others are subtle attributes.

The human conjunctiva is maintained by blinking the eyelid, and the combination of this with the convergence of the eyes characteristics makes them a very expressive channel of communication. For example, when two people talk, the speaker may not look directly at the listener, and only when he/she comes to the end of the sentence do the eyes look directly at the listener for a response. These definitions of visual behaviour have been described as:

Term	Definition
Onesided look	Gaze by one person's in the direction of another's face
Face-gaze	Direction of one person's gaze at another's face
Eye-gaze	Directing of ones person's gaze at another's eyes.
Mutual look	Two persons gaze at each other's face
Eye contact	Two persons look into each other's eyes and are aware of each other's gaze

Gaze avoidance	Avoidance of each other's gaze
Gaze omission	Failure to look at another without intention to avoid eye contact [11].

One of the expressive characteristics of the eyelids concerns the revealing of the white of the eye. For example, in the expression of fear, the white of the eyeball may appear both above and below the iris, so that the white of the eye is predominant. Also, when the eyes are drowsy, the lids are half closed.

2.2 The Bone and Muscle Motions.

The skull is the basic building block onto which the muscle and skin tissue is attached. The mandible is a mobile bone and articulates with the temporal bone to form the temporomandibular joint which creates a complex three-dimensional envelope of movement. Most of the articulations are controlled by the lateral and medial pterygoids, digastric, temporalis and masseter muscles. The depression movement is the most obvious movement and is limited to approximately forty degrees down, while the upper limit is reached when the teeth are brought into contact with each other. The temporomandibular joint also has four other important movements - elevation, protrusion, retraction and lateral. However to simplify the computer model the temporomandibular joint is treated as a single horizontal rotary action.

The majority of the facial muscles are concerned with movement with little force. This is in contrast to the majority of muscles in the rest of the body. The face has two distinct types of muscle: elliptical muscles that squeeze and linear muscles that pull. These can be broadly divided between the upper and lower face to give five major groupings:

i. Uppers and downers, that move the face upwards towards the brow and conversely towards the chin.

ii. Those that contract horizontally towards the ears and conversely towards the centerline of the face.

iii. Oblique muscles that contract in an angular direction from the lips, upwards and outward towards the cheek bones.

iv. The elliptical muscles that run round the eyes and mouth.

v. Sheet muscle for example the platysma muscle that extends downward into the neck close beneath the skin.

The upper facial muscles are responsible for the changing appearance of the eyebrows, forehead and the upper and lower lids of the eyes. The muscle on the forehead contracts towards the point of static insertion into the skull, and as a result the surface tissue buckles and wrinkles creating furrows in the epidermis.

The mouth region, defined as the area under the nasal cavity and extending to the ear, has the most complex muscle interaction and motions of the face. The primary muscle that has no bony attachment is the orbicularis oris which is a squeezing/sphincter type muscle that squeezes the lips into puckered distortions. These shapes are used in the vocalization of words, and during eating and drinking. The muscle acts in a similar manner to the action to the closing of a string bag that concertinas together.

2.3 Employing Facial Recognition.

The essential individuality of a face depends upon a great deal more than just physical attributes. Research into facial recognition reveals key features of the face that are particulary useful for computer synthesis [12].

Leon Harmon, one of the pioneers of computer facial processing, developed some early techniques for using image processing for recognition [13]. One of the experiments involved the computer selection of target faces from a selection of stored facial images. This meant establishing a pool of features that could be judged quantitatively and these features had to be independent of one another, so that each description was of pertinent and distinctive value. Originally thirty-five separate features were used to describe a total of two hundred and fifty six photographs of frontal, lateral and three-quarter views of faces. Each feature could be graded on a scale usually from one up to a maximum of five. For example, the nose length could be described as being short 1, medium 2 or long 3. The results indicated that twenty-one specified features could accurately identify a person's face.

The feature classification described by Harmon is by no means the first attempt at facial feature classification. Alphonse Bertillon, known for his classification of fingerprints, meticulously described and named features with associated photographs and sketches in a system that is still in use today. For example, the ridge of the nose in profile may be described as concave, convex or rectilinear [14].

Much of the investigation by Harmon was used in an interactive generation of two-dimensional facial images on a computer employing a heuristic strategy by Gillenson [15]. The priority in this environment was to define a reasonably minimal feature set that could be geometrically manipulated by software algorithms to form the desired shape. A total of seventeen features were described: chin, cheek lines, ears, chin lines (auxiliary), eyebrow outlines, eyelids lower, eyelids upper, eyes, forehead lines, hair outline, lower lip, upper lip, mouth lines, naso-labial lines, neck, upper nose and lower nose.

An important note of caution must be made at this juncture concerning facial feature recognition. While the direct qualitative classification of facial features by shape, positioning and size provides useful structures for computer synthesis, it is not the only information used for facial recognition. More recently Bruce and Young have argued that seven distinct types of information can be employed when people recognize faces; pictorial, structural, visually derived semantic, identity-specific semantic, name, expression and facial speech codes [16]. Therefore, the hypothesis that facial recognition is dependent solely upon the arrangement of the features in respect to one another, as in the investigations of Gillenson, is only one approach typically used by police artists and in photo-identikit procedures [17]. Unfortunately, Gillenson's selection of features is not based upon any strict fundamental ordering, for example, most salient feature to least important feature. In fact he points out that the feature list was based on his own personal judgement.

Feature lists are employed in this present research not because they are ideal, but because the quantitative breakdown is easier to use in computer synthesis than the other processes described by Bruce and Young [18]. However, what must be kept in mind is that the evidence clearly indicates that the feature list approach cannot assume priority to the exclusion of all others when determining whether the face can be recognized. Other investigations by Ellis have tried to establish whether the face is treated any differently from other forms of recognition, but the conclusion is that there is no unambiguous evidence to support the theory [19].

Lists have been produced such that the saliency of facial features has been loosely ranked in order of importance by Laughery as: eyes, nose, lips/chin, hair and ears [20]. Obviously this is dependent upon the situation in which they are observed [21].

2.4 Previous Definitions of the Facial Form.

Defining an average facial form is subjective, but it has been the desire of artists through the centuries. Perhaps the earliest consideration of facial proportion was given by the Greeks who suggested that proportion was synonymous with beauty [22]. They utilized the golden section in architecture, sculpture and painting, and they believed all beautiful objects could be analysed by dividing the object into successive parts such that the smallest division was in the same ratio to the largest as the largest was to the whole. Consequently in the perfect face the brow would be one third of the way down from the hairline and the mouth one third of the way up from the chin.

The Renaissance artists of the sixteenth century such as Durer and Da Vinci were also preoccupied with human proportion. They did not use the golden section but division into sevenths. Durer believed that the nose was the critical factor in determining facial shape and made numerous drawings to illustrate this fact (Figure 6) [23].

Some more recent applications in ergonomic design by Dreyfuss indicated gross facial sizes by describing averages from maximum and minimum measurements of a large cross-section of adults [24]. This procedure can be used to describe limits to the overall head size in the computer synthesis.

Identifying and modifying the position of features by this method is also useful for computer synthesis (see 3.4.1). A predefined mesh over the existing form can then be distorted and an ageing process applied to modify the cranial features [25].

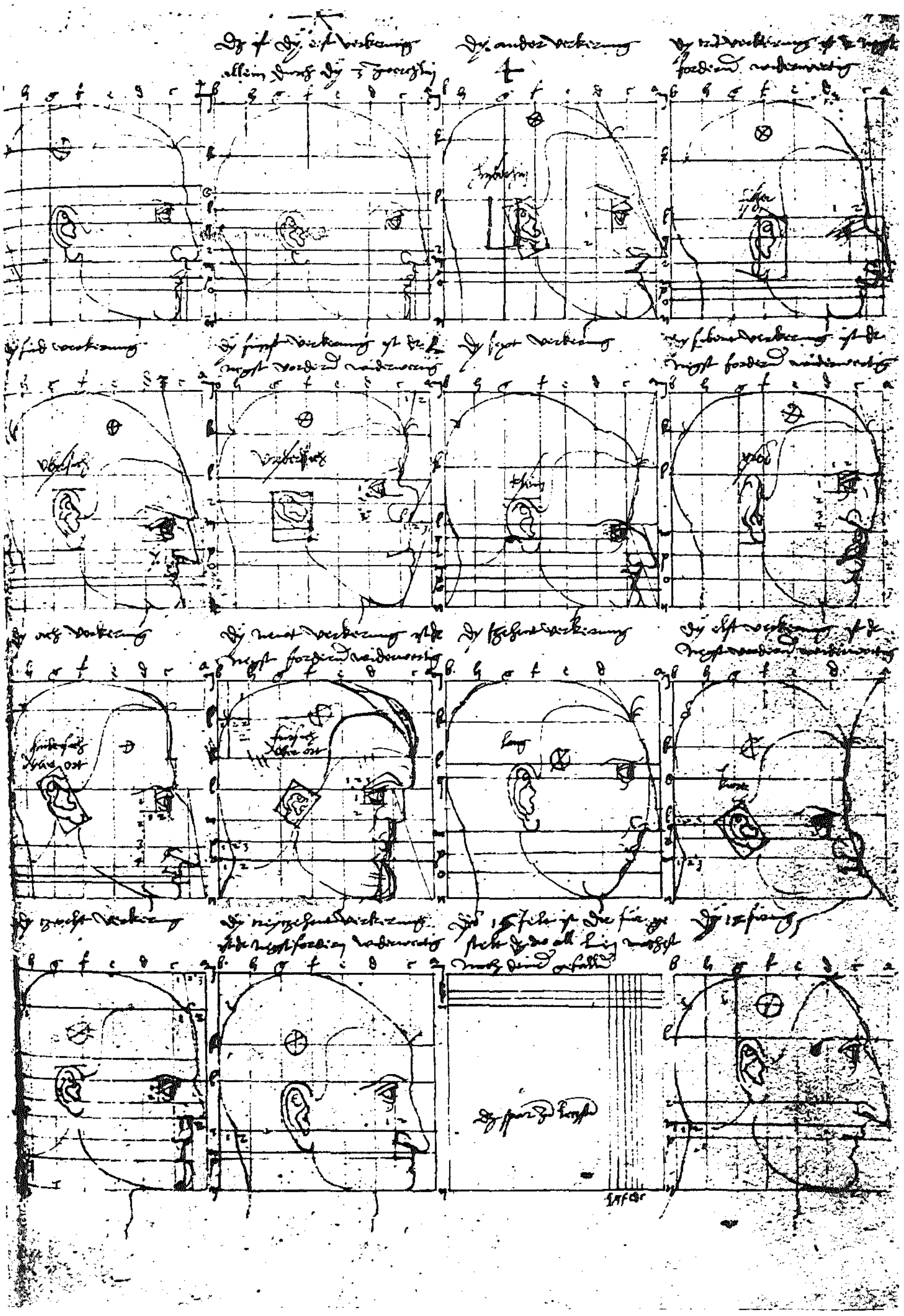


Figure 6

A page from Dürer's note book: The Four Books on Human Proportion. The lateral view of the face is divided into horizontal sections at key features of the face.

2.5 Notation Systems for the Human Face.

The motional attributes of the face are considered to be independent of the static characteristics by Bruce and Valentine and by Ekman and are considered to be so in this present research [25]. A large body or research has established, to some degree, that facial expression of emotions is universal, not only within racial groups, but also from culture to culture [26].

The facial motions require some mode of specification, grading and notation for computer synthesis, and the research by psychologists of non-verbal communication reveals key information.

Labanotation which has been used to notate the choreography of dance, is one potential mode of motion specification [27]. Labanotation basically records the motions of the human body over time and allows a description of any gesture as symbols that can easily be translated into alphanumeric equivalents [28]. The specification of motion for individual body parts is similar to the exposure sheet used in traditional animation to choreograph animation [29]. Unfortunately, Labanotation has little viable application to the notation of synthetic computer facial articulations.

Birdwhistell designed a system to code all aspects of human movement. It divided the body up into eight major sections: (i) total head, (ii) face, (iii) trunk (iv) shoulder, arm and wrist, (v) hand and finger, (vi) hip, leg and ankle, (vii) foot, and (viii) neck [30]. For the face the scheme had a large vocabulary of pictorial symbols for the expressions of the face. For example, the ´square smile´ would be represented as a array of two by three rectangles to represent the exposed teeth. While this scheme provides visual cues for synthesis, it lacks generalization and does not consider the biomechanics of the facial actions. Consequently it is not a suitable notational system for computer synthesis.

Significant work by the psychologists of non-verbal communication Ekman and Friesen developed the Facial Action Coding System (FACS), a notational-based system that is unlike previous systems such as Labanotation [31]. Individual muscles, or groups of muscles, are described as Action Units (AU) that distort the skin tissue. A full listing of the sixty-six independent Action Units and the associated thirty-five muscle names is given in Appendix A.2.

FACS is the best technique for the extraction of facial parameters useful for computer synthesis. The sixty-six independent muscle actions can give rise to several thousands of facial distortions. The facial muscle activities can be observed on faces, but contracting any individual Action Unit alone is not usually visually communicative. The Facial Action Coding System is not concerned with the muscle motion characteristics connected with expressions, it just notates and grades the individual muscle actions from static poses. This present research employs action units to create the static muscle contractions to mimic posed expressions. In addition a biomechanical understanding of the muscular contraction permits this present research to create time variant expressions that are visually more communicative, and therefore expands the functions of the Facial Action Coding System.

The six categories described by Ekman and Friesen - anger, surprise, disgust, fear, sadness and happiness - all use multiple combinations of the Action Units. Collectively the Action Units describe the set of all possible basic actions performable on the human face. For example, Action Unit 1 describes as the ´inner brow raiser´ in the upper face, contracts the inter frontalis muscle. AU9, known as the ´nose wrinkler´, activates the levator labii superiors alaeque nasi causing the nostrils to dilate, pulling the skin around the base of the nose upward and sometimes raising the upper lip.

The advantages of this type of description over those previously described are two-fold. Firstly, AU´s are designed to be closely connected with the anatomy of the face and this permits a more biomechanical approach and, secondly, FACS is based upon single muscle contractions, and on contractions by small groups of muscles that distort the face.

This present research ascribes to individual muscles, or groups of muscles, parameters that remain consistent between one face and another, in the same way that FACS is universal across a spectrum of facial types. Importantly, any differences between the computer parameters and real facial expression can easily be compared and corrected using the principles of Action Units.

2.6 Establishing Facial Expressions for Computer Synthesis by the Definition of Parameter Sets.

Real facial expressions can be formed and deciphered with the aid of FACS. The next task is to establish those characteristics required for computer control.

Parameterization is a desirable method of generating and controlling complex articulated models. Inanimate objects, such as the geometric primitives, cube, cone, sphere, can be described in terms of width, length, height, diameter, colour, weight and material, that represent basic parameters. The advantage of this approach is that it allows concise criteria to encapsulate every member of that group or class.

Few living forms can be determined by such precise parameters. Trees and other recursively generated forms seem to be the only objects belonging to such bounded sets, which can be grown from a small kernel of data that is easy to reproduce [32]. Unfortunately, the inherent nature of the face does not allow the formation of such discrete criteria and therefore belongs to an unbounded class of objects. Terminating the description of an unbounded set of objects is vague, and this is usually discerned in the resulting visual image.

Isolating the appropriate parameters to use for the face is perplexing but fundamental. For facial expressions the task is to establish salient time variant parameters. The growing, evolving and ageing processes, are long time variant parameters and are not considered by this present research. However they do constitute a vital area of research.

There has been little research in the recording of dynamic information during expressions. Baker-Shenk investigated behaviours other than linguistic functions to establish whether or not they are valid modes of discourse [33]. FACS was utilized with a slight modification to record the onset, intensity and offset of the actions. Video tape sequences recorded not only the speakers´ faces but also eye gaze, head, torso and arms. These were noted on a multi-level ´timeline´ that recorded what happened and when it occurred. Microanalysis of the results indicated a strong regularity between question types and Action Units. For example, ´yes-no´ types of questions are predominantly associated with the Action Units 1+2+5x and head forward/downward. This is particulary useful for the development of functional synergies, where grouping activities together enables the system to encapsulate more activities of a gesture or expression [34]. Unfortunately, the investigation did not reveal analytical information concerning the motion activity of the individual muscles.

The individual muscles beneath the skin (and the deeper layered muscle) have not been accurately measured. Work by the nineteenth century physiologist Duchenne applied electrical currents to freshly guillotined heads to observe the facial distortions, and later Duchenne applied the same technique to an old inmate of an alms house, who could feel no pain, to create artificial expressions [35]. His experiments tended to be arbitrary, as it was difficult to isolate individual muscles among the large groups of muscles which were usually activated simultaneously.

A number of recordings have been made of the surface displacement of the mouth region for the analysis of articulatory gestures for the deaf and hard of hearing. Investigations by Summerfield recorded the mouth on video tape and three major problems were encountered [36]. Firstly, an axis frame had to be defined to which the measurements of movement could be referred. Secondly, movements of the primary articulators such as the lip and jaw had to be separated from the effects of global movements. Thirdly, the measurements had to be sensitive since, relative to the size of the head, significant articulatory excursions are small and seldom exceed about twenty-five millimeters. Despite these inherent problems, reasonable results were obtained to describe the surface displacements of the skin.

A similar investigation by Fromkin into the lip positions in American English vowels measured the displacement activity of the lips by X-rays, plaster casts and photographs [37]. While both Fromkin and Summerfield´s investigations produced excellent quantitative results, the displacement activities of individual muscles were not recorded, only extensions of the lips and teeth. The measurement of muscle activity has to be performed on a living human, which is extremely difficult [38]. Fromkin did some preliminary electro-myographic studies from the dissection of the lip muscles of a cadaver and the results were not documented and used only to support the evidence produced by living humans [39].

Finally, investigations by Bruce and Valentine have indicated that we are far more capable of recognizing expression in a motion format than from a static photograph [40]. In fact our ability to perceive motion is quite acute, so that with only limited information, complex motion can be observed [41]. Therefore, to convey convincing facial expression on the synthetic facial image the animator requires only a few selected motion characteristics.

2.7 Conclusions.

The two attributes of the face, the physical features and motion characteristics, can be considered two completely separate issues. Both present a multitude of singular and distinctive qualities from which it is possible to establish some fundamental parameters.

The diversity and complexity of the facial bone, muscle and skin structure does not permit clearly defined, or bounded set of parameters to encapsulate all facial types. Despite this inherent lack of potential parametric definitions, a limited mode of face construction can be defined. The eyes, nose, mouth, forehead and chin are salient features that can be manipulated by the geometric transformations - translation, scaling, shearing and rotation - until the desired face shape is derived. Defining exactly where and how these transformations operate on the individual feature is open to question, and it is doubtful whether a broad range of facial types could be constructed employing this technique. This present research therefore, treats each face as a singular and unique structure.

Describing parameters for the motion attributes is the most complex issue. To achieve any form of motion realism, be it simulation or animation, a clear understanding of the underlying anatomy of the face is required. The cranium and mandible (the one articulating structure) supports the three fundamental muscle types, linear, sphincter and sheet which are further divided into the upper and lower face. These muscles have been notated and graded with the Facial Action Coding System (FACS) that breaks down complex facial actions into individual muscle actions known as Action Units (AU´s) and is the most comprehensive system for facial notation.

The Facial Action Coding System (FACS), is not however, concerned with the muscle motion characteristics. It simply notates and grades the individual muscle actions from static poses. But provided the muscle motion activity can be described in terms of the individual biomechanical behaviour of the linear, sphincter and sheet muscles, the

Facial Action Coding System provides the most powerful and comprehensive system useful for synthetic computer facial animation.

Reducing of the activity of the face into small modular units resolves the problem of formulating complex parameter sets. Parameters are only required to define the three individual muscle types that are independent of the facial type. In addition facial expressions that comprise groups of muscle acting together can be derived independently of the particular facial type.

2.8 References.

[1] Fred Parke, 'A Parameterized Model for Facial Animation', *IEEE Computer Graphics and Applications*, Vol. 2, No. 9, November 1982, pp. 61-68.
Phillipe Bergeron, 'Techniques for Animating Characters', Advanced Computer Animation Course Notes, ACM SIGGRAPH, 1985, pp. 61-80.

[2] Stephen Platt, 'A System for Computer Simulation of the Human Face', MSc Thesis, University of Pennsylvania, 1980.

[3] R Warwick and P L Williams, *Gray's Anatomy*, 35th Edition, Longman, London, 1974, pp. 496-503.

[4] John Liggett, *The Human Face*, Constable, London, 1974, p. 4.

[5] R Warwick, *Gray's Anatomy*, pp. 490-496.

[6] Ibid., pp. 1159-1169.

[7] Liggett, *The Human Face*, pp. 233-239.

[8] John Liggett, *The Human Face*, Constable, London, 1974, pp. 263-264.

[9] Stephen R Peck, *Atlas of Facial Expression*, Oxford University Press, New York, 1987, p. 27.

[10] Liggett, *The Human Face*, p. 263.

[11] Robert G Harper, Arthur N Wiens and J D Matarazzo, *Non Verbal Communication "State of The Art"*, John Wiley, New York, 1978, pp. 171- 212.

[12] Leon Harmon, 'The Recognition of Faces, *Scientific American*, Vol. 229, No. 5, November 1973, pp. 70-82.
H D Ellis, 'Recognizing Faces', *British Journal of Psychology*, Vol. 66, 1975, pp. 409-426.
Vicki Bruce and Andy Young, 'Understanding Facial Recognition', *British Journal of Psychology*, Vol. 77, 1986, pp. 305-327.

[13] Harmon, 'The Recognition of Faces', pp. 70-82.

[14] Alphonse Bertillon, *Signaletic Instructions*, Werner, Chicago, 1893.

[15] Mark Gillenson, 'The Interactive Generation of Facial Images on A CRT using a Heuristic Strategy', PhD Thesis, Ohio State University, 1974.

[16] Bruce, 'Understanding Facial', pp. 306.

[17] Gillenson, 'The Interactive', pp. 15-17.

[18] Bruce, 'Understanding Facial', pp. 306.

[19] Ellis, 'Recognizing Faces', pp. 409-426.

[20] R Laughery, 'Recognition of Human Faces: Effects of Target Exposure Time, Target Position, Pose Position, and Type of Photograph', Journal of Applied Psychology, Vol. 55, 1971, pp. 477-483.

[21] Bruce, 'Understanding Facial', pp. 305-327.

[22] C Bragdon, The Beautiful Necessity, Reprinted Quest, London, 1978.

[23] L Strauss, The Human Figure by Albert Dürer, Dover Publications, New York, 1972.

[24] Henry Dreyfuss, The Measure of Man Human Factors in Design, Second Edition, Whitney, New York, 1967.

[25] James Todd, L Mark, R Shaw and J Pittenger, 'The Perception of Human Growth', Scientific American, Vol. 242, February 1982.

[26] Vicki Bruce and Tim Valentine, 'When a Nod's as Good as a Wink: The Role of Dynamic Information in Facial Recognition', Practical Aspects of Memory: Current Research and Issues, Vol. 1, Wiley, Chichester, 1988, pp. 169-174.
Paul Ekman, 'About Brows: Emotional and Conversational Signals in Human Ethology', Journal of Phonetics, No. 11, 1983, pp. 169-202.

[27] Paul Ekman, Darwin and Facial Expression: A century of Research in Review, Academic Press, London, 1973, pp. 169-220.

[28] Thomas Calvert and John Chapman, 'Aspects of Kinematic Simulation of Human Movement', IEEE Computer Graphics and Applications, Vol. 2, No. 9, November 1982, pp.

41-52.

[29] J Menosky, 'Recording for the Future, Science, 1982, pp. 60-63.
Victoria Preston-Dunlop, Practical Kinetography Laban, MacDonald and Evans, London, 1969.

[30] Roy Madsen, Animated Film: Concepts, Methods, Uses, Interland, New York, 1969, pp. 143-149.

[31] R Williams, Contact: Human Communication, Thames and Hudson, London, 1981.
L Birdwhistell, 'Kinesics and Context', University of Pennsylvania Press, Philadelphia, 1970.

[32] Paul Ekman and Wallice Friesen, Manual for the Facial Action Coding System, Consulting Psychologist Press, Palo Alto California, 1977.

[33] Peter Oppenheimmer, 'Real-Time Design and Animation of Fractal Plants and Trees', Proceedings ACM SIGGRAPH, Vol. 20, No. 4, 1986, pp. 55-64.

[34] Charlotte Baker-Shenk, 'Nonmanual Behaviour in Sign Languages: Methodological Concerns and Recent Findings', Nonmanual Behaviours, Vol. 2, 1983, pp. 175-184. This research has indicated that facial expression communication can only be clearly understood with associated cues such as speech and gesture that is emphasized in melodramatic performance. Often in dramatic situations, the vocal expression is not so important as the gesture and facial expression.

[35] Keith Waters, 'Towards More Autonomous Control for Facial Animation', Proceedings British Computing Society, Displays Group, 1987, pp. 1-10.

[36] Liggett, The Human Face, p. 206.

[37] Quentin Summerfield, 'Analysis, Perception and Analysis of Visible Articulatory Movements', Journal of Phonetics, Vol. 11, 1974, pp. 63-76.

[38] Victoria Fromkin, 'Lip positions in American English Vowels', Journal of Language and Speech, No. 7, 1964, pp. 215-225.

[39] Ekman, Manual for the Facial. They investigated

individual muscle actions for FACS. Not all the muscle actions could be separated and consequently, they observed the changes in the surface of their own faces when a needle was placed into the muscle and delivered an electrical current. The work with this technique was only used in extreme situations, as needle insertions were painful.

[40] Fromkin. 'Lip Positions', p. 224.

[41] Bruce, 'When a Nod', pp. 169-174.

[42] Gunnar Johansson, 'Visual Motion Perception', Scientific American, Vol. 232, No. 6, June 1975, pp. 76-88.

CHAPTER THREE

Chapter Three

Modelling the Face as a Three-Dimensional Topology

3.1 Constructing the Face as an Irregular Natural Free-Form

3.2 Data Acquisition Techniques

3.2.1 The Lofting of Cross Sections

3.2.2 Light Beam and Laser Techniques

3.2.3 Mechanical Digitizers

3.2.4 Stereophotogrammetry

3.2.5 Discussion

3.3 The Construction of Three-Dimensional Faces for this Present Research

3.3.1 The Upper Face Construction of the Eyes, Eyelids and Brows

3.3.2 The Lower Face Construction of the Lips and Teeth

3.4 Geometric Transformations and Flexible Moulding

3.4.1 Solid Deformation

3.4.2 Parametric Patch Model Definition

3.5 Conclusions

3.6 References

3.1 Constructing the Face as an Irregular Natural Free-Form.

The first task involved in facial animation is to create the model, and this can be achieved by a number of the techniques discussed below. This present research adopts the approach that the face is best assembled from groups of polygons, which are the fundamental building blocks that support more complex and intricate modelling procedures. Consequently, the modelling techniques are essential to the system but are determined independently.

Describing three-dimensional body surface topographies such as the feet, legs and face while preserving a relationship to underlying forms is an area of current research [1]. This is not a simple task, as body surfaces are typically doubly curved, and the forms have no known equations permitting their duplication by computer aided techniques. However, compound surfaces are not unique to the human form, and are frequently found in CAD/CAM (Computer Aided Design and Manufacture) and CAGD (Computer Aided Geometric Design) applications, such as the design of car bodies, ships hulls, aircraft fuselages, bottles and numerous small machined elements such as engine components [2].

Manufactured objects provide engineers with elaborate compound surfaces that require an automated design process, and one approach for CAD/CAM involves boolean operators [3]. Because machined components belong to a class of objects having certain well-defined properties, three-dimensional models can be constructed from geometric approximation of elementary mathematical primitives such as cubes, cones and spheres. Furthermore, describing these 'primitives' as geometric solids offers a convenient way of constructing components step-by-step by employing a repertoire of elementary solid primitives as building blocks. Some or all of the operations of union, intersection and complement are provided as the tools for manipulation. A parallel, but equally powerful technique to the boolean operation is the set theoretical solid modelling procedure [4]. This exploits polynomial inequalities to define half spaces, and is ideal for the description of manufactured components that are simple elements joined, subtracted from or unioned to, one another. CSG (Computer Solid Geometry) has been exploited by Latham for the visualization of sculpture. Here primitives can be described in three-dimensions within the context of an evolutionary tree and by the utilization of boolean operators to add, subtract and union forms together until the desired form is created [5].

While the derivation of the surfaces of machined objects can be formulated from mathematical geometric solids, the face does not consist of clearly defined geometric objects and cannot be conveniently encapsulated within such a process. Although the complex compound curves can be described with boolean operators, the construction of the face using this technique would be far too cumbersome to model, maintain and control.

An alternative strategy is to delineate the face as a collection of curvilinear quadrilateral patches, where each patch can be completely specified by means of a mathematical formula. This area of research has received a broad range of attention [6]. The patch definition, in contrast to boolean operators, offers distinct advantages, primarily due to the simplicity of specification. However, despite the simple designation, patches require initiating at least four x,y,z triplets of a regular mesh, and the demand is identical to the explicit definition of polygons. Therefore, the modelling of the face has to begin with the input definition of at least the vertices of a polygonal representation.

Polygons are a fundamental modelling component in a majority of systems. The face can be treated in the same way, and once the model has been constructed, additional procedures can be applied, for example general or local free-from deformation [7].

3.2 Data Acquisition Techniques.

Whatever the application, the data points that represent the surface are partly determined by the input techniques employed. General raw data of complex compound surfaces can be captured by a number of different techniques, each with specific advantages for particular applications;

3.2.1 The Lofting of Cross-Sections.

As the definition indicates, cross sections can be stacked on top of one another and corresponding slices connected together. For a more comprehensive survey of contouring techniques see Sabin [8].

Lofting is a general technique involved in medical imaging, and the resultant data is usually considered as three-dimensional computer-generated images that surgeons use for diagnostic purposes [9]. Computer Axial Tomography (CAT), X-rays, Electro-Magnetic Resonance (EMR) and Single-Photon Emission Computed Tomography (SPECT) create scans that are slices of data at regular intervals (usually less than a mm apart) through the human body and reveal internal bounding surfaces, such as bones, and the softer organs of the human body.

The process then involves reconstructing a composite three-dimensional object from the data. Each slice is processed to isolate specific tissue types such as bone and softer organs. Then, by stacking the slices in the same order that they were scanned and linking successive slices together, a solid form is constructed.

Linking can be achieved by a number of techniques, such as connecting contours of consecutive slices with triangles, voxel modelling, or marching cubes, that prevent the excessive loss of data between slices [10]. The object can then be viewed in near real-time while being illuminated by light sources as a replica of the original form.

A less elaborate but analogous lofting technique was used by the present author in an early three-dimensional model of the Statue of Liberty (Figure 7) [11]. In this instance

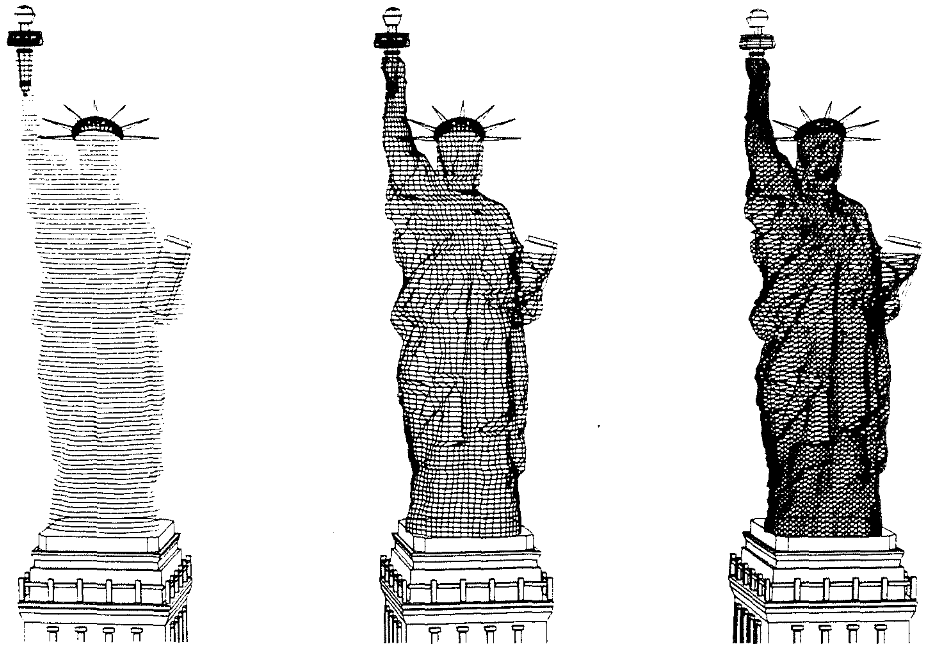


Figure 7(a) **Figure 7(b)** **Figure 7(c)**

A three-dimensional model of the statue of Liberty. Three techniqes were employed to visualize the surface, all the cross sections drawn (7(a)), non planar quadrilateral facets (7(b)) and the complete triangulated facet structure (7(c)).

a physical model was assembled, cast and set in a regular block of dyed plaster approximately 200mm by 200mm by 1000mm. Then regular 5mm slices were cut to reveal the cross-sectional detail. Once they were digitized, stacked in three-dimensions and lofted together, a facsimile of the original was generated in the computer.

The advantage of this approach was that it allowed data to be collected from the undercutting in the folds of the cloth and shapes that would be obscured from an optical system. However, it is evident from the images created, especially for the head and face, that key data points were lost. This was primarily due to the coarseness of the cut and could be resolved by maintaining finer divisions. While this would reveal more detail and result in a finer image fidelity, it is obvious that more computational processing is required to display the images. Therefore a compromise needs to be made between image fidelity and computational overheads.

3.2.2 Light Beam and Laser Techniques.

Both light beam and laser employ the same techniques in the acquisition of three-dimensional raw data [12]. Typically the human subject to be scanned sits in a chair, and the face is immobilized, the chair is then rotated by a stepper motor under microprocessor control to give synchronization between the subject´s position and the television image. A laser beam or light beam is shone onto the surface of the form creating a number of vertical traces. These traces are lines of intersection between the beam and the surface of the form. By knowing the rotation angle of the object, the laser angle and the distance between the camera and the center of rotation a relationship between the real world space and the screen co-ordinates can be determined.

Figure 8 of the author´s head, illustrates the facial form created by this process. Some thirty thousand data points were recorded using this automatic process which is far too many for a reasonable computer display of a synthetic face. Therefore the individual vertical line segments were pre-processed to simplify the data. First, the line segments were passed through a B-spline routine to smooth out any irregularities in the motion of the face during scanning (both the motion of the turntable and the movement of the subject). Once this was achieved, each segment was sliced horizontally according to the principles of CAT scanning to reveal sections of the subject that could then be lofted together. The advantage of this technique was that any

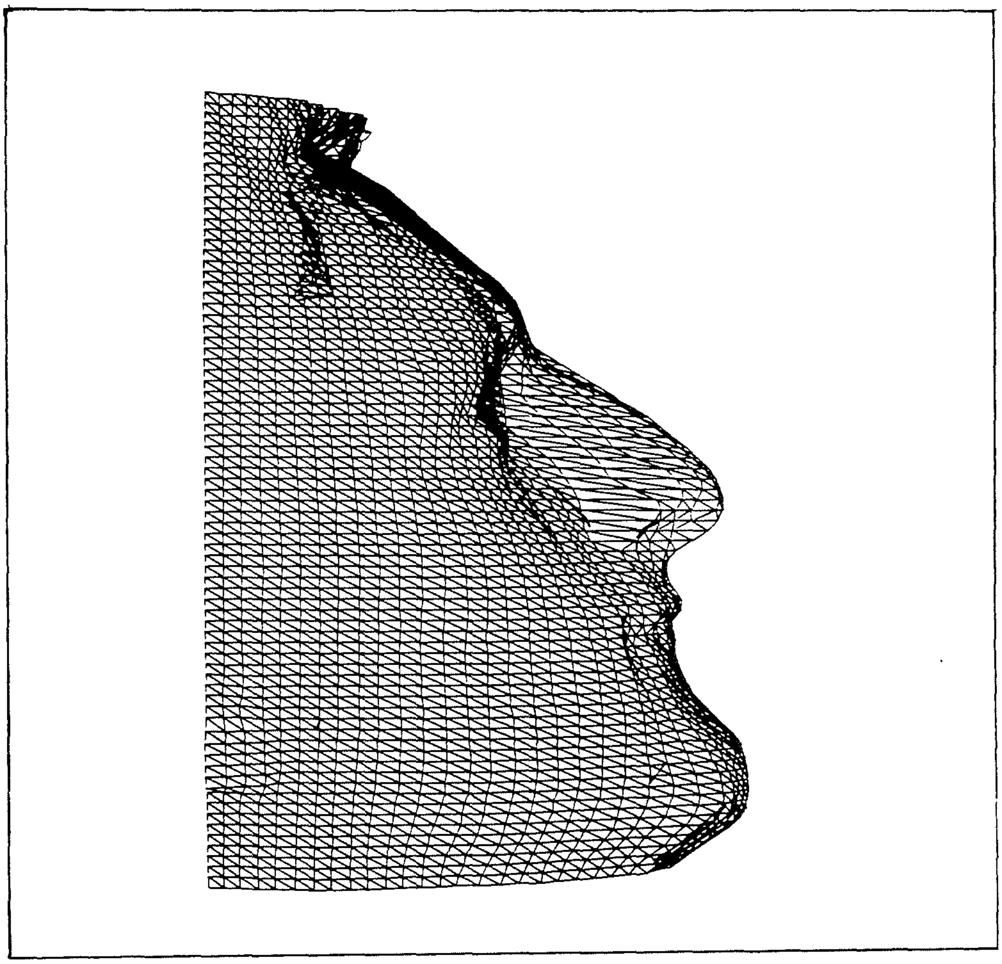


Figure 8 (a)

A faceted reconstruction of the three-dimensional laser scans of the author's face. A horizontal slice separation created cross-sections of the face which were then triangulated together.

Figure 8 (b)

A rendered image of the author's face from laser scans.

slice resolution could be specified (usually less that 5mm) and this consequently reduced the computational processing overhead.

While this produces very accurate static data representations, the data structure lacks obvious and specific manipulation. For example, to establish the jaw rotation would require the exact definition of vertices that make up the jaw and lip segment. This may not be perplexing for a particular slice resolution, but as soon as the detail is increased the group of vertices that constitute the jaw would have to be respecified. This is not a convenient data structure for the model of a face.

3.2.3 Mechanical Digitizers.

Mechanical digitizers come in a variety of forms, some are pens, others are mechanical arms. Mechanical arms are concerned with calculating the tool-tip position in three-dimensional space from the joint angle readings of potentiometers at the joints. The mathematics concerns homogeneous matrix transformations of concatenated rotation and translation matrices, that is, the inverse approach to that adopted in automatic robotic manipulators which establish the tool-tip at pre-specified points in three-dimensional space [13].

Once the mesh topology has been described onto the subject, the tool-tip is touched onto the surface and the point registered in the computer. Using this technique individual facets can be constructed in a desired order, thus surmounting the data base manipulation problem met within laser techniques.

The limitations of this technique for faces are three-fold. Firstly, most mechanical arms are limited to certain degrees of freedom (some are just two sliding joints and one revolute). This can be as limiting as the line of sight process found in laser techniques and is therefore no improvement on the optical processes already discussed. Secondly, the face is a flexible form, and if a variety of expression positions was required, the subject would have to hold the static poses for long periods while the digitization took place. Finally, the surface of the face is predominately soft and the tool-tip would have to be placed carefully onto the surface to prevent surface distortion during digitization.

3.2.4 Stereophotogrammetry.

The measurement of vertex positions on a three-dimensional surface is difficult without access to any specialized equipment, such as lasers or mechanical digitizers. The photographic process is the simplest and most fundamental process for the capture of three-dimensional data.

The two-dimensional photographic image can be viewed as a projection of the three dimensions onto a two-dimensional plane [14]. This mapping can be described mathematically using homogeneous coordinate techniques and a mapping transform matrix [15]. Described in this chapter is a technique used to input all the images in the research (except the two laser scans) with a simplified principle used by Bergeron and Parke [16]. This is achieved with standard photographic techniques, the digitizing tablet and the computer.

3.2.5 Discussion.

The techniques outlined above have specific advantages for particular applications. One of the fundamental aspects is whether a system uses an exterior optical or interior CAT scanning of the object. With an optical process the system can only capture surface data that is visible to the line of sight. Consequently, any details beneath the surface are lost, but more importantly, any under-cutting or shadowing will cause the truncating of the object. Additionally, two separate objects, such as the arm raised up beside the head, cause intractable problems such as the cast shadows masking the data on the opposing surfaces.

Slice scanning, such as the CAT or EMR scans, on the other hand, reveals the data lost by shadowing. But, because it can only slice at pre-specified intervals, it is indiscriminate and unsympathetic to the shape, and consequently vital detail can be lost. Also the bifurcation of two shapes from one has to be manipulated as a special case. Detail can be regained by slicing at even finer levels, but this can only be satisfactory if some form of image processing is applied after scanning, as there will undoubtedly be an enormous amount of similar data that becomes computationally expensive to process.

The photographic technique is an optical process that relies on two views that can simultaneously capture the data points. Therefore the object under view has to be convex, with little or no undercutting from the line of sight. The face is predominately a convex object with only a limited number of undercuts. It is a single body, with the exceptions of the eyes, mouth organs, ears and hair, and finally it can be treated (although not with strict accuracy) as a symmetrical object about the vertical meridian. The primary disadvantage concerning the photographic process is that unlike the laser and CAT scanning, it is not an automated process. It is necessary to describe the surface data points explicitly onto the surface of the model such that they can be used as accurate reference points for digitizing.

3.3 The Construction of Three-Dimensional Faces for this Present Research.

For the application of facial topographies in this present research a polygon description is used and stereophotogrammetry is the data acquisition technique employed. The reasons for this are two fold:

i. Firstly it was necessary for the features of the eyes, eyelids, tongue and teeth to be modelled as discrete structures. The techniques discussed above concern the scanning of data for the whole object with no facilities to identify and separate individual characteristics.

ii. Scanning techniques or lofting techniques provide a vast quantity of data much of which is redundant. A more optimal data structure can be created by specifying the salient nodes explicitly, thus reducing the overall data storage and subsequent manipulation times.

The first task is to establish a topological network for the face, by dividing the main areas of the face into regions. The objective is then to describe facets with either quadrilaterals or triangles. A useful feature of triangular description for facets is that three vertices will always lie in the same plane, whereas four vertices need not be co-planar.

Some basic guidelines can be established for this process:

i. Where the curvature is high, more facets are required for the description.This is especially noticeable around the area of the mouth and nose.

ii. By following general latitude/longitude lines and orbital features, such as the areas round the mouth and eyes, the problem of extraneous distortions in the facet structure can be avoided. Furthermore, it is important that polygon descriptions do not span the natural creasing of the facial form, as this results in flattening of the structure.

iii. The dissimilation of vertices with a line that traverses facets from the ear to the corner of the lips (roughly following the line of the mandible) describes the rotation of the jaw. Subsequently, the mandible rotation can be established from those nodes that fall below the dividing line.

iv. The face is a flexible surface, and a variety of crease lines appears on the face. By getting the subject to contort his/her face, clear indications of major facet divisions can be discerned. Polygons must be layered in such a way that allows the face to flex naturally and not distort the form when it is articulated.

v. Since the face is approximately symmetrical, it is possible to model only half the face and then mirror the polygons in a vertical plane of symmetry.

vi. Excellent results can be obtained with only a limited number of polygons. This is in contrast to lofting or laser triangulation where thousands of data points construct numerous polygons. Excluding the eyes and teeth, a respectable half face can be constructed from less than three hundred vertices and five hundred polygons, indeed it is possible to construct faces from as little as one hundred vertices and two hundred and fifty polygons. While more detail can be obtained from extra points, such a process increases the data acquisition problem, and more importantly, the computational overhead to process the data.

Once the facets have been drawn onto the subject, two cameras with long focal length lenses (to reduce perspective distortion) are set up ninety degrees to each other at equidistance from the subject who sits facing one camera while the other is positioned to the side (Figure 9). Because the face is a variable surface over time, simultaneous photographs are required if an interpolation principle is to be used in animation.

The photographs can then be digitized to form two separate two-dimensional views of the same object. Knowing that the cameras are set up ninety degrees from each another allows

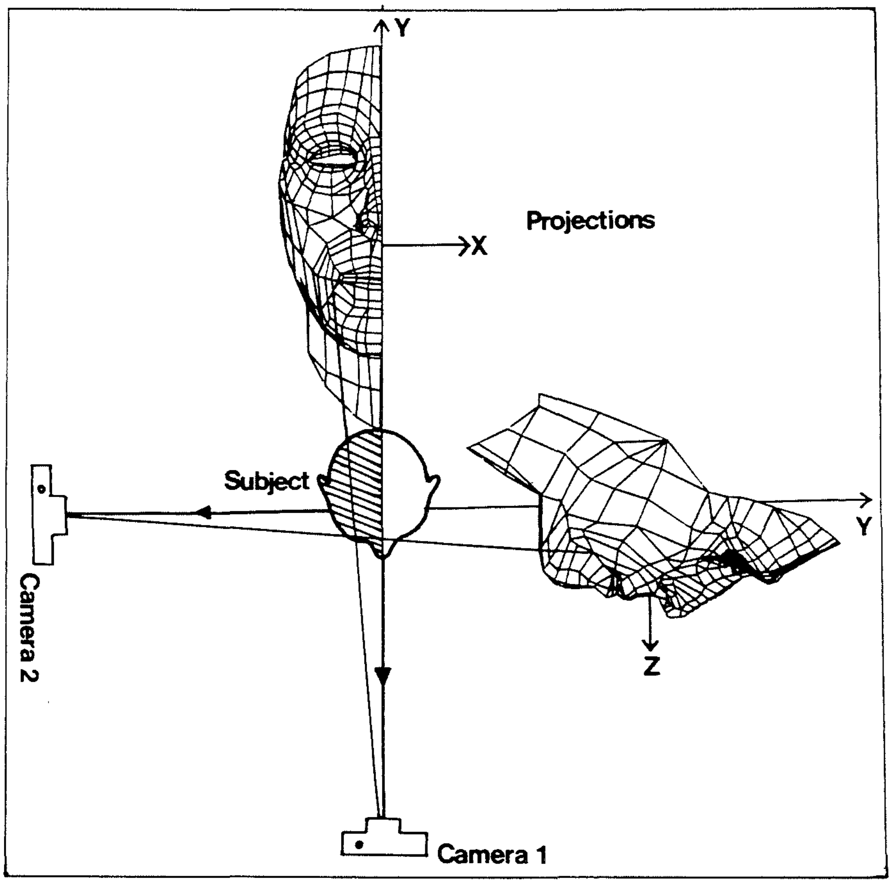


Figure 9

Camera layout for stereophotogrammetry. Two synchronized cameras ninety degrees to each other can capture one x,y view and one x,z view.

the substitution of the x component from the lateral view into the frontal (x,y) view to correspond to the z component. This allows the largest degree of accuracy to be achieved. However, at such a large angle, some data may be obscured from the line of sight. A modification of the substitution code, to take into account the angle at which the cameras are set, allows different views to be used. The major draw back concerns errors which occur when the angle between the cameras becomes small. To avoid some errors the images created in this present research were all synthesized from two cameras ninety degrees to one another.

To create a representation of polygons, pointers are used to reference the vertex list. With this representation, each vertex in the polygon mesh is stored just once. For example, in the vertex list N = ((x1, y1, z1) .... (xn, yn, zn)), the polygon is defined as a list of pointers P = (1,5,9,11). This is a standard modelling technique used in many computer graphics applications [17].

3.3.1 The Upper Face Construction of the Eyes, Eyelids and Brows.

Once the facial topology has been constructed, the positions of the eyelids and eyes can be established from the data structure, one of the primary features being the construction of the eyes, eyelids and brow which serve to attract the observer´s attention [18].

The eye is fundamentally a ball, consequently it can be constructed simply from a profile of swept revolution with the back open as this will never be seen. The pupil dilation can then be achieved by varying the profile facet count for the pupil.

Eyeball focusing is critical if the eyes are not to stare aimlessly at objects. In effect only two rotations are required to orientate and focus the eyes. Once again general descriptions as to the limits of rotations are given by Drefuss [19]. The upper limit of the visual field being approximately 50-55 degrees and the lower limit to the visual field 70-80 degrees. For the horizontal axis binocular vision is limited to approximately 60 to 65 degrees in both directions with a further limit to monocular vision of approximately 90 to 95 degrees. This gives an overall visual field of nearly 190 degrees in the horizontal plane and almost 125 degrees in the vertical field. These limits can be set to avoid unreal situations when the eyes are over-articulated.

For the focusing of the eyes, Figure 10 illustrates the calculations required in two dimensions, and in the complete model the same calculation is extended to the vertical angles. The results allow the eyes to focus on objects and this enhances their believability, especially with objects close to the face when the eyes clearly do not have parallel activity.

The highlight on the eye plays an important role in creating the illusion of reality. Perfectly acceptable results can be obtained from the rendering algorithms. However, more explicit control can be achieved by calculating the light position in relation to the eyeball, then extracting the corresponding facet and shading this white. This technique can be used when the eye is orientated away from the light source.

The construction of the eyelids is accomplished in conjunction with the eyeball profile and the use of curves. The components of the eyelids are the static bottom lid and the slightly larger upper lid that concertinas into the upper eye-oris. The variables in constructing the eyelids are:

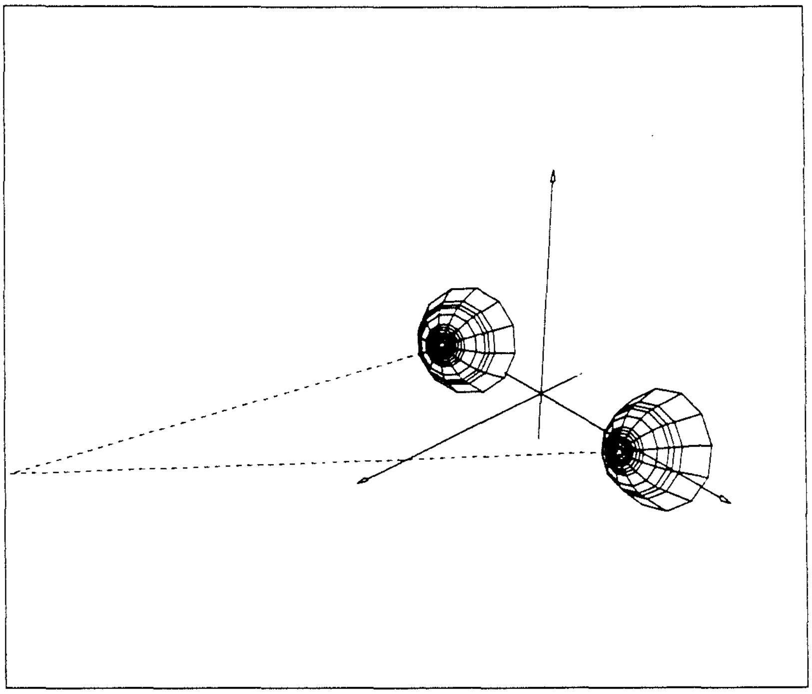


Figure 10 (a)

A pair of eyes focusing in three dimensions.

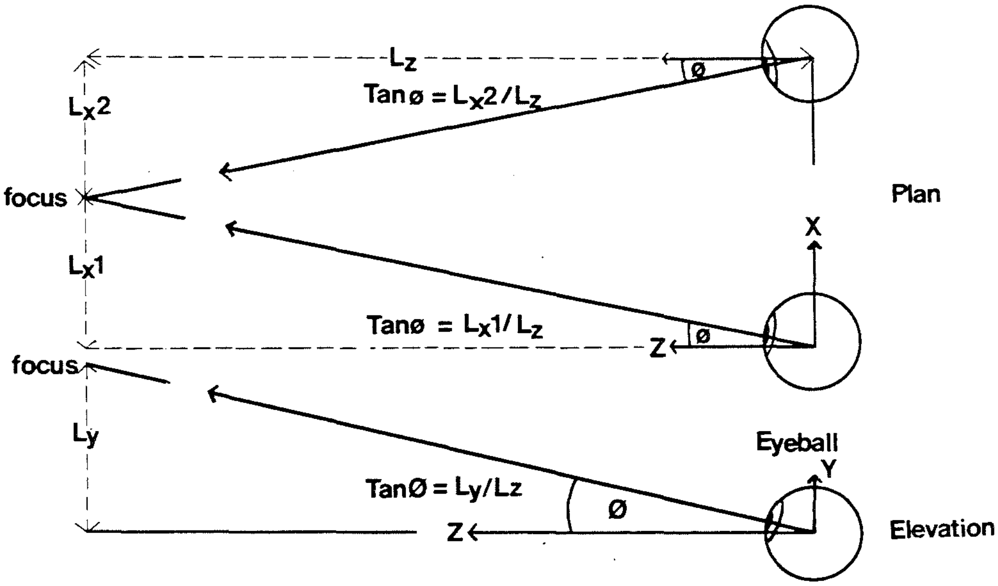


Figure 10 (b)

The plan and elevation calculations allowing the eyes to focus on positions in three dimensions.

i. Positioning of the lids.
ii. Opening values for the upper and lower lids.

The eyelid profile does not curve exactly to the eyeball´s spherical shape, but tapers at each end blending into the corners of the eyes. Consequently, it is necessary to construct the lids from free-form curves. In the computer model the eyelids are constructed from four basic profiles and an additional profile that maintains the curvature of the top lid by remaining in the mid position between the extremes. Video sequence V.1 animates the characteristics.

The eyebrows are keys to larger expression characteristics [20]. The variety of expressive shape is not limited to the up and down motion, but in overall shape along their length and in the relationship of one brow to the other. Fundamentally, the eyebrows are constructed from the existing data base. They can then be extracted and, to avoid the brows sliding over the face, the underlying topology can also be displaced. An important feature is the area between the brow and eyelid. This contracts and expands, darkening and lightening to give the expressive characteristic. It was assumed in the first practical implementation, that moving the brows vertically would suffice. However, this proved not to be the case due to the complexity and subtlety involved in brow motion. It was discovered that a variety of parameters is required to maintain the correct expressive message:

i. Separation
ii. Curvature
iii. Raising
iv. Longitudinal creasing.

Even with the topology moving, it still looked as if the brows were sliding over the facial form. To solve this problem, creasing of the forehead needs to be introduced to emphasise the expression. The creases must be apparent when the model is static, and then deepen as the brows are raised.

3.3.2 The Lower Face Construction of the Lips and Teeth.

The lower face consists of the jaw, lips, teeth and tongue. Research into speech therapy indicates that the teeth are paramount in the perception of visible articulatory gestures [21]. Certainly, during speech, the white of the teeth is revealed in all ´s´ sounds and the tongue in ´l´ sounds. For the model in this present research, only the teeth are considered. Because the teeth are only partially viewed, only a fundamental model was constructed. However more detailed information could result in a more accurate model. The general curvature of the mandible was followed thereby creating a Bezier curve from which the upper and lower teeth were modelled. The upper teeth remained static, while the lower teeth followed the motions of the mandible.

3.4 Geometric Transformations and Flexible Moulding.

The fundamental form can be adequately constructed from the preceding procedures. However, it is possible to distort the face geometrically from any arbitrary starting shape to another position. This is a powerful form of modelling that can be considered to be a flexible moulding approach. Flexible moulding is a process that relies upon the manipulation of already existing structures with algorithms that predominately bend, deform and shape the structure as if it were plastic.

Geometric transformations can be applied in both two and three dimensions, and it is possible to consider a living form (such as a growing organism) as a collection of two-, or three-dimensional x,y,z coordinate triplets in either polar or cartesian space [22]. Geometric transformations can then describe how coordinates can be systematically altered by mathematical formulas to emulate growth or some other natural process. Transformations can be classified by identifying common characteristics, for example the transformation:

$$x' = \cos(\text{theta}) * x + \sin(\text{theta}) * y$$
$$y' = -\sin(\text{theta}) * x + \cos(\text{theta}) * y$$

where:

theta is the angle of rotation,
x,y the position before rotation,
x',y' the position after rotation.

is a rotational transformation, and the angles and distances between vertices are preserved.

$$y' = y$$
$$x' = x + (y * \tan(\text{theta}))$$

where:

theta is the angle of shear,
x,y the position before shear,
x',´y' the position after shear.

This is an affine transformation that maintains parallel lines after the distortion.

One of the first to recognize that geometric transformations could be applied to the morphological deviation in the head and face was Albert Durer in 1613 [23]. Dürer´s work, and that of other renaissance artists such as Leonardo Da Vinci, was founded upon observation. It was not until 1917 that Thompson, in his classic work On Growth and Form applied geometric and mathematical transformations to the head [24]. His primary evidence for the progression of evolution and the development from infancy to adulthood was his ability to represent apparently complex changes in morphology as the deformation of a grid positioned over an evolving or growing organism.

Thompson gave no physical or biological explanations for the phenomena he modelled. In fact his approach, in most instances, was subjective: he merely noted how the transformations operated and did not analyse their effects. While this may appear to be an ad hoc approach, it is worth noting that a reliance on perception seems to be present in almost every attempt that has ever been made to study the morphology of either living or inert forms [25].

An attempt to measure, analyse and predict craniofacial growth in humans was carried out by Walker and Kowalski in 1971 [26]. Their technique involved the digitization of two-dimensional images of the bones of the cranium, known as cephalograms. Once the images were aligned and recorded in two dimensions in the computer, statistical analysis was carried out and mathematical progressions for specified vertices were predicted. The observed growth trajectories associated with the digitized cranial nodes were called vectorgrams, and the variables were such that;

$$Pij(Tk) = [Xij(Tk), Yij(Tk)]$$

where:

vertex position $j = 1 \ldots. N$
individual $i = 1 \ldots. N$
time $k = 1, \ldots ki; < \ldots < Tki$

Consequently the vectorgram for the ith individual for vertex j was:

$$Vij = [Pij(t1), \ldots Pij(Tki)]$$

Subsequently a matrix for an individual can predict how the person will grow based upon a summation of previous data collected from that person. While the results appear promising, substantial amounts of data about the craniofacial features and positions are required before the vectorgram can be constructed and manipulated.

The investigations of Todd et al further supported Thompson's observations, by the selection of the 'best-fit' to describe the prospective transformation [27]. Further detailed geometric analyses by Todd et al carried out on the perceived distortions validated the results. They discovered that most effective geometric deformations are the conformal transformations. Here the distortions preserve the angular coordinates of every point in a polar coordinate system:

$$\text{theta}' = \text{theta}$$

$$R' = R(1 - k \cos(\text{theta}))$$

where:

theta is angle in polar co-ordinates,

R is the radial distance,

k is a strain constant,

O',R' the returned polar co-ordinates.

this deformation is called the cardioid-strain transformation, so named because it transforms a circle into a heart shape called a cardioid and is illustrated in Figure 11. The results of the Todd et.al investigations indicated that the cardioid-strain transformations were the closest approximation to the actual growth sequence of the human head from infancy to adulthood [28]. They also developed additional transformations to orientate the face to fundamental axes, such as the Frankfurt horizontal, from which the facial angle can be determined and manipulated by shearing transformations.

For the transformation of evolution the cardioid was re-established as:

theta' = theta
R1 = R (1 + k ¶theta¶)

where:

theta is angle in polar co-ordinates,
R is the radial distance,
k is a strain constant,
theta', R' the returned polar co-ordinates.

Figure 12 illustrates the transformation from a Primative man to a futuristic being. These techniques of geometric

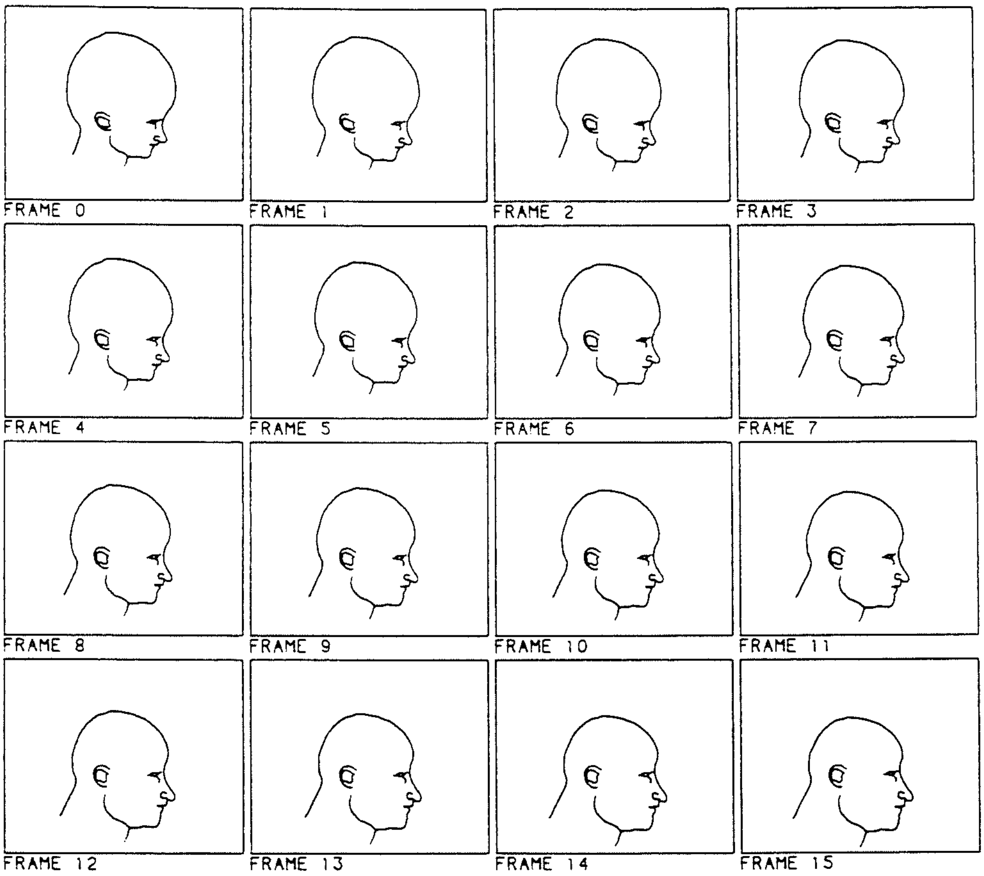


Figure 11

Cardioid strain transformation from infancy to adulthood. Strain value from - 0.22 to 0.33.

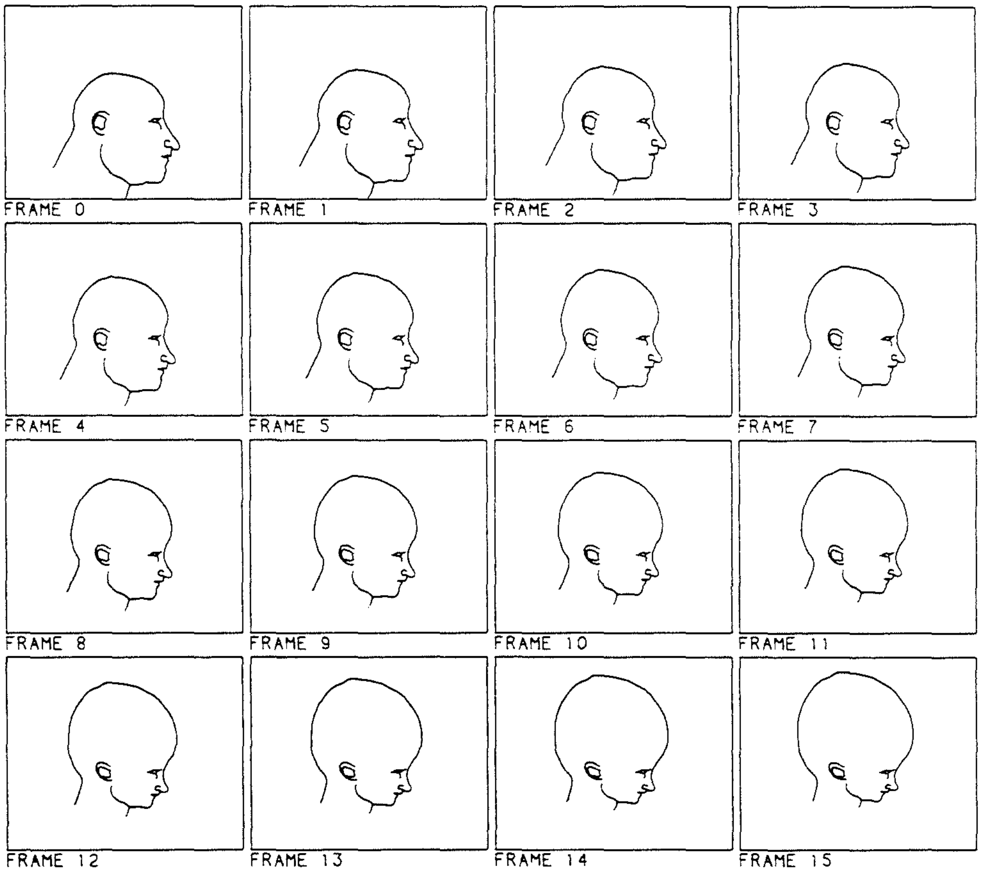


Figure 12

Cardioid strain transformation from neanderthal man to a futuristic being. Strain 0.33 to - 0.31.

transformation of the human cranium, have important consequences for the user once the face has been constructed in three-dimensions, as they provide a technique for ageing the underlying structure of the human face.

A different system of geometric transformation was utilized for the development for the interactive generation of two-dimensional facial images on a CRT by Gillenson [29]. The system was designed as a computer photo identi-kit, and rather than using biological cranial features, individual facial features could be heuristically specified and modelled. The system begins by displaying an average face and a pair of large-scale scaling routines age the face. Then further gross changes are made on individual features through a hierarchical manipulation scheme. A total of seventeen features (seven of which are paired) are stored in the system and can be geometrically distorted by scaling, rotation and translation such that any reasonable variation can be achieved. The whole system hinges on the user´s faculty of recall recognition and the system´s selection of feature manipulation.

The early investigations by Parke in three-dimensions exploited some conformable features to change the facial structure [30]. The process of modification used both interpolation between faces of identical topologies and scaling parameters. These geometric deformation parameters determined the aspect ratio of the face, the shape of the eyelids, the width of the nose, the scaling of the area between the mouth and chin and the width of the jaw. For example, the pronounced forehead, could be established by interpolation from another facial mesh and then modified globally the aspect ratio of the face.

What is apparent with mathematical geometric distortions is that it becomes difficult to encapsulate a broad range of facial types, and specific deformations are difficult to delineate. In three-dimensions it is possible to overcome these deficiencies by "hard-wiring" the system such that certain characteristics will only perform the appropriate tasks, as with the work of Parke [31]. This obviously limits the system, as every face would require an explicit definition of the topology.

It is evident from the deficiencies in existing methods that an alternative approach is required to increase the breath and generality of three-dimensional geometric transformations. It is essential that they function on non-specific topologies to remove the hard-wiring obstacle, and to facilitate the unconstrained construction of the model.

3.4.1 Solid Deformation.

When the facet description of the face has been generated, further manipulation of the form can be induced by algorithmic deformations [32]. This is considered to be a flexible moulding process as the model must be realized before it can be manipulated.

Dürer in his pursuit of facial proportion, suggested that the nose was the critical factor in determining the facial shape. With his drawings he divides the face into horizontal sevenths at key positions:

i. Top of the head
ii. Base of the hair line
iii. Top of the eye
iv. Base of the eye
v. Top of nose flange
vi. Base of nose
vii. Midline of lip
viii. Base of the lip
ix. Base of the chin

These positions are then manipulated and a new facial form drawn. Then additional characteristics other than the vertical separation are also utilized such as the facial plane, forehead protrusion or chin regression to create a new face. Due to the variety of potential facial forms created by this manipulation technique, and investigation was briefly carried out into the potential of algorithmic manipulation.

An existing three-dimensional model was taken and the same horizontal divisions were delineated from the lateral view only to create planes from which it was possible to determine which side a vertex was found. By relocating the plane separation a linear relationship can be established for any arbitrary node for its subsequent distortion;

$$y' = y + ((y - y1)/(y2 - y1)) * (ny1 - ny2)$$

where:

y is the old y position,

y' is the new y position,

$y1$ is the start first plane location,

$y2$ is the start second plane separation,

$ny1$ is the finish first plane separation,

$ny2$ is the finish second plane separation.

Figure 13 illustrates the geometric distortions that resulted in the manipulation of the sections. Extensions to this fundamental principle can be applied to include a horizontal displacement, twisting, shearing, scaling and other basic procedures of this nature.

This type of flexible molding does not consider the boundary conditions between sections that can cause irregularities. What is further required is a volumetric approach that deforms under the influence of other key boundaries or control points.

The process of deformation illustrated in Figures 14(a), 14(b), 14(c) and 14(d) can be considered on a volumetric basis such that the form is distorted from a rest position with a number of control points. This is uniquely different from the explicit control of the surface form. Free-form deformation has been used for modelling and animation, but both areas have the same problem which concerns specific manipulation and how to control the guiding vertices.

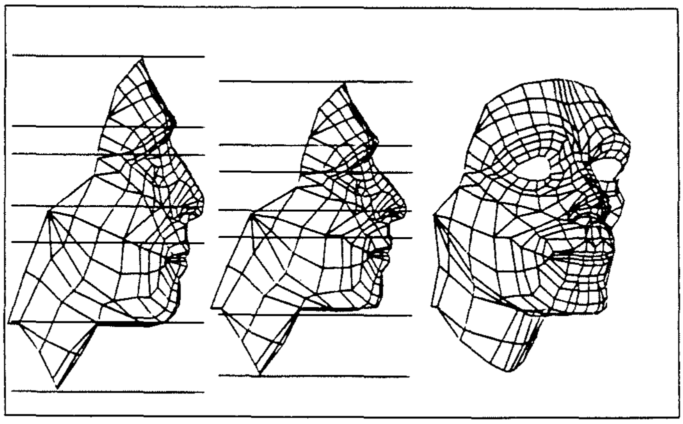


Figure 13 (a) Compressing the sections together.

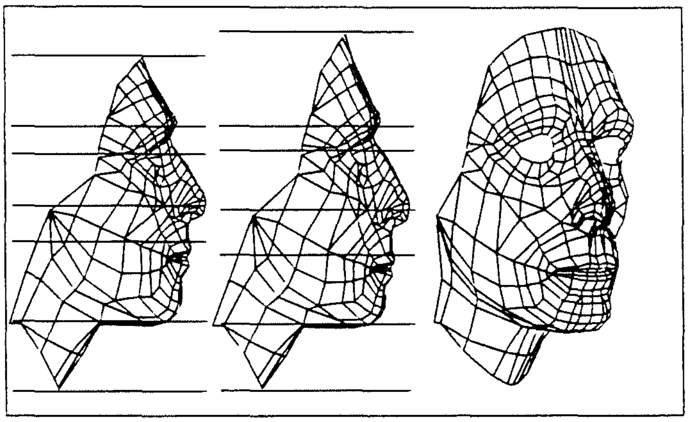


Figure 13 (b) Extending the sections.

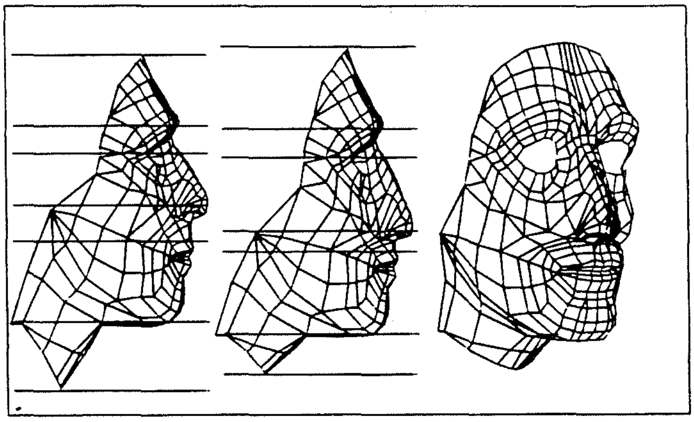


Figure 13 (c) Extending the nose zone.

Three-dimensional geometric section manipulations after Dürer. This illustrates the face before manipulation (left), the deformation (centre) and the three-dimensional result (right).

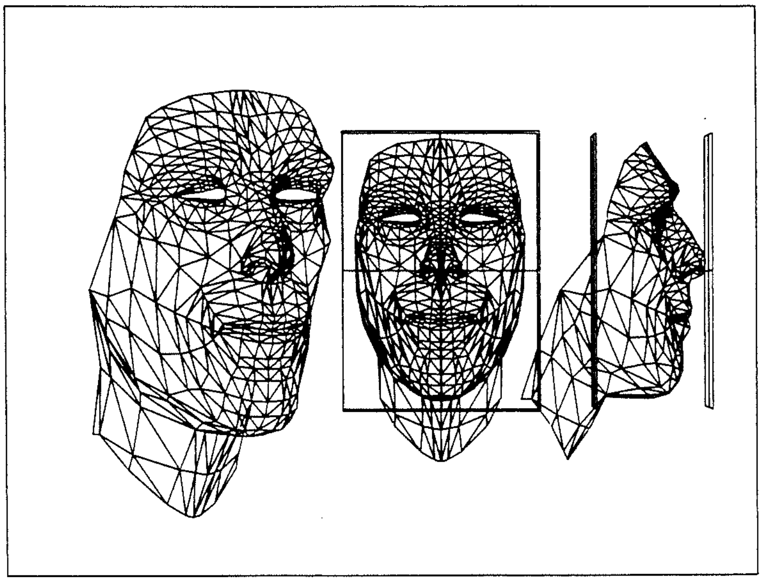


Figure 14 (a)

Free-form deformation raising the centre section of the face upward towards the brow.

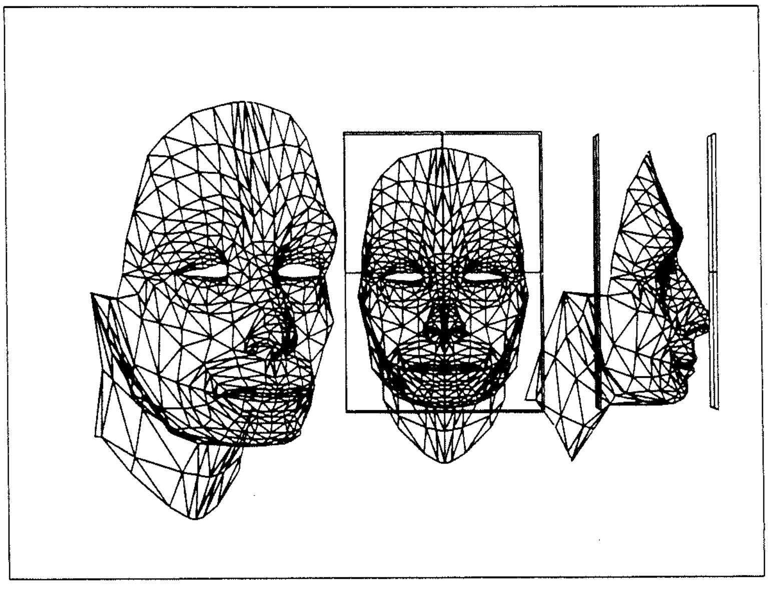


Figure 14 (b)

Free-form deformation lowering the centre section of the face downward toward the chin.

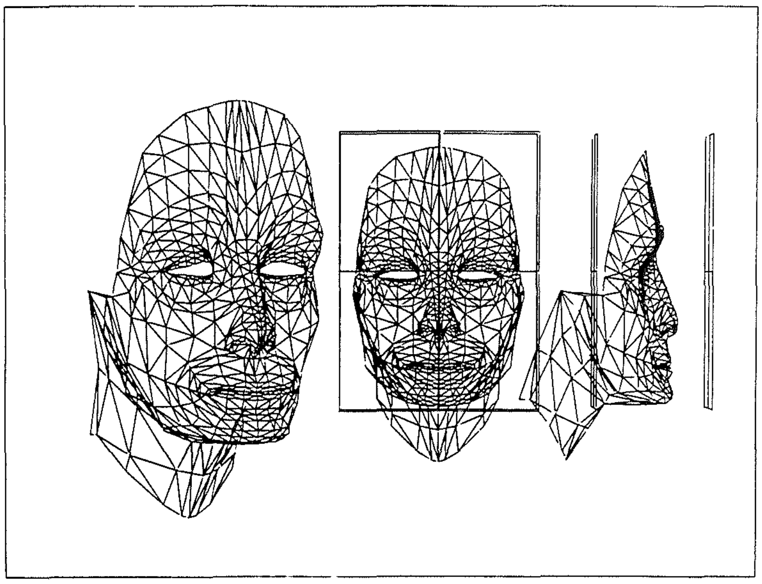


Figure 14 (c)

Free-form deformation depressing the centre of the face inward.

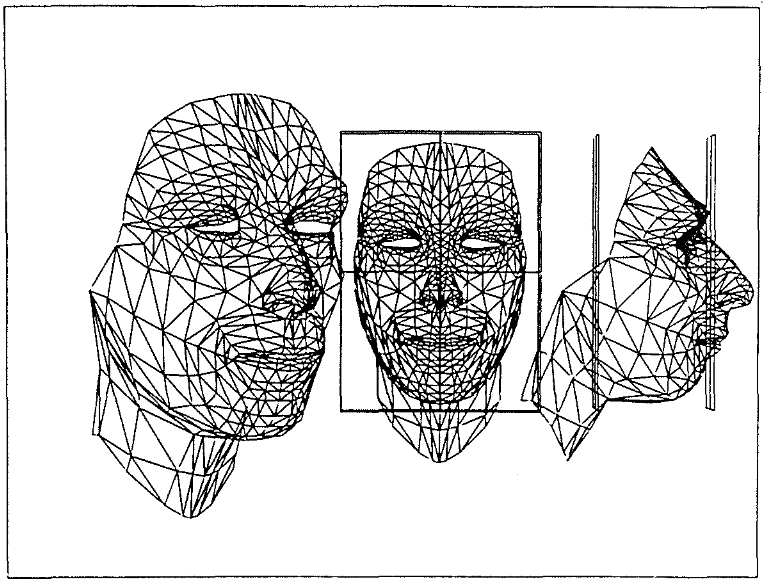


Figure 14 (d)

Free-form deformation pulling the centre of the face outward.

3.4.2 Parametric Patch Model Definition.

An alternative strategy is to delineate the object as a collection of curvilinear quadrilateral patches, where each patch can be completely specified by means of a mathematical formulae. This area of research has received a broad range of attention [33]. A typical example is a Bezier patch that consists of sixteen control points arranged in a four by four grid [34]. The resulting surface is regulated by control points, and displacing the guiding vertices distorts the underlying surface. This provides a particularly valuable device for the modelling of non geometric surfaces, and the strengths and limitations of the technique is symbolized in the ubiquitous teapot constructed from twenty-eight parametric patches [35]. The predominant draw-back is the location description for the guiding vertices that are established, either by adapting an object from an arbitrary starting shape using interactive techniques, or by fitting a surface to a given set of data points measured from a physical model. The latter technique is more typical, and is exploited more extensively in the construction of natural free-forms such as the feet [36].

Once the basic polygon description of the face has been established, further detail can be established by using the vertices of the polygon descriptions as the control points for patches. A number of techniques have been described for the formation of patches, and in this present research Bezier patches were employed [37]. While any level of detail can be obtained from patch descriptions, they are eventually subdivided into smaller facets for rendering. Consequently, it is important to determine whether subdivision is necessary for the image fidelity, as the results can be only marginally superior to the explicit polygon descriptions. Patches are therefore an additional technique used in this research (see 5.10) and are not necessarily part of the basic strategy for the modelling of the facial topology.

3.5 Conclusions.

The different techniques - lofting, optical, mechanical digitizing and stereophotogrammetry - partly determine the representation of raw data captured. This is critical factor because the nose, eyes and a mouth are salient features that have to be carefully constructed.

The lofting of cross sections is a crudest technique for modelling the face. Not only does it produce large quantities of data, but it is unsympathetic to the form it is capturing. Critical nodes of the face can be completely missed as a slice level is predetermined before the operation begins.

It has been shown that the optical process of laser scanning is a very powerful technique for capturing face data. Many thousands of points can be captured in a few seconds, but much of the data is redundant (in the order of a factor of seven) as the scans are not capable of selective processing. The task therefore, once the data has been recorded, is the extraction of the key nodes from the very large quantities of data.

Both mechanical digitizing and stereophotogrammetry require the definition of the polygonal topology to be determined before digitizing begins. While this is a labour intensive process, it does allow the explicit definition of only the polygons required. Consequently, stereophotogrammetry was the techniques employed by this present research and the features of the eyes, eyelids, eyebrows, tongue and teeth were also created as a polygonal description.

Once the face has been constructed, soft moulding transformations provide powerful manipulations to modify the basic shape. It has been demonstrated that geometric transformation can be applied in two dimensions that mimic the growth and evolving processes. General deformations in three-dimensions can be described, but these are not a complete solution capable of establishing a wide range of different facial types. The present research indicates that in three-dimensions the facial features do not possess the same topological characteristics. Consequently, general geometric transformations can not encapsulate all facial forms from a single generic mould.

3.6 References.

[1] P Duncan, ´Anatomical Definition and Modelling´, Engineering in Medicine, Vol. 15, No. 3, 1986, pp. 109-116.
Marilyn Lord, ´Curve and Surface Representation by Iterative B-Spline Fit to a Data Point Set´, Engineering in Medicine, Vol. 16, No. 1, 1987, pp. 29-35.

[2] P Bezier, ´Mathematical and Practical Possibilities of Unisurf´, Computer Aided Geometric Design, Academic Press, New York, 1974, pp. 127-151.
Adrian Bowyer, ´Symbolic Ray Tracing´, Computer Animation Computer Graphics Online Publications, 1987, pp. 127-133.

[3] I Braid, ´Six Systems for Shape Design and Representation´, University of Cambridge CAD Group, Document No. 87, 1975.

[4] Bowyer, ´Symbolic Ray Tracing´, p. 128.

[5] William Latham, ´Visualization of Sculpture´, Proceedings BCS Displays Group, December, 1987, pp. 39-45

[6] I D Faux and M J Pratt, Computational Geometry for Design and Manufacture, Ellis Horwood, New York, 1979.

[7] Thomas W Sederberg and Scott R Parry, ´Free-From Deformation of Solid Geometric Models´, SIGGRAPH Proceedings, Vol. 20, No. 4, 1986, pp. 151-160.

[8] Malcolm Sabin, ´A Survey of Contouring Methods´, Computer Graphics Forum, Vol. 5, 1986, pp. 325-340.

[9] Duncan, ´Anatomical Definition´, pp. 109-116.

[10] Henry Fuchs, ´Optimal Surface Reconstruction from Planar Contours´, Communications of the ACM 20, 1977, pp. 693-702.
S Bright, ´Shading of Solid Voxel Models´, Proceedings Eurographics UK, 1986.
William E Lorensen and Harvey E Cline, ´Marching Cubes: A High Resolution 3D Surface Construction Algorithm´, Proceedings ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 163-169.

[11] John A Vince, Computer Graphics For Designers, Pinter, London, 1985, p. 132.

[12] F Schmitt, H Maitre, A Clainchard and J Lopez-Krane, ´Acquisition and Representation of Object Surface Data´, Biostereometrics SPIE, Vol. 602, 1985, pp. 42-49.
J P Moss, A D Linney, S R Gindrod, S R Arridge and J S Clifton, ´Three-Dimensional Visualization of the Face and Skull using Computerized Tomography and Laser Scanning Techniques´, European Journal of Orthodontics, Vol. 9, 1987, pp. 247-253.
David Addleman and Lloyd Addleman, ´Rapid 3D Digitizing, Computer Graphics World, November 1985, pp. 41-44.
Cyberware Laboratory Inc., Computerized Three-Dimensional Reproductions, Sales Pamphlet, 2062 Sunset Drive, Pacific Grove, CA., 93950, 1986.
Studec, 3D Videolaser, Sales Pamphlet, Licence ENST, Paris Cedex 11, 1987.
Simon Arridge, James P Moss, Alfred D Linney, and David R James, ´Three-Dimensional Digitization of the Face and Skull´, Journal of max.fac. Surgery, Vol. 13, 1985, pp. 136-143.

[13] Haruhiko Asada and Jean-Jacques E Slotine, Robot Analysis and Control, John Wiley, New York, 1986.

[14] James D Foley and Andries Van Dam, Fundamentals of Interactive Computer Graphics, Addison-Wesley, Reading, Massachusetts,1981, pp. 269-279.

[15] Fred I Parke, ´Measuring Three-Dimensional Surfaces with a Two-Dimensional Data Tablet´, Computer and Graphics, Vol. 1, 1975, pp. 5-7.

[16] Phillipe Bergeron, ´Techniques for Animating Characters´, Advanced Computer Animation Course Notes, ACM SIGGRAPH, Vol. 10, 1985, pp. 61-80.

[17] Foley, Fundamentals, p. 508.

[18] David Noton and Lawrence Stark, ´Eye Movements and Visual Perception´, Scientific American, Vol. 224, No.6, June 1971, pp. 114-121.

[19] Henry Drefuss, The Measure of Man Human Factors in Design, Second Edition, Whitney, New York, 1967.

[20] Paul Ekman, ´About Brows: Emotion and Conversational Signals in Human Ethology´,Journal of Phonetics, No. 11, 1983, pp. 169-202.

[21] Michael Brooke, ´Analysis, Synthesis and Perception of Visible Articulatory Movements´, Journal of Phonetics, Vol. 11, 1982, pp. 63-76.

[22] Sederberg, ´Free-From Deformation´, pp. 151-160. Keith Waters, ´Expressive Three-Dimensional Facial Animation´, Proceedings Computer Graphics Online, 1986, pp. 188-210.

[23] L Strauss, The Human Figure by Albert Durer: The Complete Desden Sketchbook, Dover Publications, New York, 1972, p. 234.

[24] D´Arcy Wentworth Thompson, On Growth and Form, Abridged Version, Cambridge University Press, London, 1961.

[25] James T Todd, Leonard S Mark, Robert E shaw and John B Pittenger, ´The Perception of Human Growth´, Scientific American, Vol. 242, February 1982, p. 106.

[26] Geoffrey Walker and Charles Kowalski, ´A Two-Dimensional coordinate Model for the Qualification, Description, Analysis, Prediction and Simulation of Craniofacial Growth´, Growth, Vol. 35, 1971, pp. 191-211.

[27] Todd et al, ´The Perception´, p.107.

[28] Ibid.

[29] Mark Gillenson, ´A Heuristic Strategy for Developing Human Facial Images on a CRT´, Pattern Recognition, Vol. 7, 1975, pp. 197-196.

[30] Fred I Parke, ´A Parametric Model for the Human Face´, Technical Report UTECH-CSc-72-120, Dept. of Computer Science, University of Utah, 1974.

[31] Ibid, pp. 19-34.

[32] Sederberg, ´Free-From Deformation´, pp. 151-160.

[33] Faux, Computational Geometry, pp. 126-151.

[34] William M Newman and Robert F Sproull, Principles of Interactive Computer Graphics, McGraw-Hill, London, 1979, pp. 315-320.
Foley, Fundamentals, pp. 529-530.

[35] Frank Crow, ´The Origins of the Teapot´, IEEE Computer Graphics and Applications, 1987, pp. 8-19.

[36] Lord, ´Curve and Surface Representation´, pp. 29-35.

[37] P Bezier, ´Mathematical´, pp. 127-151.

CHAPTER FOUR

Chapter Four

The Muscle Model Process

4.1 Introduction.

Breaking down muscle activity into individual modules known as Action Units (AUs) is a valuable tool for computer synthesis [1]. These discrete units are used as fundamental building blocks for the development of a parameterized facial muscle process.

Facial topologies have been shown to vary enormously between individuals, and no one topological network can be satisfactorily employed to encapsulate all facial types for computer synthesis. Consequently, existing methods of facial parameterization have the inherent problem of hard-wiring the performable actions.

The development of a muscle process that is controllable by a limited number of parameters and is non-specific to facial topology allows a richer vocabulary and a more general approach to the modelling of the primary facial expressions.

4.2 Factors Determining the Modelling of Muscles and Skin.

The description of the physical and motion attributes of bone, muscle and skin of the face, are instrumental in the development of the muscle model process. Only by building upon an understanding of the motivators of the musculature action can any clear definition of parameter sets be built for the synthetic face. The following fundamental motion attributes can be extracted:

i. The muscles of the face can be grouped according to the orientation of the individual fibres of the muscle. Three types of muscle can be discerned as the primary motion muscles;
   (i) Linear/parallel muscles pull in a linear direction.
   (ii) Elliptical/circular sphincter type muscle that squeezes.
   (iii) Sheet muscle behaves as a series of linear muscles spread over an area.

ii. With one exception the muscles are attached to the skull at one end, and embedded into the soft tissue of the skin at the other. The exception is the orbicularis oris which has the one end embedded in the group of muscles converging at the modiolus at the corner of the mouth.

iii. When the skin articulates it acts like a rubber sheet, and deforms over and around the underlying structures. With the progression of age the epidermis and dermis lose their elasticity resulting in the deepening of flexure lines that are either folds in the dermis associated with habitual joint movement, or lines of attachment to the underlying deep facia.

iv. Defining the surface skin as a mesh determines that each node has a finite degree of mobility (DOM). The primary factors in determining the nodal mobility are:

(i) The tensile strength of the muscle of skin.
(ii) The proximity to the muscle node of attachment.
(iii) The depth of tissue at the node and the proximity to the bone.
(iv) The elastic bounds of the relaxed tissue, and the interaction of other muscles.

v. With the linear muscles only a proportion of the force is effective along the line of contraction, especially as the fibres become oblique in relation to the node of attachment. The displacement can be roughly approximated as the length of the muscle fibre x the cosine of the angle of the muscle fibre to the tendon or surface tissue [2].

vi. Most human muscle fibres obey the all-or-none law, which states that a stimulus produces either a maximal response for that condition, or no response at all. Further stimulation of a single motor nerve fibre of an intact muscle invokes a response in one motor unit, and all the muscle fibres of that motor unit respond [3]. The motor unit therefore also obeys the all-or-none law as well as the muscle fibre. A similar principle of all-or-none has been exploited in a finite state approach to the synthesis of bio-engineering control [4]. The system used a cybernetic actuator that has two binary inputs with one continuous output like a logical gate array to initiate the functioning of a mechanical limb. Gradations of contraction in a whole human muscle result from altering the number of active motor nerve fibres, and hence the number of motor units activated. This is the same scenario as that for the facial muscle.

vii. There are two types of musculature contraction, isotonic where the muscle shortens under tension, and isometric, where a considerable degree of tension may be produced but no shortening is produced [5]. Facial muscle is isotonic in nature and is infinitely variable.

viii. Finally, the skull is immobile with the exception of the mandibular joint, and has to be treated as a special case.

4.3 The Computer Model of Muscles for the Face.

With the three muscle types, it is evident that the three-dimensional structure is endowed with viscous, elastic and other biomechanical properties that result in the displacement of the skin. The simulation of such interactions would be formidable, and is not the objective of the present research. What is required is not the exact simulation of neurons muscles and joints, but a model with a few dynamic parameters that mimic the primary biomechanical characteristics. What follows is the description of the modelling of the three primary types of muscle: linear, sphincter and sheet. Video sequences and diagrams illustrate the activity of the various synthetic muscle contractions.

4.3.1 The Linear Muscle Vector.

Muscles, like vectors, can be described with direction and magnitude, both in two and three dimensions. The direction is toward the point of attachment on the bone, and the magnitude is the contraction of the muscle dependant upon the muscle spring constant and tension. The linear, or parallel muscle, contracts the surrounding skin towards the static node of the bone until, at a finite distance away, the force dissipates to zero.

The parameters employed in the first muscle approximation were: the vector contraction, being a variable between 0 and 1, the position of the head and tail of the vector described in three dimensions and the fall-off radius. The parameters describing the first approximation did not provide the desired effects, as linear muscles do not result in a circular contraction towards a node. It was necessary therefore, to establish, firstly the area of flesh influenced by the muscle contraction, secondly, the length of the muscle and thirdly, the position of the muscles in three-dimensional space relative to the underlying bone structures.

The zone of influence (the disturbed area of skin) depends upon the muscle contraction, such that as the muscle tension increases so does the area of skin influenced. Measuring real faces is at best difficult, as the range of surface characteristics varies a great deal from face to face. The individual muscle contractions employed by FACS provide a source for an approximation of the zone of influence. It can be established that most muscles of the face create convex zones that vary from fifteen to one hundred and sixty degrees.

The three-dimensional computer muscle vectors were positioned in the model from anatomical descriptions [6]. The muscle vector consists of an explicit x,y,z head and tail, which is located with a reasonable degree of accuracy, as the points of facial muscle attachment are usually spread over a small area of a few millimeters.

This new muscle model presents a different problem. When the zone of influence for the muscles has been described, it is necessary to compute how the adjacent tissue, such as the node P in Figure 15(a), is affected by a muscle vector contraction. It is assumed there is no displacement at the point of attachment to the static muscle node (the bone) and that maximum flexation occurs at the point of insertion into the skin. Consequently, a dissipation of the force is passed through the adjoining tissue, both across the sector Pm, Pn and Vl, Ps.

To calculate the displacement of an arbitrary node P located on the mesh to a new displacement P', within the segment Vl Pr Ps, a displacement towards Vl along the vector $\overline{P\ Vl}$, is created. This creates $P'(x',y')$ such that:

$$x' \propto f(K,A,R,x)$$
$$y' \propto f(K,A,R,y)$$

and the new location x', y', is a function of an angular and radial displacement where:

K is the muscle spring constant,
A is an angular displacement factor,
D is the vector $\overline{Vl\ P}$ distance.

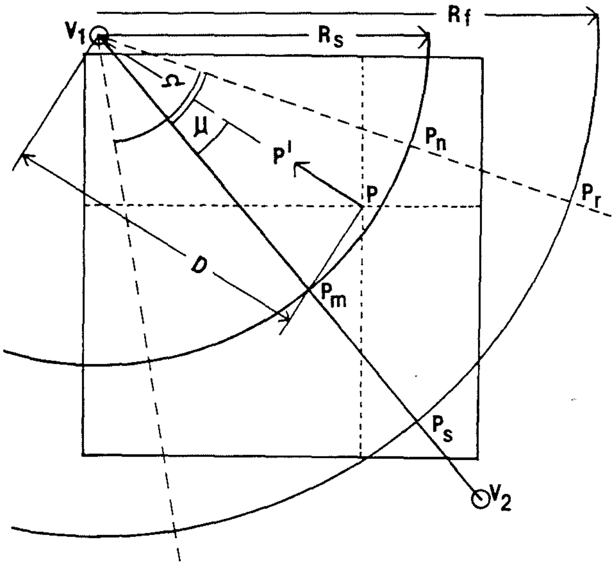


Figure 15 (a)

Linear vector muscle diagram.

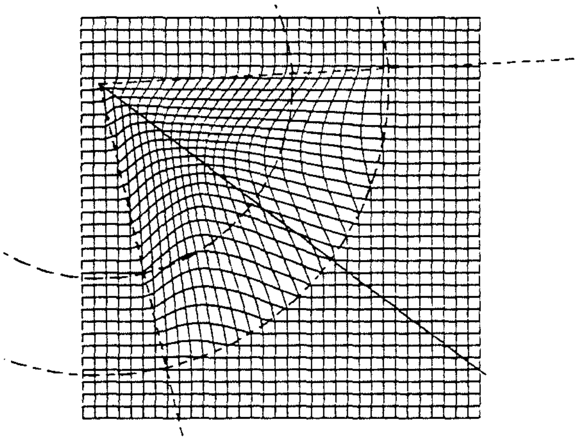


Figure 15 (b)

A three-dimensional linear vector muscle contraction in the x,y plane. Zone of influence $\Omega = 35.0$, fallstart $R_S = 7.0$, fall fin $R_f = 14.0$, muscle spring constant $K = 0.75$, elasticity $E = 1.0$.

The angular displacement factor A is defined as:

$$A = \cos(\mu)$$

where μ is the angle between the vectors $\overline{V1\ V2}$ and $\overline{V1\ P}$. The radial displacement factor R is defined as:

$$R = \cos(1 - D/Rs)$$

for P inside segment V1 Pn Pm, and

$$R = \cos(D - Rs) / (Rf - Rs)$$

for the node P inside zone Pn Pr Ps Pm.

Using a cosine interpolant results in the linear action of a vector muscle contraction illustrated in Figure 15(b). By applying the same principles to the third dimension the point $P(x,y,z)$ is displaced to the point $P'(x',y',z')$. The video sequence V.2 illustrates the animation of the linear vector muscle contraction.

4.3.2 The Sphincter Muscle.

The sphincter muscle that squeezes the skin tissue can be described from a single point around which the surface is drawn together like the tightening of a string bag. This can be described as occurring uniformly around a point of contraction, consequently, the angular displacement is no longer required, and a major and minor axis are employed to describe the elliptical shape of the muscle.

$$x' \propto f(K, Lx, x)$$

$$y' \propto f(K, Ly, y)$$

where:

K is the muscle spring constant,

Lx is the semi-major axis of the ellipse,

Ly is the semi-minor axis of the ellipse.

The equation of an ellipse centred on the orign states that the sum of the distances from the foci is a constant [7]:

$$\frac{x * x}{a * a} + \frac{y * y}{b * b} = 1$$

where:

x,y is a point on the ellipse,

a is the a constant for the x-axis,

b is the a constant for the y-axis.

Therefore the node p in Figure 16(a) can be found to lie within the ellipse and the displacement of p to p' inside the ellipse can be calculated:

$$F = 1-((\mathrm{Sqrt}(x * x + C * y * y))/Lx)$$

where:

C is a constant for the ellipse

$$\frac{Lx * Lx}{Ly * Ly}$$

F is the resultant displacement factor in the range $0.0 \leq F \leq 1.0$.

therefore:

$$x' = F * x$$

$$y' = F * y$$

This activity is illustrated in Figure 16(b).

4.3.3 The Sheet Muscle.

Sheet muscle comprises strands of fibre that lie in the form of a sheet. Good examples of this muscle type are the frontalis major that lies on the forehead and is primarily

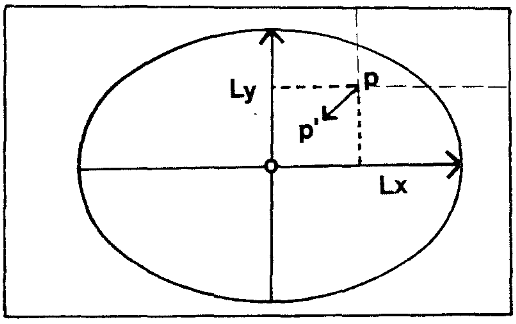


Figure 16 (a)

Sphincter muscle diagram.

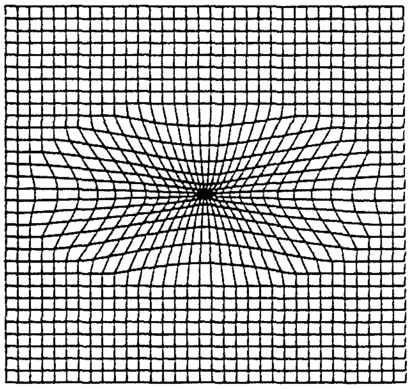


Figure 16 (b)

A three-dimensional sphincter muscle contraction in the x,y plane. Semi-major x axis Lx = 9.0, Semi-major y axis Ly = 4.0, Spring constant K = 0.9.

involved in the raising of the eyebrows. Another important muscle of expression is the corrugator which is at the apex of the chin and is responsible for the raising of the bottom lip.

In terms of the basic function there is no longer the requirement for the angular displacement activity as the muscle does not emanate from a point source and does not contract to a localized node or rather a group of separated muscle fibre nodes. In fact the muscle is a series of almost parallel fibres spread over an area. Consequently it requires the definition of a displacement parallel to the direction of the central muscle vector. This can be define as:

$$x' \propto f(K,D,x)$$

$$y' \propto f(K,D,y)$$

where:

K is the muscle spring constant,

D is a dissipation of the force dependant in which zone the node falls. In the zone A B C D in Figure 17(a) the following expression is employed:

$$D = \cos(1 - Lt/Rf)$$

where:

Rf is the right angle range of the muscle contraction orthogonal to the muscle contraction,

Lt is the length at orthogonal to the direction of the muscle contraction between the muscle vector and node.

if the node falls into the zone C D F E an additional falloff is employed where:

$$D = \cos(1 - (Lt/Rf) * (Vi/VL + Vf))$$

where:

Vi is the parallel vector length to the node,

VL is the length of the vector,

Vf is the extension of the muscle.

This results in the activity illustrated in Figure 17(b).

4.4 The Elastic Properties of the Computer Model.

The flesh of the face varies in elasticity with age, consequently, the cosine interpolant can be modified by raising the function to a power illustrated in Figure 18. This produces the effect of lowering the elasticity of the mesh but produces only small variations in the skin distortion.

The cosine function is only a first order approximation for the modelling of the visco-elastic nature of skin. Alteratively the interpolant can be modified by any non-linear function that is variable between 0 and 1.

The examples in Figure 18, with the associated interpolants, allow a more flexible approach to the modelling of the elastic nature of the skin which can mimic the elastic properties of older or younger flesh. The video sequence V.2 demonstrates the linear activity of the interpolant while the video sequence V.3 clearly demonstrates the effect of non-linear muscle activity. Finally, the video sequence V.5 demonstrates the effect on the mesh by animating the interpolant.

The linear muscle vector parameters are listed in the table T.1, the sphincter muscle in the table T.2 and sheet muscle in the table T.3. These parameters apply in both two and three dimensions. For the linear muscle the definition of the head and tail position is required, for the sphincter

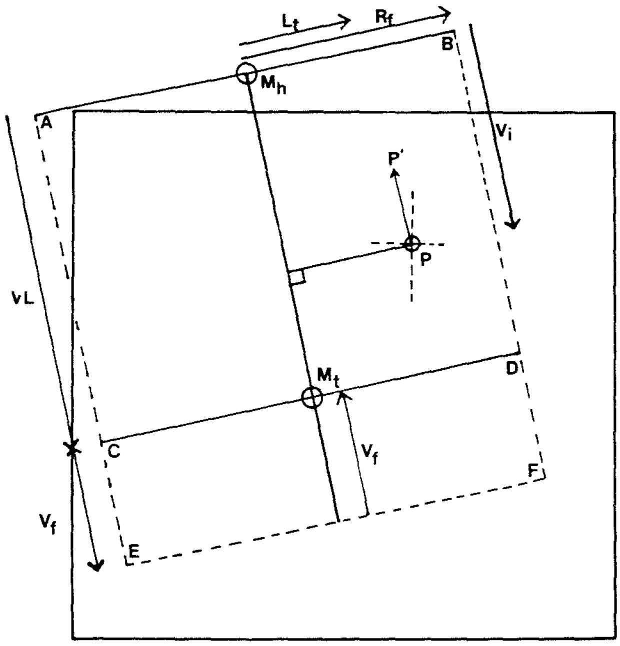


Figure 17 (a)

Sheet muscle diagram.

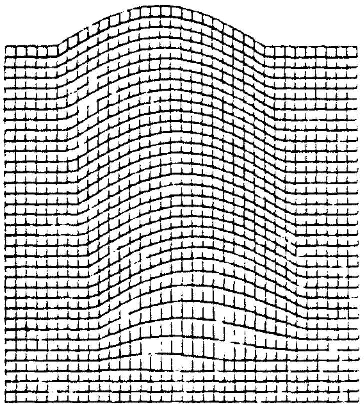


Figure 17 (b)

A three-dimensional sheet muscle contraction in the x,y plane. Right angle range Rf = 4.4, extension $V_f = 3.0$, spring constant K = 0.3.

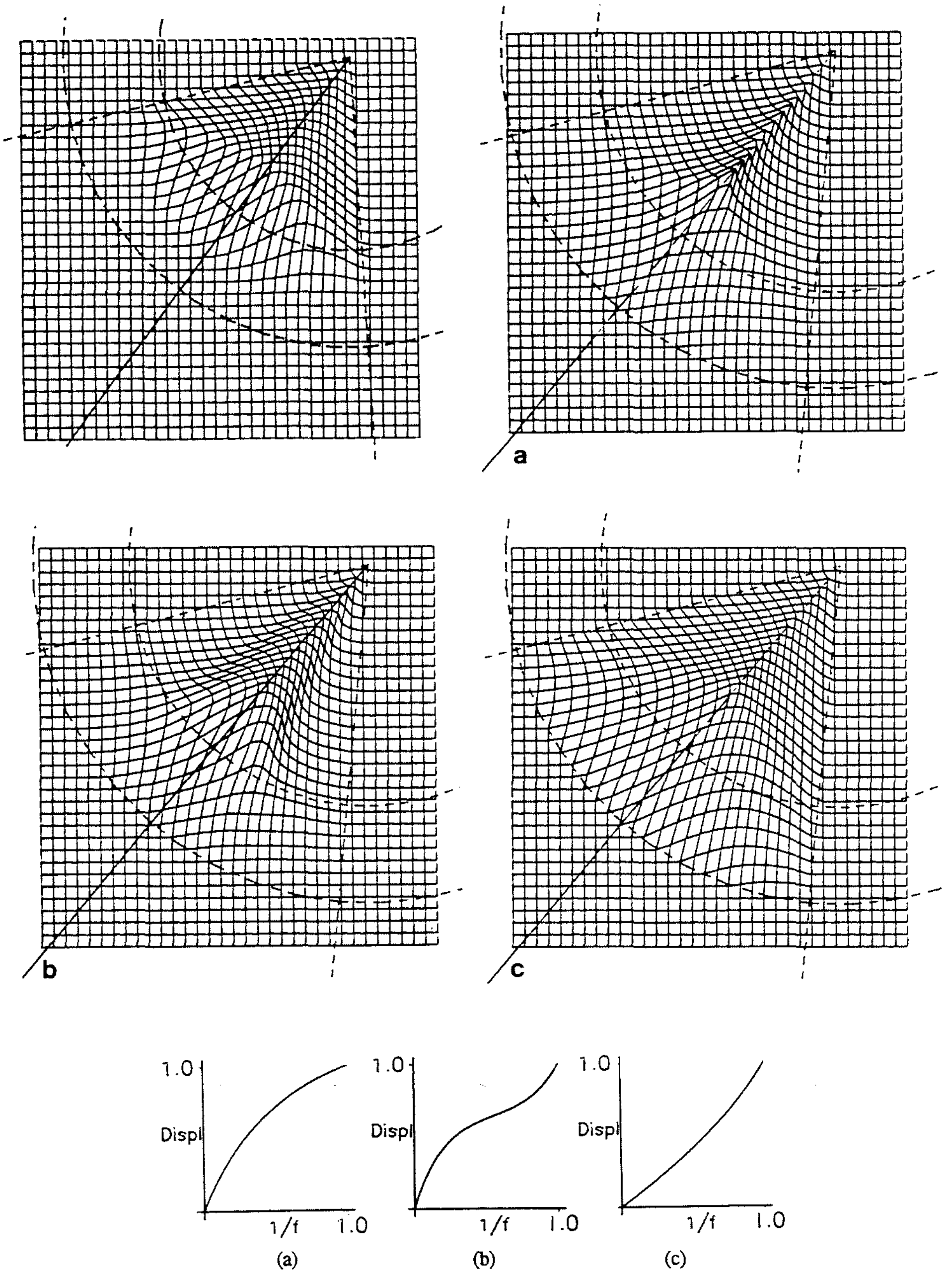


Figure 18

Top left illustrates the muscle contraction cosine raised to a power 10.0. The remaining three illustrations are related to a variable function between 0-1 to vary the elasticity of the mesh. Top right relates to the function (a), bottom left to (b), bottom right (c). All the contractions lay in the x,y plane with a zone of influence $\Omega = 35.0$, fallstart $R_s = 7.0$, fall fin $R_f = 14.0$, muscle spring constant $K = 0.3$.

the centre of the ellipse and for the sheet a head and tail location for the central fibre.

Linear Muscle Parameters		Value Range
Zone of influence	Z	$0 \leq \text{value} \leq 180$
Fallstart	Fs	$0 \leq \text{value} \leq 1000$
Fallfin	Ff	$0 \leq \text{value} \leq 1000$
Muscle spring constant	K	$0 \leq \text{value} \leq 1$
Elasticity	E	$0 \leq \text{value} \leq 1$
Muscle Tension	T	$0 \leq \text{value} \leq 1$

Table 1. Linear Muscle Parameters.

Sphincter Muscle Parameters		Value Range
Major x length of ellipse	Lx	$0 \leq \text{value} \leq 1000$
Major y length of ellipse	Ly	$0 \leq \text{value} \leq 1000$
Elasticity	E	$0 \leq \text{value} \leq 1$
Muscle Tension	T	$0 \leq \text{value} \leq 1$

Table 2. Sphincter Muscle Parameters.

Sheet Muscle Parameters		Value Range
Fallfin	Ff	$0 \leq \text{value} \leq 1$
Elasticity	E	$0 \leq \text{value} \leq 1$
Muscle Tension	T	$0 \leq \text{value} \leq 1$

Table 3. Sheet Muscle Parameters.

4.5 The Generation of Computer Synthetic Facial Expressions.

The objective of the muscle model is to generate realistic expressions on the computer model using the group activity of the muscles. It is necessary therefore, to explain how these expressions can be modelled from FACS.

The face can be considered to be divided into three sections, the upper face, the mid section and the lower portion. The action units are manipulated on the computer model within these sections to prove that they function correctly, and that they can be manipulated together to form complete facial expressions.

Appendix A.2 lists fifty two Action Units and twelve actions that are considered additional [8]. The list has also been compiled with the appropriate muscles for the action alongside. From this list activities for the upper, mid and lower face can be discerned. Figure 19 illustrates the static facial topology with all the muscle relaxed and the eyelids and jaw closed.

4.5.1 The Upper Face Region.

The muscles of the upper face predominately modify the shape of the eyebrows of which there are seven visibly distinctive actions [9]. Action Units 1,2 and 4 can either function independently, but also in the combinations of 1+2, 1+2+4, 1+4, 2+4. Each of them can occur unilaterally or bilaterally and the extension of each can be multiply variable. Figures 20, 21, and 22 illustrate the Action Units 1, 2 and 4.

4.5.2 The Mid Face region.

The muscles of the mid facial region predominately modify the shape of the face from the eyes to the top of the lip. This area has many activities and only the primary actions are considered. The Action Units 41 to 46 concern the activity of the eyelid, while Action Units 5, 6, and 7 concern the contraction of the inner and outer portion of the orbicularis oculi to raise the cheeks, lower the brows and tighten the eyelids. Action Units 9 to 14 concern the contraction of the flesh around the nose, the corner of the mouth and contractions in an angular direction upward and outward.

The actions are usually observed in combination with jaw actions AU 26. Figure 23 and Figure 24 illustrate the Action Unit 26 and Action Unit 9.

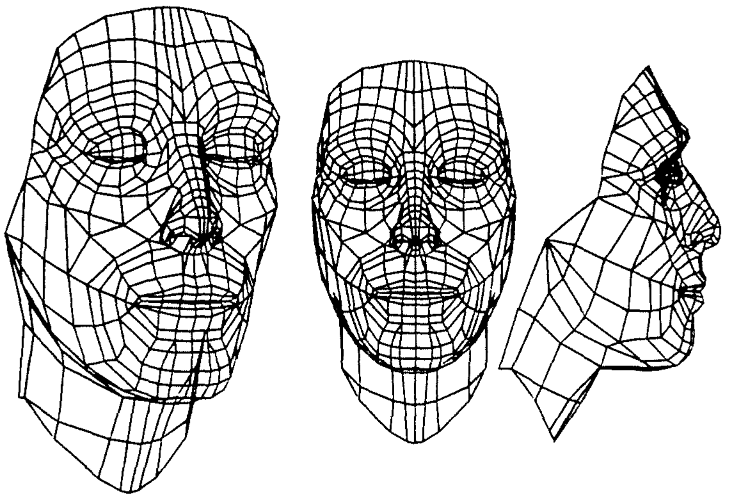


Figure 19

A neutral facial expression in three dimensions. All the muscles are relaxed with the jaw and eyelids closed.

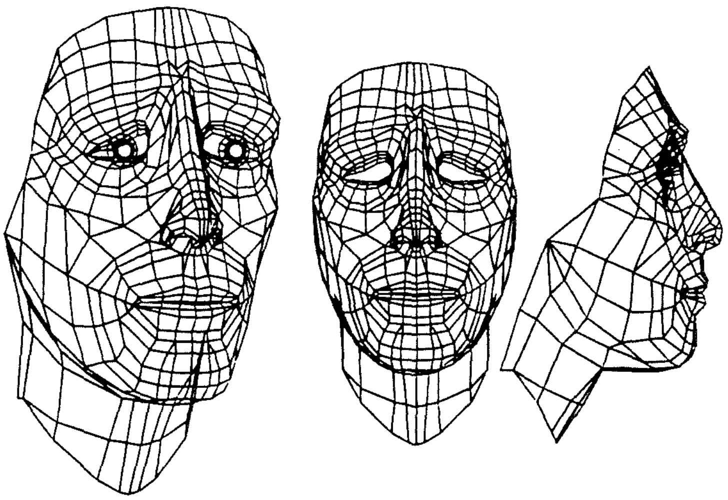


Figure 20

Action Unit 1 the inner brow raiser. The inner frontalis muscle contracts the skin at the root of the nose raising the inner portion of the brow upward.

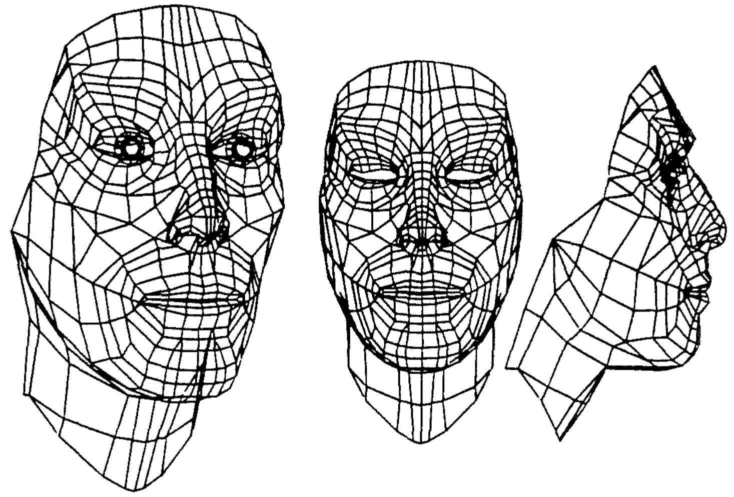


Figure 21

Action Unit 2 the outer brow raiser. The laterial portion of the frontalis major contracts the skin at the outer edge of the brow upward.

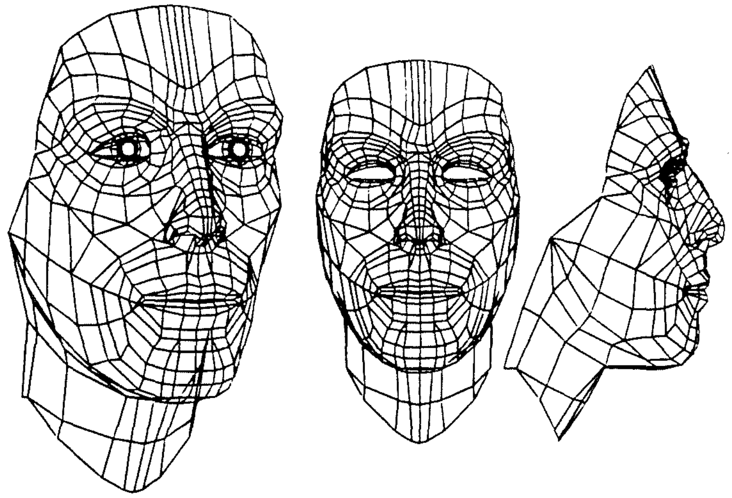


Figure 22

Action Unit 4 the brow lowerer. The corrugator contracts in an angular direction towards the root of the nose drawing the brows downward and together.

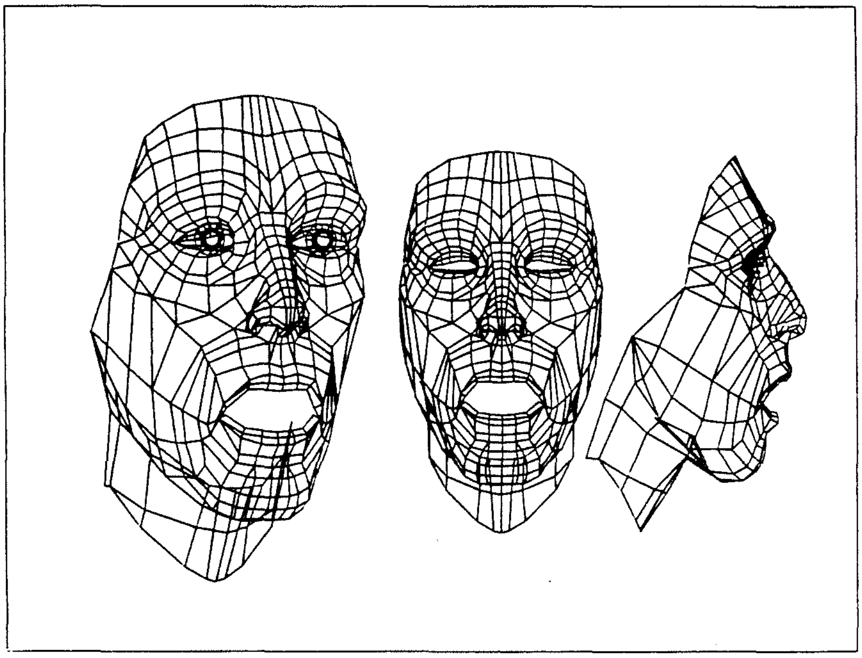


Figure 23

Action Unit 26 jaw drop.

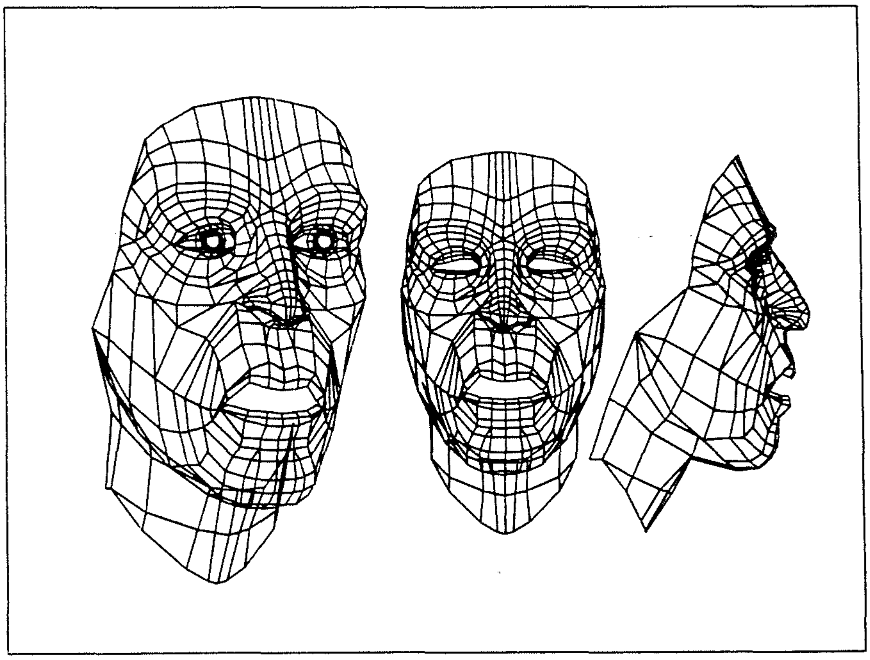


Figure 24

Action unit 9 the nose wrinkler. The levator labii superioris alaeque nasi contracts the skin around the nose base upward, causing the nostrils to dilate and sometimes raising the upper lip.

4.5.3 The Lower Face Region.

The action of the lips and jaw dominate the activities in the lower face and single muscle actions are rarely observed in isolation. Lip puckering (AU18), lips part (AU25) and the lip corner depressor (AU15) are typical actions for the lips.

During speech the vast majority of the muscles around the lips are activated and the sphincter muscle orbicularis oris is responsible for lip funneling (AU22), lip tightening (AU23) and lip pressing (AU24). The combinations of muscle actions to form these mouth shapes are further discussed in 5.10.

Other Action Units important to the mouth parts include: Lips part (AU25), chin raiser (AU17), lip stretcher (AU21), lip corner depressor (AU15), lip presser (AU24). The Action Unit contractions 39 and 16 are illustrated in Figures 25 and 26.

The single action that is noticeably missing, and has not been discussed, is the action of the jaw. The musculature contraction around the mandibular joint is complex and has been ignored because it involves muscle acting on bone to

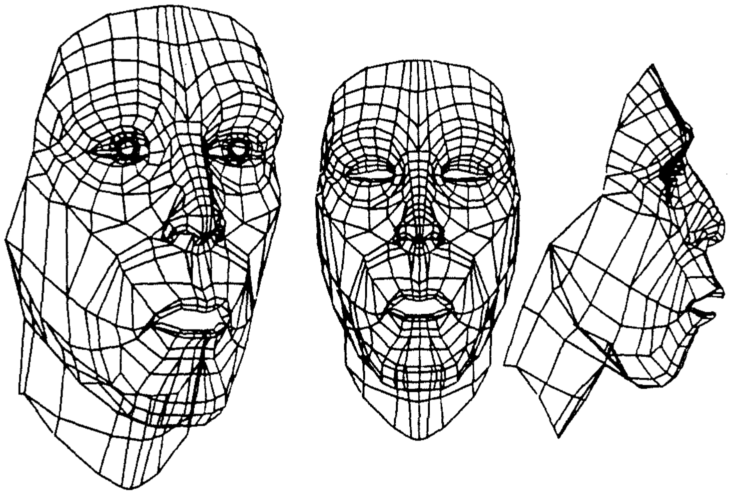


Figure 25

Action Unit 39 the lip funneler. The sphincter muscle that surrounds the lips contracts the skin together like a string bag causing the secondary action of lip pouting.

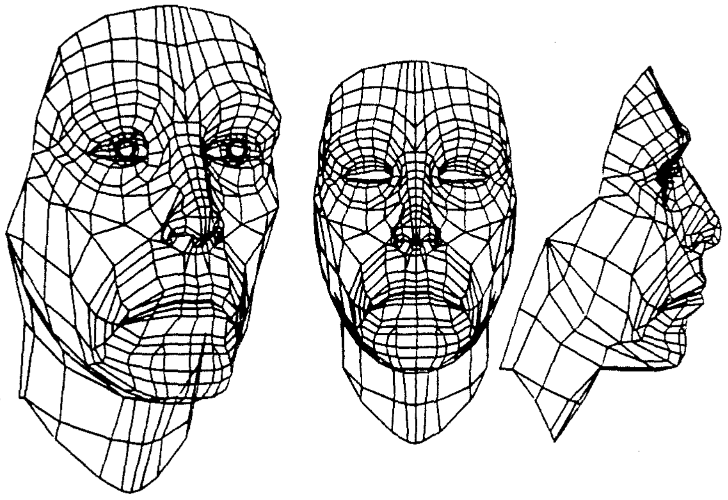


Figure 26

Action Unit 16 the lower lip depressor. The action of the depressor labii inferioris contracts the skin at the corner of the mouth downward towards the chin.

initiate a revolute movement of the jaw resulting in the displacement of the skin. To avoid the modelling the intricate biomechanics, key vertices from the corner of the mouth to the rotational axis of the ear, are extracted from the database to form the jaw. These vertices are then rotated about the mandibular joint to imitate the lowering of the jaw. The first attempts at this action produced square mouth shapes. This was remedied by rotating the corners of the mouth by half the jaw rotation to form a rounded appearance to the mouth opening. This is consistent with the observed effects by Brooke and Summerfield [10]. The muscles that point towards the corners of the mouth and those on the chin where also manipulated to maintain their correct three-dimensional location. This process for the jaw does perform with considerable ease and efficiency and is considered the only really effective solution.

4.6 The Six Generic Human Facial Expressions.

There has been extensive research by the psychologists of non-verbal communication to establish a basic categorization of facial expressions generic to the human race [11]. Ekman, Friesen and Ellsworth proposed happiness, anger, fear, surprise, disgust/contempt, sadness and as the six major primary affect categories [12]. Other expressions such as interest, calm, bitterness, pride, irony, insecurity and skepticism can be displayed on the face but have not been as firmly established as happiness, anger, fear, surprise, disgust and sadness [13]. This present research therefore describes and illustrates how six expressions can be created with the muscle model process. Figure 27 illustrates the neutral face with all the muscle relaxed and the eyelids and jaw closed.

4.6.1 Surprise.

Surprise is perhaps the briefest of expressions [14]. It is sudden with onset and quick to return to its former position. In the upper face the brows are curved and high (AU1 + AU2). In the mid section of the face no there are no distinctive muscle actions. In the lower face the jaw drops causing the lips and teeth to part, the more extreme the surprise the wider the jaw becomes (Figure 28).

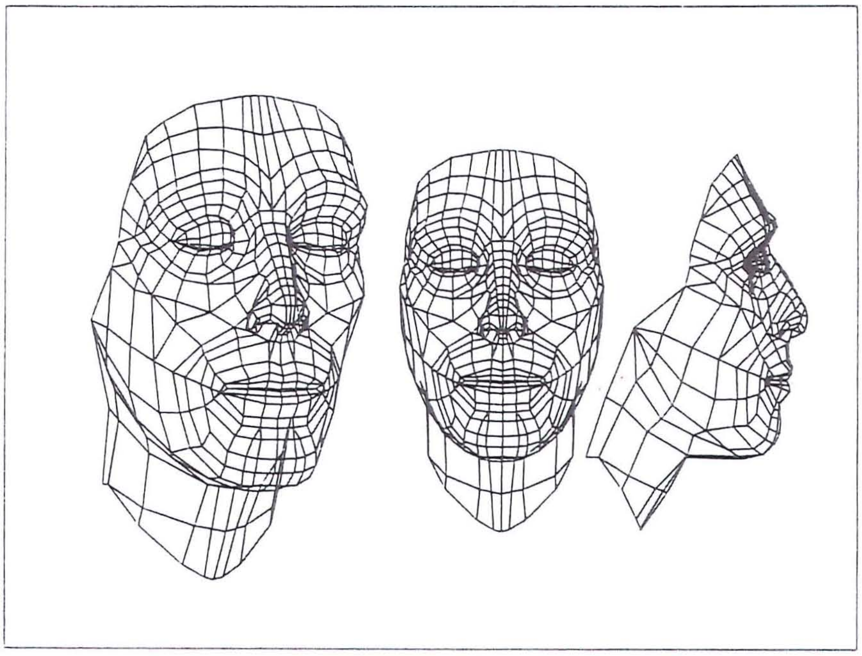


Figure 27(a)

Neutral Face.

Figure 27(b)

Rendered image of the neutral face. All the muscles are relaxed and the eyelids and jaw are closed.

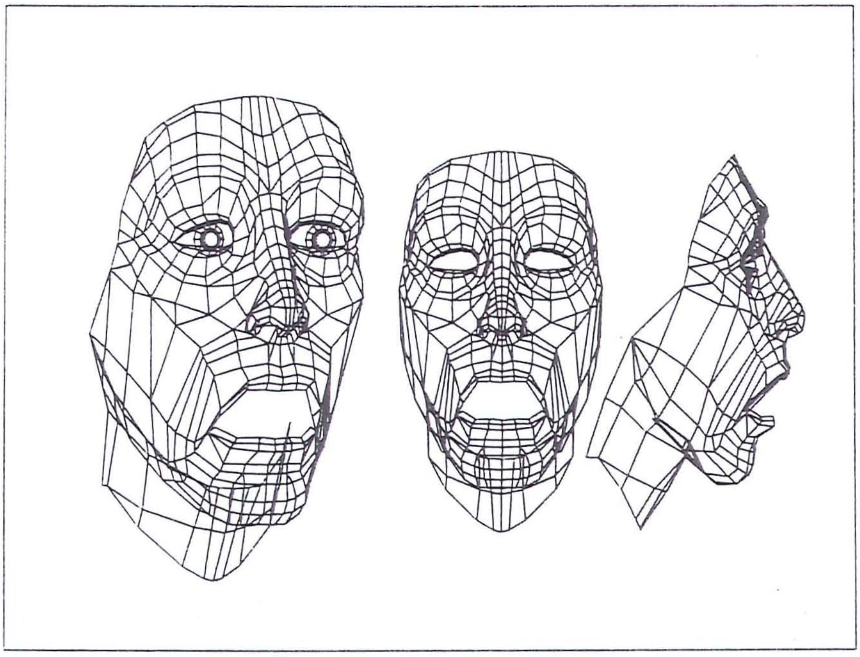


Figure 28 (a)

Surprise.

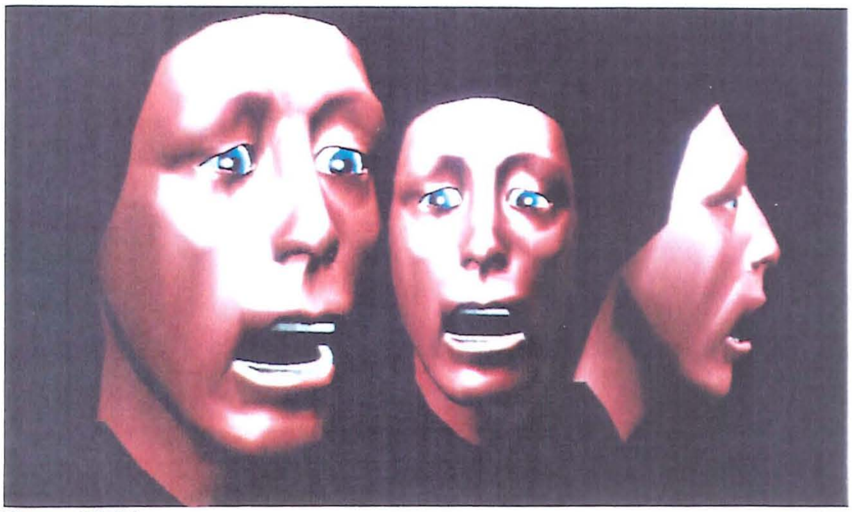


Figure 28 (b)

Rendered image of the expression surprise. The brows are curved and high the eyelids wide with the pupils dilated and the jaw open.

4.6.2 Fear.

Fear varies in intensity from apprehension to terror [15]. In the upper face the brows appear raised and straightened (AU1 + AU2 + AU4), the eyes are tense during fear with the upper lids raised and the lower lid tense, in the mid section the corner of the lips may be drawn backward (AU20) tightening the lips against the teeth. Usually in the lower face the teeth are exposed by the lower lip being pulled downward (AU15 and/or AU16) (Figure 29).

4.6.3 Disgust.

Disgust is an expression of aversion, such as the taste of something you want to spit out [16]. In the upper face there could be the lowering of the brows (AU4), however the primary cues to the expression are found in the mid region of the face round the nose and upper lip. Usually the upper lip is raised (AU9 and/or AU10) that draws up the flanges of the nose. The lower lip may be drawn downward or raised (AU17) (Figure 30).

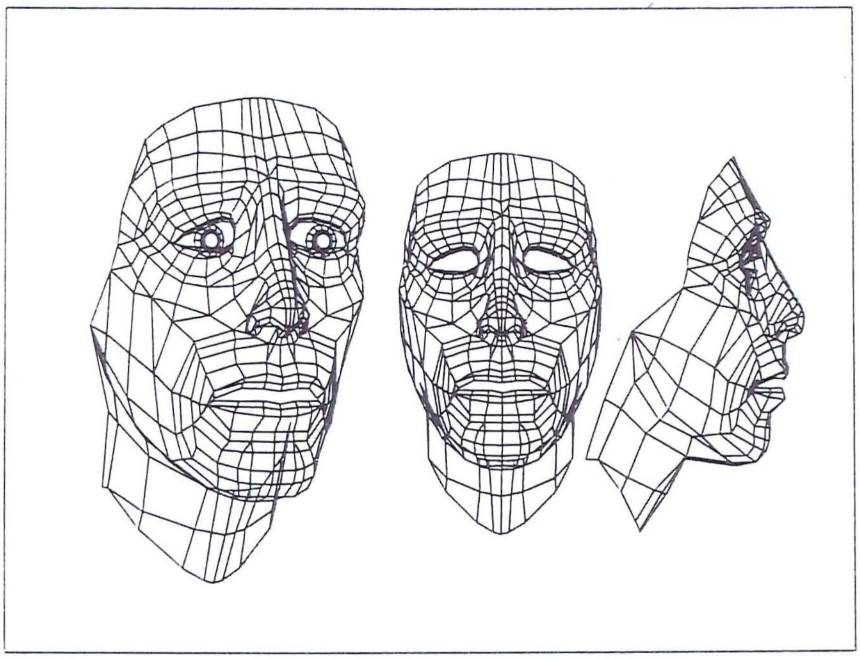


Figure 29 (a)

Fear.

Figure 29 (b)

Rendered image of the expression fear. The inner portion of the brows are raised, the eyes are wide with the pupils dilated. The jaw is open with the lips drawn back.

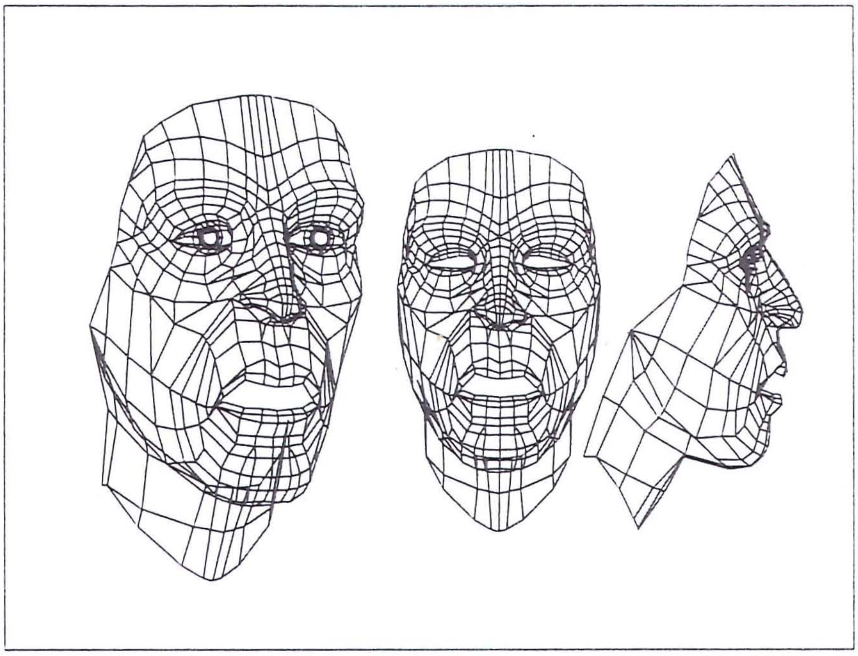


Figure 30 (a)

Disgust.

Figure 30 (b)

Rendered image of the expression disgust. The alaeque nasi muscle contracts the upper lip pulling the skin upward at the base of the nose causing the nostrils to dilate, while the jaw and eyelids are partly open.

4.6.4 Anger.

In the emotional state of anger a person is most likely to harm someone purposefully [17]. The brows are drawn down and together (AU2 + AU4), while the eyes stare in a penetrating fashion with the eyelids wide (AU5). In the mid region of the face the flanges of the nose can be drawn upward (AU10). In the lower face region there can be two distinctive types of motion: the lips closed but pressed hard against the teeth (AU24), or where the lips part to bare the teeth (AU25) (Figure 31).

4.6.5 Happiness.

Happiness is a positive emotion, sometimes in response to pleasurable sensations and can vary in intensity from mildly happy, to joy or ecstacy [18]. In the upper face the brows hardly change, while the eyelids are slightly compressed by the cheek raising up (AU6). The most prominent action is the raising of the corners of the lips that widens into a broad grin (AU12) and this is usually in conjunction with deepening naso-labial folds (AU11) (Figure 32).

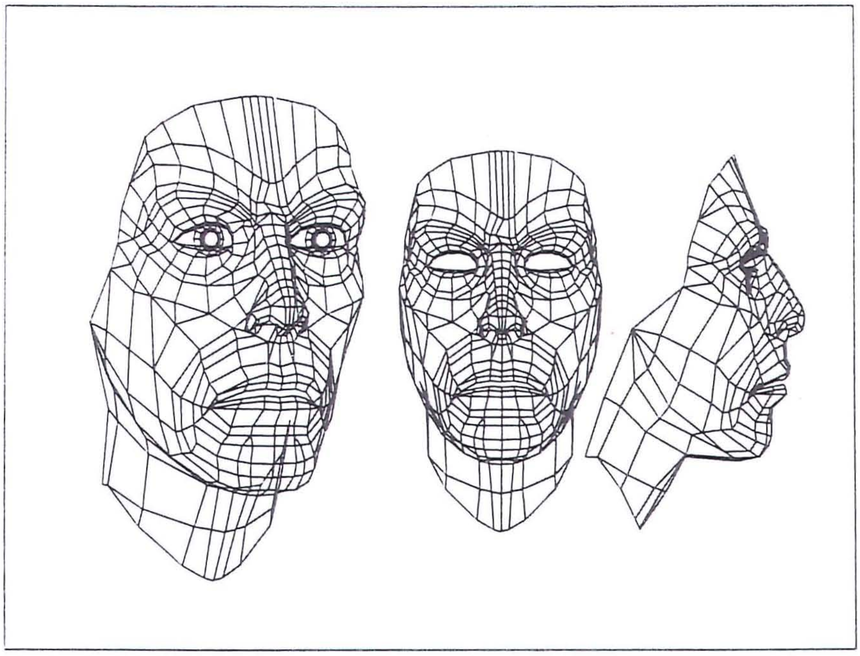


Figure 31 (a)

Anger.

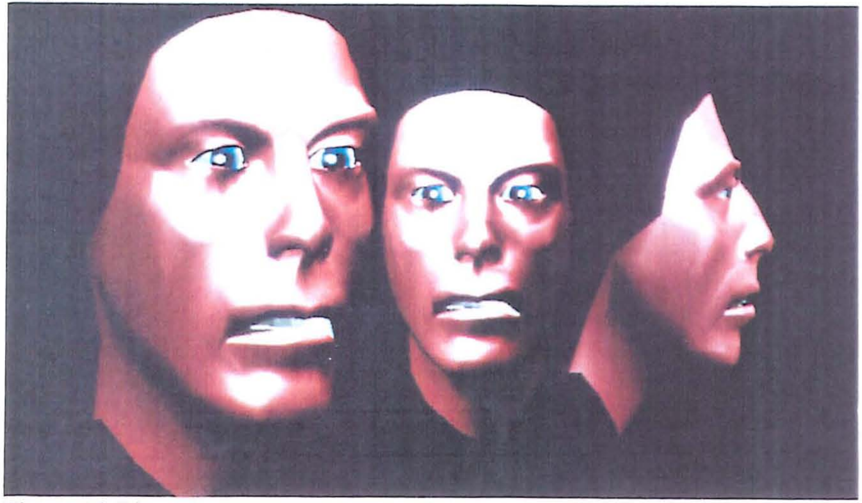


Figure 31 (b)

Rendered image of the expression anger. The corrugator muscle draws the inner portion of the brows together and down. The jaw is not open but the lips are pulled tight and downward by the action of the anguli depressors.

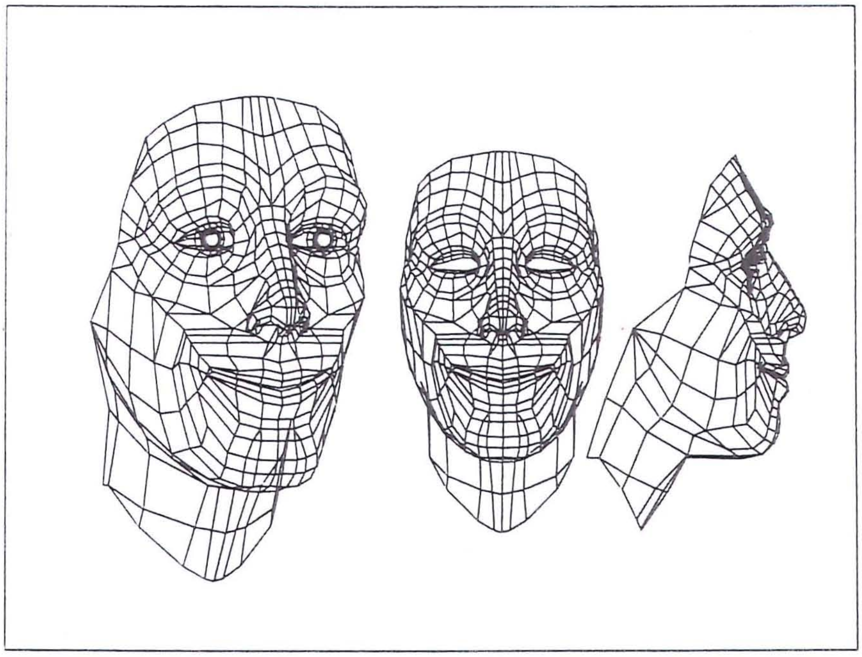


Figure 32 (a)

Happiness.

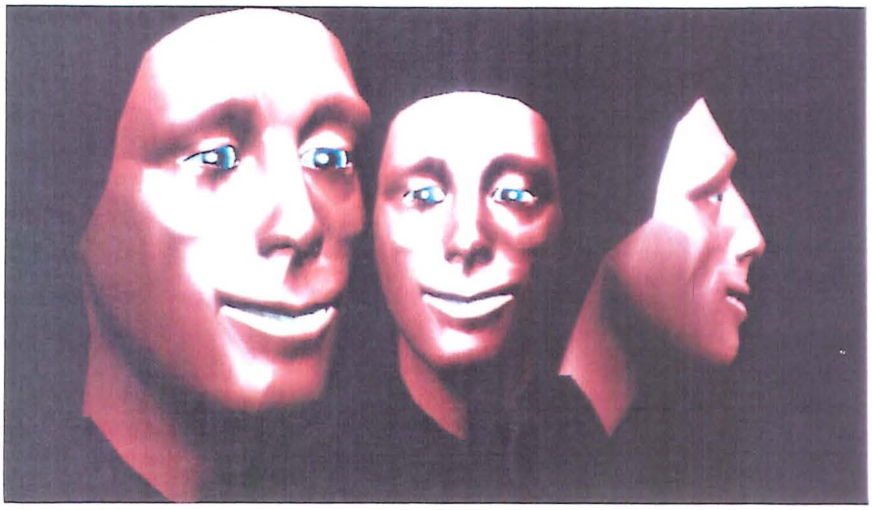


Figure 32 (b)

Rendered image of the expression happiness. The corners of the lips are drawn upward and obliquely by the zygomatic major muscles.

4.6.6 Sadness.

Sadness is endured stress, and more often than not prolonged in duration, the opposite of the briefest surprise [19]. In sadness the inner portion of the brow is raised and drawn together (AU1 + AU2 + AU4). The eyes are usually cast downward and the lower lids slightly raised. The mouth displays subtle motions that are akin to the expression of disgust where the corners of the mouth are pulled down (AU15) (Figure 33).

4.8 Muscle Confluence.

The muscle model has so far been described in terms of independent muscle actions. While this is part of the picture, there are further complications when the muscles interact. Every muscle has a limit to its length or range of contraction, and this varies from muscle to muscle. In a dynamic environment muscles interact, pulling in a wide range of directions with varying force. Therefore, when the resultant skin displacement is modelled in the computer, the interactions of muscles have to be taken into consideration.

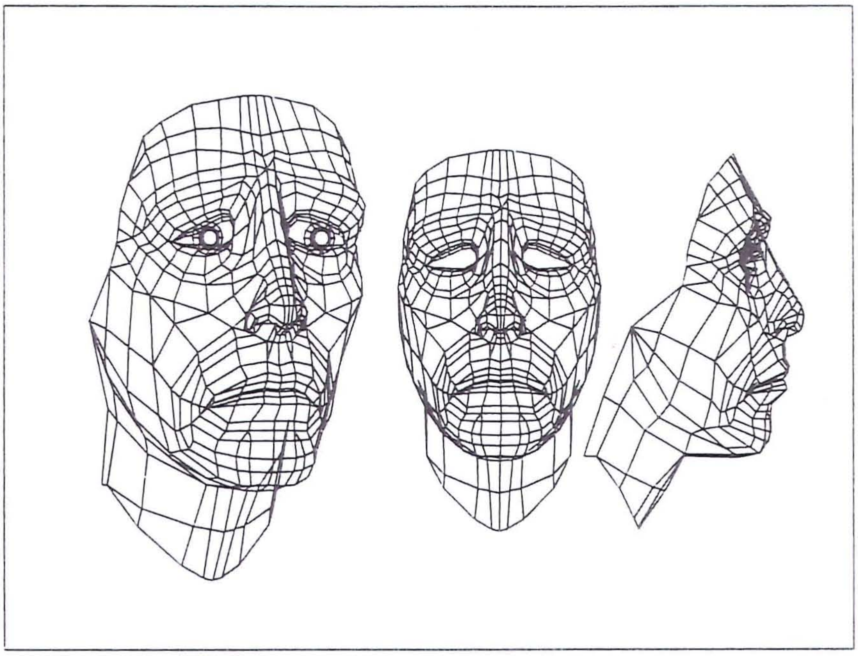


Figure 33 (a)

Sadness.

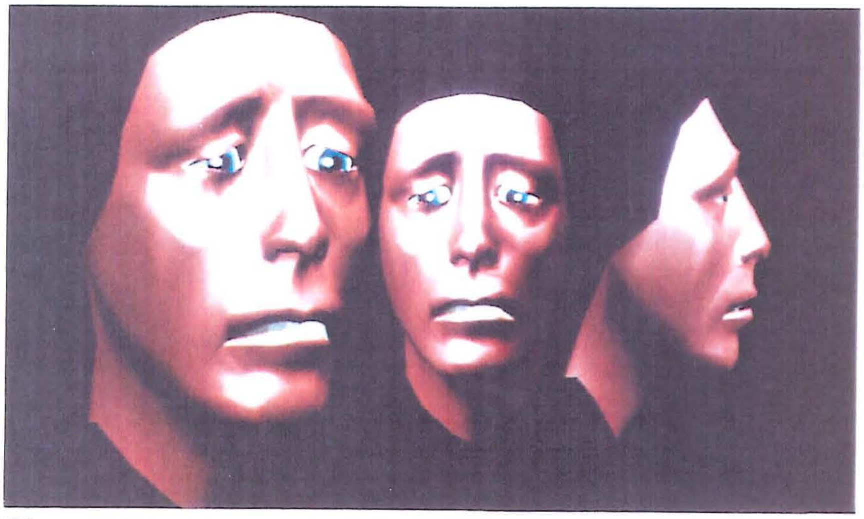


Figure 33 (b)

Rendered image of the expression sadness. The inner portion of the frontalis muscle is raised. The corner of the lips are drawn downward by the action of the anguli depressors.

A fundamental musculature interaction can be considered with two linear muscles with overlapping zones of influence. To demonstrate the examples a regular three-dimensional mesh is employed. Those nodes of interest can be found in the zone;

$$z(A) \cap z(B) = z(C)$$

where:

$z(A)$ and $z(B)$ are the two muscle zones,

$z(C)$ is the intersection between the zones.

and the sum of two vectors can be established by the parallelogram law:

$$A+B = C$$

Therefore any node within $z(C)$ the resultant force from each vector is established such that the resultant force on the node by the coincident muscles can be calculated;

$$F[p(C)] = F[p(A)] + F[p(B)]$$

where:

p is the node under consideration,

F	is the Force applied to a vector,
A and B	are the influencing vectors,
C	is the returned vector.

Figure 34 demonstrates the interaction of such a muscle combination. For the computer model it is necessary to compute, before displacement, those nodes involved in the confluence. Then the process is to compute, per frame time, the resultant vector acting on the node. This muscle confluence activity is illustrated in the video sequence V.4.

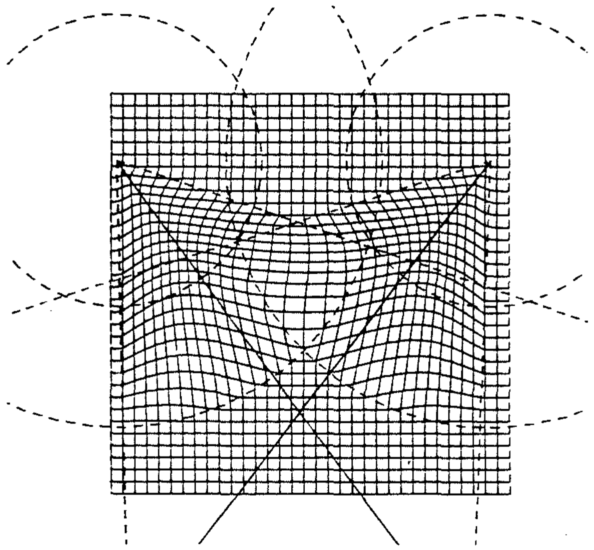


Figure 34

The confluence of two linear muscles contracting the skin in opposite directions.

4.9 Conclusions.

The three computer muscle functions - linear, sphincter and sheet - provide approximations to the bio-mechanics of muscle activity. The parametric controls for the muscles are extendable to any non rigid object and are independent of specific topological considerations. The muscle model process has been designed to perform complex articulations in reference to the notation of the Facial Action Coding System. Not all of the action units have been demonstrated in this present research, however with seventeen muscles and the jaw rotation, the six major expressions can be generated along with many other combinations.

The images are generated from muscle contractions, in accordance with FACS, to produce static posed expressions. The muscle model however, can produce more than an all-or-nothing contraction as required for by FACS. The muscle model can vary the elasticity and the tension parameters, this permits two important characteristics. Firstly, muscles can generate subtle expression blends and nuances, and secondly, the muscle contractions are variable over time. The muscle model process therefore, enhances FACS capabilities from static to motion attributes as well.

The muscle model process is the first order approximation for the computer synthesis of muscle activity in the face. The approach is fundamental, but it is based on the underlying bone, muscle and skin structure of the face. Improvements to the model can be made such as bulging, wrinkling and flow around skin (see 8.3). These are, however, considered secondary actions to the basic distortion of the skin. A more comprehensive approach would be to incorporate tension nets within the muscle model process, such that a simulation of visco-elastic flow of skin could be generated (see Chapter 8.4).

4.10 References.

[1] Paul Ekman and Wallice Friesen, Manual for the Facial Action Coding System, Palo Alto Press, California, 1977.

[2] R Warwick and P L Williams, Gray's Anatomy, 35th Edition Longman, London, 1973, p. 492.
H Ferner, Atlas of Human Anatomy, Urban and Schwarzenberg, Munich, 1983.

[3] B R Landau, Essential Human Anatomy and Physiology, Scott Foresman, Illinois, 1923, pp. 156-166.

[4] R Tomovic and R B McGhee, 'A Finite State Approach to the Synthesis of Bioengineering Control Systems', IEEE Transactions on Human Factors in Electronics, Vol. 7, No. 2, 1966, pp. 65-69.

[5] Landau, Essential Human Anatomy, pp. 156-166.

[6] Warwick, Gray's Anatomy, pp. 488-503.

[7] Charles C Carico and Irving Drooyan, Analytical Geometry, John Wiley, New York, 1980, p. 55.

[8] Ekman, Manual for the Facial, 1977.

[9] Paul Ekman, 'About Brows: Emotion and Conversational Signals in Human Ethology',Journal of Phonetics, No. 11, 1983, pp. 169-202.

[10] Micheal Brooke and Quentin Summerfield, 'Analysis, Synthesis and Perception of Visible Articulatory Movements', Journal of Phonetics, Vol. 11, 1982, pp. 63-76.

[11] Daniel Goleman, 'The 7,000 Faces of Dr.Ekman', Psychology Today, February, 1981, pp. 43-49.
Paul Ekman, Emotion in the Human Face, Cambridge University Press, London, 1972, pp. 147-175.

[12] Paul Ekman, Wallice Friesen and Phoebe Ellsworth, Emotion in the Human Face, Pergammon, New York, 1972.

[13] Paul Ekman, Unmasking the Human Face, Prentice Hall, New York, 1971, p. 22.

[14] Ibid, pp. 34-46.

[15] Ibid, pp. 47-65.

[16] Ibid, pp. 66-77.

[17] Ibid, pp. 78-98.

[18] Ibid, pp. 99-113.

[19] Ibid, pp. 114-126.

CHAPTER FIVE

Chapter Five

Speech Synchronization

5.1 Introduction.

The muscle model process described in Chapter 4 demonstrates that with seventeen muscles and a jaw rotation, a variety of facial expressions can be created. These expressions are considered either as frozen expressions in time, or posed facial distortions. Both may produce very similar results, but they contain important perceptual differences. Ekman describes two types of expression in addition to happiness, anger, fear, disgust, surprise and sadness [1]. Firstly where the context is relevant, there are the emblematic expressions that have a verbal equivalent in common words or phrases, such as the eye wink signal for agreement, 'right on','sure', or a flirtatious, 'will you?'. Secondly, there are conversational punctuators, where rapid facial signals emphasize what is being said in words. Unfortunately, static images of these expressions do not clearly reveal the unique and subtle differences between them, and only the dimension of time provide sufficient information for a precise judgment on the displayed expression to be made. Another typical example of expression confusion is a 'feigned smile' to the camera compared to a smile of joy or happiness. Here the speed, duration and offset time provides vital cues to the expression.

So far the present research has assumed the models to be mute, but speech is the primary mode of human communication with facial expressions playing only a secondary role during conversation. The action of the two together provides a very powerful mode of communication [2]. Consequently, the aim of this present research is to combine the time variant activity of speech with expressive facial characteristics for computer synthesis.

Some of the first images of animated speech were created by Georges Demeny in 1892 with a piece of equipment he called the Phonoscope [3]. The principle involved photographing a subject's lips and the lower part of the face sequentially. Then by mounting the images on a rotating disk, the images could then be viewed through a spy hole. The images rotated at approximately thirty frames a second, fast enough to exploit the phenonenon of persistence of vision and give a convincing appearance of motion. Although this technique is nearly a hundred years old, it is still applicable to computer animation and a similar process is employed in film animation.

Speech synchronized animation presents the most complex problem for computer synthesis, as multiple activities not only have to be coincident with one another to produce the desired mouth shape, but they also have to synchronize with the audio in time. As a result, it is the most intricate problem concerning facial animation to accomplish convincingly.

A critical problem arises from the single fact that animation is not concerned with speech production, rather the creation of visible mouth shapes to fit to an existing sound track. Human speech is produced by making use of the lungs, throat, mouth and nose. A stream of air is exhaled by the lungs through the trachea and the remainder of the vocal tract. This makes the vocal cords vibrate, and the perceived pitch of the voice is modified by the movement of the tongue, jaw, lips and velum. Therefore existing techniques for the recognition of visible articulatory gestures on the lips are significant, as these can then be analyzed, synthesized and manipulated for the use in the computer animation. This chapter therefore, is not concerned with speech production, but the creation of animated synchronized speech to an existing sound track.

Investigations 1-4 are chronologically organized and assume that speech sequences comprise of a multitude of mouth and lip shapes that can be edited together to synchronize with the spoken track, allowing the complexities of speech production to be minimized.

5.2 The Phonetics of Speech Employed in Visual Speech Perception.

Speech comprises a mixture of audio frequencies, and every speech sound belongs to one or other of the two main classes known as vowels and consonants. Consonants and vowels fall into groups, with members of these groups called visemes. It was the early teachers of lip-reading/visual speech perception, such as Jones at the turn of the century, who further broke-down speech sounds down into small linguistic units known as phonemes [4].

Lip-reading is based upon the observation of forty-five English phonemes. These phonemes are audible in speech but cannot all be seen by the lip-reader. Traditionally, lip-reading has been considered as a completely visual process developed by the small percentage of people who are completely deaf. There are, however, three mechanisms employed in visual speech perception: auditory, visual and audio-visual, and much of the emphasis in lip-reading for those with impaired hearing is concerned with the audio-visual.

Studies of both hearing and hearing impaired people indicate scores of 38% to 68% correct visual vowel identification in speech using a variety of techniques (film, video and face-to-face) [5]. Consonants are not so reliably observed. Viseme groups such as /p,m,b/ and /f,v/ are reliably observed like the vowels, while confusion amongst individual consonants within each visemes group is more common. Within the phonetic distinctions, McGrawth reports that 30% - 50% of audible phonetic distinctions can be seen [6]. In terms of the information value, /t, d, n, s, l/ are frequently misinterpreted and classed into the same viseme group. Unfortunately, this is further compounded by the fact that they are the five most common phonemes in English, and their co-articulation in fluent speech increases the lip-reading difficulty. It is also misleading to assume that there is a correlation between visemes, phonemes, words and sentences, as this is dependent upon the particular lip-reading task involved [7].

Despite the low threshold between understanding and misunderstanding, the discrete phonetics provide a useful abstraction as they group together speech sounds that have acoustic or articulatory features in common. As a result phonetics have been used extensively in animated models in both two and three dimensions.

The principles of animated speech employing phonetics involve developing a selection of mouth shapes that mimic the phonetic structures. The sound track is searched and the phonetics identified and timed as keys on a exposure sheet. These keys can then be inbetweened or interpolated, establishing the intermediate frames of the animation.

5.3.1 Cell Animation Techniques.

The technique used by traditional animators (such as the Disney studios) for two-dimensional cartoon characters employed phonetic break-down for lip animation [8]. Depending upon the individual character, a limited number of phonetic structures (usually less than ten) would cover most lip synchronization possibilities. The following categories were employed;

i. A, E, I are open vowels and require the lips and teeth to separate.

ii. M, B, P are consonants and are closed mouth shapes that are accents on the words.

iii. U, O, W create an oval mouth shape

iv. F, V are complex articulations where the lower lip is tucked up under the top teeth [9].

Once the basic mouth shapes have been established, it is then necessary to analyze the sound track. Words that have accents in relation to the frame time are pin-pointed and marked and the accents are isolated and identified as falling into one of the previously mentioned categories. Once all the key-frames have been identified and noted on the exposure sheet, inbetween frames can be drawn to make up the remaining animation sequence.

From the animator's experience a variety of additional techniques have been noted, for example;

> Anticipate the dialogue with the head, body or gestures three or four frames ahead of the heavy modulation, but the lip synchronization should occur on the modulation. If the mouth is creating a closed mouth shape, maintain the closure for at least two frames [10].

Fundamentally the principles used by traditional animators can create quite acceptable lip synchronization. Interestingly they point out that there are no hard and fast rules to adhere to. In fact they became pragmatic in their approach, if it worked then they used it, if it did not they experimented. In computer animation similar techniques have been used extensively.

5.3.2 The Hight Lip-Reader Trainer.

The Lip-Reader Trainer is a computer-based system developed by Hight as an aid for the deaf and hard of hearing [11]. The software package ran on a micro-computer, while displaying two-dimensional bit-mapped colour images of the lips, tongue and teeth. A total of nineteen phonemes were created to form a basic phonetic library that could be loaded sequentially to form speech segments. The lip-reader program was capable of calculating lip-position bit-maps at a rate of fifteen frames a second, fast enough for the successive positions to appear animated, and as a result the lip-reader could pick up the gist of the communication.

The system was designed for lip-reading and therefore it avoids synchronization problems. However, it does illustrate the system's inability to stress intonation. The addition of more phonetic keys would appear to be the solution, but this would only cover a finite number of potential mouth shapes.

5.3.3 Moore and O'Connor's Talking Head.

This system was developed by Moore and O'Connor and like the Lip-Reader Trainer, was developed on a micro-computer as an aid for the speech impaired [12]. The application was carried out in three dimensions and included the whole face. The three-dimensional phonetic poses were synthesized from real people, who mimed some twenty phonemes (eleven vowels, eight consonants and one at rest) while simultaneous photographs were taken. These photographs were then digitized to form a three-dimensional model which could be simply rendered to give the impression of solidity. It was assumed that the nose and forehead were static and not involved with speech, and these features were only input once. The advantage of that was that only the differences between mouth positions needed to be stored.

The procedure involved establishing key poses for each phoneme, with the individual mouth shape for any moment in time calculated by interpolation. The animations were created by building a succession of key phonemes, and a number of variable parameters were manipulated to create the desired sequence. Approximately one hundred frames could be stored in the system that ran at fifteen frames a second, which is about one phoneme per frame.

The advantage of using three-dimensions is that the face can be viewed from any angle in space, which is an additional aid for the lip-reader. In their conclusion, the authors suggest two further areas of interest. First, the three-dimensional modelling of the face that would allow expressions to be formed to enhance the communication of the spoken word, and secondly, an expansion of the system to model the internal organs of the mouth (teeth, velum, tongue and larynx) as a substantial aid for speech therapy.

The major draw-backs of this environment are similar to those of the Hight Lip-Reader Trainer. Firstly, phonetic structures do not provide intonation and, secondly, explicit definitions of at least the differences between facial positions have to be stored.

5.3.4 The University of Montreal DADS and TRANNA.

In 1984 the University of Montreal produced an animated short called Tonie De Peltrie which involved many innovative techniques for the animation of three-dimensional synthetic characters and included the development of the packages DADS and TRANNA [13]. Speech synchronization was achieved by first recording the animation onto one inch video tape and then transferring it onto 16mm film, then using traditional animation techniques the phonetic sound track was recorded onto a exposure sheet. The model's mouth shapes were phonetic in structure and recorded from real-life poses. A mapping technique was used to relate the real person's face to the models, and finally idiosyncratic key mouth shapes could be created by interpolating between different phonetic poses. Once the sequence was ready, it was previewed in real-time to check for errors.

With the capability of real-time playback, and a bank of CBSE (Character Bank of Standard Expressions), it was possible to create reasonably accurate speech synchronization. The major disadvantage of the system was that the animation was limited to sequences recorded from real life, and any additions would have to similarly be re-recorded from the real world.

5.3.5 Parke's Parametric Model.

The early work of Parke at the University of Utah took a different approach to speech synchronization [14]. The basic principles involved creating a precise model that could be modified by parameters. This meant that certain vertices of the model were harnessed to particular attributes, so that the parameters controlling the mouth were used to give the required animation capability. The open vowels and the accent consonants were controlled mainly by jaw rotation. The oval mouth was a combination of jaw rotation, lip scaling and translation of the lips away from the teeth. An exposure sheet of the phonetics was then created from the sound track and translated into command files that controlled the mouth parameters over time.

This parametric approach is by far the most effective mode of controlling mouth shapes. It avoids the complexity of having to store explicit definitions and allows a multiplicity of mouth shapes to be created.

The main disadvantage of this technique concerns the hard-wiring of the performable actions. The facial model has to be specified before the motion characteristics are attributed to the model. This means that each model will be unique in structure and in its motion capabilities.

5.4 Analysis and Summary of Techniques.

Cell animation is mainly concerned with characterization, and this provides a greater latitude in lip synchronization. For example, a beach ball or door knob with six or seven key phonetic mouth shapes could be successfully animated. But as the face approaches human proportions, and indeed if the model is attempting to mimic human motions, the phonetic approach does not suffice. In fact we make many critical comparisons between a live actor speaking and a synthetic face and cannot accept the visual dichotomy. This is a critical problem because the higher the image fidelity of the synthetic face, the greater the invitation to compare it with real people's faces.

Speech is more than mechanically shaping the mouth using phonemes as a guide. Phonetics are the mouth shapes usually formed in precise connection with individual linguistic units. But fluent speech, in contrast, does not maintain such discrete mouth shape definitions as it co-articulates from word-to-word. Additionally, the intonation of what is being said influences the mouth shapes, especially if the speech has an emotional content or is spoken in a loud voice.

Summerfield succinctly points out that lip-reading, and therefore phonetics, cannot completely replace auditory speech perception:

> Hollywood's conception of the secret agent who perfectly divines her opponent's plan by observing a muttered conversation across a dimly lit cafe is fantasy. Research suggest that the agent may fare better if, implausibly, she could persuade her adversary to face her at a distance of 1.5m in bright illumination so that his arms and torso were visible in addition to his face, to remove any disguises such as a beard or moustache that might obscure the mouth, to speak naturally but slowly, and possibly to wear lip-stick to emphasize the pattern of the lip movements. Even then the agent would probably find the twenty-four acoustically distinguishable English constants fell into at most a dozen optically distinguishable groups, and maybe only as few as four [15].

Lip-reading can only provide an impoverished indication of what has been spoken. This is again emphasized by the homophonous groups of consonants (a number of consonants that cannot be distinguished) such as : /th, dh/, /t,d,n/, /s,z/. Skilled lip-readers may be able to distinguish some of these groups, but fluent speech co-articulation can easily undermine these distinctions. Summerfield concludes that lip-reading is of limited use on its own, because the relationship between the visible appearance of phonemes and their sound is ambiguous. Finally, he suggests that bi- or multi-modal cues are vital, as we, as humans, are predisposed to relate what we see to what we hear [16].

It is clear that the techniques employing phonetic speech structures provide only a partial solution to speech synchronization. Elementary two-dimensional faces, simplistic three-dimensional characters, exaggerated puppets and other inanimate structures can be successfully synchronized with speech phonetics. However as the models approach, or indeed attempt to mimic, real people speaking, phonetics do not suffice and an alternative technique is required.

5.5 Recording and Measuring Facial Articulatory Movements.

If speech phonetics are not to be employed in the computer synthesis, it is necessary to obtain data of the mouth during speech articulations to establish parameters for the duration and displacements of individual parts of the mouth in three dimensions. There have been some attempts to measure and quantify facial articulation during speech.

5.5.1 Static Vowel Recordings by Fromkin.

An early investigation by Victoria Fromkin in 1964 attempted to determine parameters for lip positions in American English vowels [17]. Data was collated from standardized simultaneous frontal and lateral photographs, lateral X-rays and plaster casts, and the dimensions that were measured were as follows;

i. Horizontal width of the mouth opening.
ii. Vertical height of the opening.
iii. Area of the opening.
iv. Distance of the forward most points of the lips.
v. Protrusion of the upper lip.
vi. Protrusion of the lower lip.
vii. Distance between the upper and lower front teeth.

Information concerning the width, height, area and protrusion of the lips can be extracted from her data. These results, in a graphic form, appeared to vary in width and height during successive vowel pronunciations by the same individual. It was useful for this research to establish the latitude in the generation of mouth shapes for the synthetic speech synchronization. Table 4 and Table 5 therefore, are a summation of Fromkin's results by the present author to quantify the variation in height and width of an individual producing vowels

	Mean Width	Mean Height	Mean Area (squared) (0.7 * H * W)
a	41 mm	11 mm	315.7 mm
e	40 mm	9 mm	252.0 mm
i	40 mm	8 mm	224.0 mm
o	24 mm	6 mm	100.8 mm
u	16 mm	4 mm	44.8 mm

Table 4. Vowel Mean Height, Width and Height

	Pv Height	Pv Width
a	34.6 %	19.4 %
e	30.7 %	11.9 %
i	27.2 %	22.2 %
o	41.5 %	20.0 %
u	19.0 %	23.5 %

Table 5. Vowel Percentage variation of height and width.

While these results appear to be of value for the formation of lip positions of vowels, it must be noted that variations between individuals were pronounced: an overall range of 17mm for the width of the mouth opening, a range of 11mm for the height, and a corresponding 30 mm squared for the area for the opening of the mouth. The mean statistical variation for individuals producing five repeated static vowels was

21% with an outer range of 37%. For repeated samples from the same individual, 17% variation and an upper limit of 27%. Consequently, lip shapes can be seen to vary enormously between individuals and also with successive acts by a single individual.

Even a much lower level of tolerance of 5% would result in difficulties of determining the definition of the lip width of /a/, /e/, and /i/. The reasons for such a high percentage discrepancy are two fold. Firstly, measurements have to be sensitive since, relative to the size of the head, significant articulatory excursions are small and rarely exceed about 25mm, and the measurements have to isolated from the global motions of the head. Secondly, and perhaps more importantly, talkers cannot generally produce specified graded articulatory gestures, especially static phonetic shapes.

Despite the inherent variability of the results, valuable general conclusions can be summarized:

i. A simple model in two dimensions could possibly use only two parameters, height and width, as the major and minor axis of an ellipse.

ii. A relationship of the mouth width could predict the protrusion of the lips. Where the lip width reduces the lips become more protruded.

iii. The muscular actions for the production of vowels can be summarized as follows: The orbicularis oris has a draw string effect on the back rounded vowels. The pulling back of the lips are actions of the zygomaticus and risorius muscles. Raising and lowering the jaw has the largest effect on the front vowels.

5.5.2 Brooke and Summerfield's Speech Synthesizer.

A method for the synthesis of visible articulatory gestures has been developed by Brooke and Summerfield [18]. This was a computer-controlled graphical display of a talking face in two dimensions, where nine VCV (vowel, consonant, vowel) syllables were recorded onto a U-matic tape for subsequent analysis of the perception of visible articulatory gestures. The results they obtained are of particular interest, as they reveal quantitative measurements about the extremities the mouth and lips as they articulate.

Recording was achieved by seating the subject facing the camera at a distance of three meters with a plane mirror located adjacent to the subject's head with its horizontal axis at 45 degrees to the principle optical axis. Three facial reference points (on the nose, temple and check bone) and twelve articulatory points were marked onto the subject's lips such the they were visible to the camera so as to allow plausible representations of spread, rounded and protruded lips. The twelve articulatory points measured by Brooke and Summerfield are;

1. Horizontal extremity of the jaw
2. Center of upper outer lip margin
3. Tip of "Cupid's Bow" on upper, outer lip margin
4. Intermediate point on upper, outer lip margin
5. Corner of upper, outer lip margin
6. Center of lower, outer lip margin
7. Intermediate point on lower, outer lip margin
8. Corner of lower, outer lip margin
9. Center of upper, inner lip margin
10. Center of lower, inner lip margin
11. Corner of upper, inner lip margin
12. Corner of lower, inner lip margin [19].

The nine VCV analyzed were of the type /aCV/, where C was one of /m/, /b/, and /p/ and V was one of /a/, /i/, and /u/.

The three vowels are generally classified as producing maximal jaw opening, lip spreading, and lip rounding. For example, the production of /aba/ over a time span of 800 ms, the vertical displacement of point five (corner of upper, outer lip margin) was in the region of 7 mm - 8 mm, while the displacement of point ten (center of lower, inner lip margin) ranged from 17 mm - 21 mm with a slightly larger displacement for point 1 (horizontal extremity of the jaw). An approximate 2 mm horizontal displacement of point five (corner of upper, outer lip margin) was also recorded.

Some important observations of results from their investigations have been extracted:

i. Overall excursions of all the nodes rarely exceed 25mm.

ii. The corner of the lip margins have an approximately 50% motion in the vertical plane during open rounded articulations.

iii. Despite the rapidity of the mouth articulations they demonstrate near sinusoidal motions. This correlates with the work of Keslo et al where LED's (light-emitting diodes) were placed on the surface the speakers while producing /ba/ or /ma/ syllables. The resultant articulator trajectories over time are implied to be functionally equivalent to those of a mechanical mass-spring system [20]. Consequently, repetitive motions of the lips display a near sinusoidal steady-state motion as if created by a simpler non-dissapiative mass-spring dynamics. Other researchers have reported similar biomechanics of the lips and tongue [21].

iv. There is a decrease (overall 56% - 51%) in the lip-reading of vowels with the removal of the teeth in the synthetic two-dimensional model. This is more pronounced with those vowels, whose production requires the teeth to be visible.

5.6 Introduction to Investigation 1 to 4.

The Brooke and Summerfield synthesizer and the investigations of Fromkin, were primarily concerned with an analysis and generation of two-dimensional frontal displays of the face. Their investigations and results provide an indication of how speech can be synchronized, but this present research is concerned with the three-dimensional facial articulations where not only are there differences in the perception of the visible articulatory gestures, but also in the computer generation of synthetic facial images.

The investigations in this chapter are chronologically organized and assume that speech sequences comprise of a multitude of mouth and lip shapes that can be edited together to synchronize with the spoken track. Each investigation involved a different technique which resulted in an animation sequence that can be analysed.

Investigation 1 directly employs phonetic the shape of the mouth and lips in an attempt to comprehend the algorithmic problems of animated speech. Investigation 2 uses speech mapping to overcome the problems encountered in Investigation 1 and Investigations 3 and 4 employ the muscle model process, demonstrating firstly that lip shapes can be generated algorithmically in two dimensions, and secondly, that these can be extended to three dimensions.

5.6.1 Investigation 1: Phonetic Guides.

The aim of Investigation 1 was to synthesize in three dimensions a sentence of fluent speech, using phonetic keys. The primary objective was to comprehend, establish and observe a process that would cope with the complexity of the image synthesis. The sentence 'And now for something completely different', was used as the first test sequence.

In three dimensions the phonetic position poses of the face were constructed. The fundamental set consisted of the nine three-dimensional phonetics; /a/, /b/, /f/, /s/, /c/, /e/, /m/, /o/ and /l/ (Appendix A.3). Of these phonemes /s/ and /c/ double up, since they are based on the static phonetic positions. The phonetics were grouped in the same way used by the traditional animators such that the sets: /m, b, d, p/, /o, u, w/, /f, v/, /e, i/, /s, t, z/ were the fundamental divisions.

The sound track was recorded on U-matic video tape, key mouth phonetics were manually searched for and the frame was time recorded on a exposure sheet.

Interpolation was specified frame times and between two phonetic faces created by a function Speak (see Appendix A.4). This function employed cosine interpolation to establish the intermediate frames. The whole process of lip synchronization was carried out in a non real-time environment. Static frames were computed and sequentially recorded onto three quarter inch video tape. Typically, a wireframe line test of just a few seconds animation would take fifteen to twenty minutes to record, and a fully rendered animation sequence five to seven hours. The resultant mute animation can be seen in video sequence V.6.

5.6.2 Results and Discussion.

As expected, this first attempt at speech synchronization was unsatisfactory. The mouth only partially synchronized to the audio, and the face appeared to be over-articulating. However important lessons were learned. The three primary causes for the poor animation were established. Firstly, the three-dimensional phonetic keys digitized were from static poses, and fluent speech does not comprise complete phonetic shapes, but a blend between one mouth shape and the next. Secondly, the keys created were over-stressed by the subject, causing the face to over-articulate during the sequence. Thirdly, as the image fidelity of the face employed was high, an observer inevitably made comparisons with real people speaking and could not accept the shortcomings of the synchronization.

This failure to achieve synchronization using phonetic structures meant that further analysis of real people speaking was required, and a new model for the face that did not attempt to model human characteristics had to be developed.

5.6.3 Investigation 2: Speech Mapping.

It was evident from Investigation 1 that the phonetics of speech could not approach synchronization with facial images that approach human proportions. Therefore the exaggerated characters of Margaret Thatcher and the Queen from the Spitting Image puppet makers were used. Speech of a real person was recorded onto U-matic tape from the frontal view only, and it was assumed from the investigation of Fromkin that the mouth width, height and area were paramount in the perception of visible articulatory gestures [22]. An actor was placed approximately two meters from the camera while a close-up of the mouth was recorded as he recited: 'And now for something completely different' and 'Would the right honorable gentleman please shut up!'.

For each segment mouth shapes were traced from the video screen at each key position of the mouth and the frame time noted onto an exposure sheet. Each key mouth shape was then digitized in two-dimensions and a mapping was created such that the model's mouth was mapped from two to three dimensions. To achieve the animation the key mouth shapes maps were interpolated using a cosine function. The video sequence V.7 and V.8 illustrate the speech mapping technique.

5.6.4 Results and Discussion.

The result of the animation in video sequences V.7 and V.8 was a great improvement over the previous attempts at three-dimensional phonetic key-framing. The reasons for this are: firstly, data was collected from real people speaking and, secondly, the models that were animated were exaggerations of human figures, and consequently one does not expect them the speak as a human might.

Speech mapping can be seen in conclusion as a powerful and accurate approach for the fundamental synchronization of the lips to an existing sound track. It does however, depend upon the individual model and requires each individual mouth shape to be created.

5.7 The Recording and Analysis of Fluent Speech.

It is evident that speech mapping is a successful technique for speech synchronization, but it requires the animator to manipulate the face data-structure, resulting in the hard-wiring of specified vertices. What is required for the next step is a more general approach to the geometric transformation of the mouth that does not rely upon the data-structure. Consequently it was necessary to make a closer analysis of fluent speech before proceeding to Investigation 3.

The previous research of Fromkin and Brooke presented the phonetic vowels principally for a two-dimensional analysis [23]. What was further required was the development of synthetic mouth shapes in three-dimensions for fluent speech. Therefore it was necessary to record real people speaking so as to identify the displacement, spread and speed of articulations in three dimensions. Consequently, a similar approach was taken to Brooke and Summerfield and speech was recorded onto video tape for subsequent analysis.

A video camera was set up as in Figure 35 with the subject facing the camera and a mirror at forty-five degrees to the

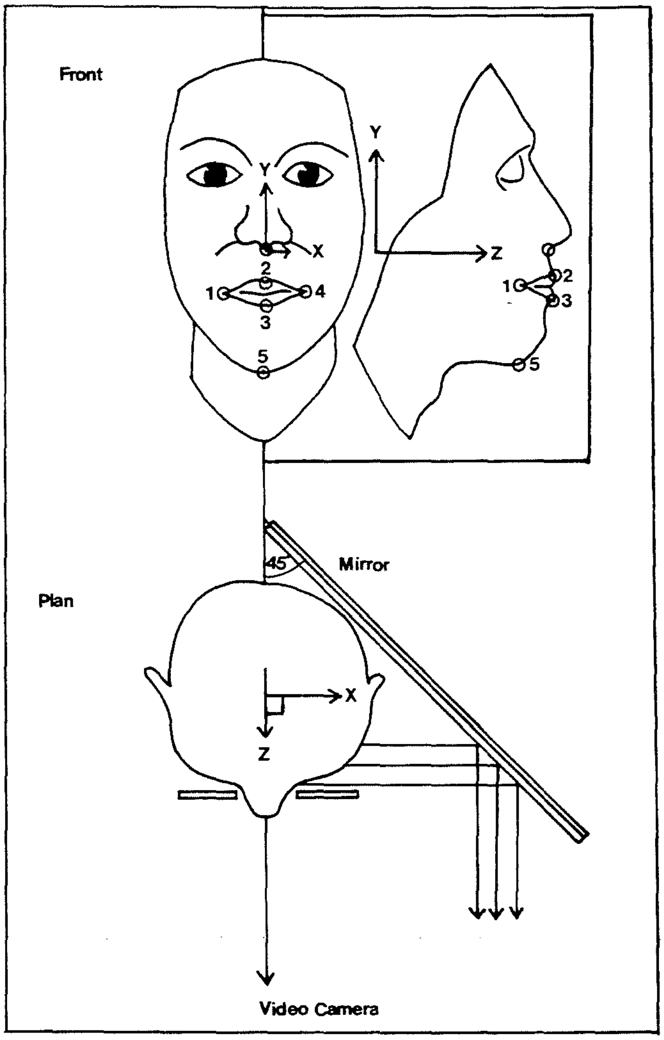


Figure 35

The video camera set-up in plan and frontal view. A plane mirror was positioned behind the subject at 45 degrees and the video camera positioned in front of the subject to capture both the x,y and y,z views simultaneously.

plane of the face, such that the lateral view of the face could be recorded at the same time. The subject wore glasses to align the face both horizontally and laterally, and these were used as a static location point against which it was possible to define the movements of the face.

The following poem was recited at normal speed in plain English:

> I wonder how many people in this city
> live in furnished rooms.
> Late at night when I look out at the buildings
> I swear I see a face in every window
> looking back at me,
> and when I turn away
> I wonder how many go back to their desks
> and write this down [24].

Three subjects were recorded reciting the same poem at twenty-five frames a second on three-quarter inch U-matic video material. Once recorded, the sequence was stepped through frame-by-frame, and the extremities of the mouth shapes were rotoscoped from a large flat screen (1.2m by 0.9m) while timings were recorded onto an exposure sheet. By scaling the images from the life model, and using reference points on the upper lip, lower lip, the chin, the corners of the mouth, and the base of the nose, displacement activities could be measured and correlated.

The teeth and tongue are an integral ingredient in the construction of sounds, but no attempt was made to record the motions of the tongue, as the complexities of its motions are, for the most part, obscured from direct scrutinization by the occlusion of the fatty tissue of the lips and the teeth.

5.7.1 Analysis of Articulatory Movement.

The following data of three principal locations, the center of the upper lip margin, the center lower lip outer margin and the right-hand corner of the lip were recorded. It was considered unnecessary to capture in every frame of the sequence only the extremities of the lip activity in relation to the frame time as the intermediate positions can be established by calculation.

The data in Appendix A.5 is three-dimensional unscaled displacements from a static rest position for the first sentence, 'I wonder how many people in this city'. The Figure 36(a) illustrates the scatter of the data in a lateral view, while Figure 36(b) illustrate the scatter of data from the frontal view. Figure 36(c) is a graph that illustrates the displacement activity of the top, mid and

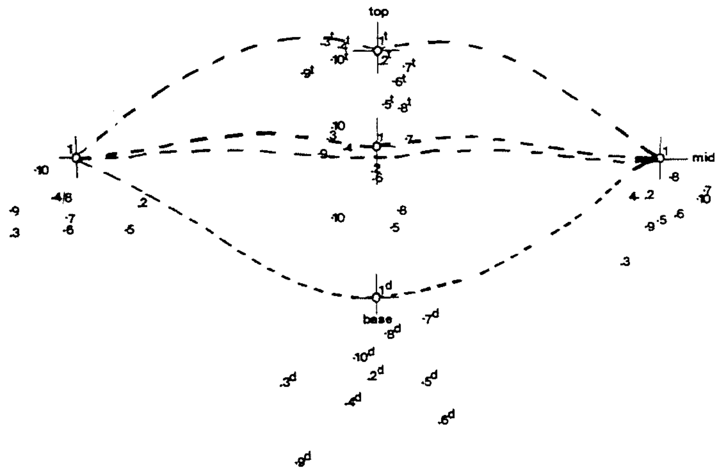


Figure 36(a)

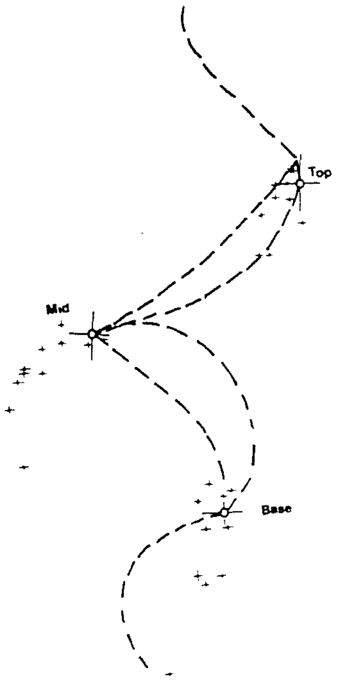


Figure 36(b)

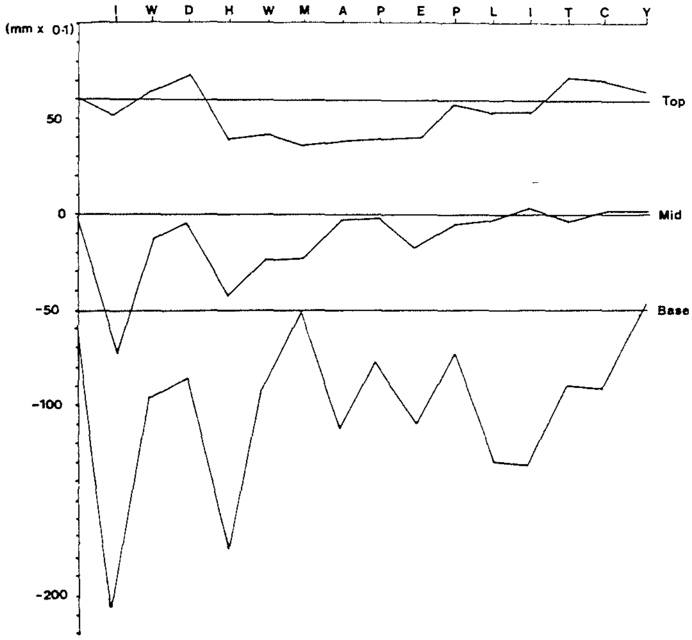


Figure 36(c)

Figure 36(a) shows the scatter of lip positions of the top lip, mid lip and bottom lip over ten key mouth positions in fluent speech and 36(b) demonstrates the scatter of the lip positions from a lateral view. 36(c) shows the displacement activity of the three lip positions - top, mid and base - from the frontal view of mouth shapes during fluent speech (see appendix A.5).

base lip positions from the frontal view. The following peculiarities are typical and are particulary difficult to measure:

i. The apparent volumetric changes of the lips can vary enormously as they squeeze and stretch.

ii. The changing shadows under the bottom lip and the upper lips shadowing into the mouth cavity that is influenced by the position of the tongue give a distinct characteristic of volume especially as the lips pout forward.

iii. The mouth and teeth are not symmetrical either in shape or motion.

It was evident from this brief analysis of articulatory movement that a mode of generating a wide range of mouth shapes was required if a general technique for three-dimensional computer synthesis was to be achieved.

5.8 Investigation 3: Geometric Manipulations of the Mouth.

The general manipulations of the face have been described by the muscle model process in section 4.3, and it was desirable to encapsulate the mouth articulations within the same strategy.

Before a three-dimensional model was undertaken, a preliminary model of the mouth was described in two-dimensions. The objective of the model was to determine geometric transformations required to distort an arbitrary mesh to any desired configuration in the same way that a real mouth can be distorted during speech. For the two-dimensional mouth the parameters were as follows:

Action	Variable	Value Range
Jaw rotation	Lip separation	$0.0 \leq value \leq 5.0$
Orbicularis Oris	Muscle squeeze	$0.0 \leq value \leq 1.0$
Left Zygomatic	Muscle tension	$0.0 \leq value \leq 1.0$
Right Zygomatic	Muscle tension	$0.0 \leq value \leq 1.0$
Left Depressor	Muscle tension	$0.0 \leq value \leq 1.0$
Right Depressor	Muscle tension	$0.0 \leq value \leq 1.0$

Table 6. Mouth Parameters.

5.8.1 Investigation 3: Results and Discussion.

Figure 37 illustrates the activity of the two-dimensional muscle model. With only the four muscles for the mouth and the assimilated jaw rotation, the phonetic structures could

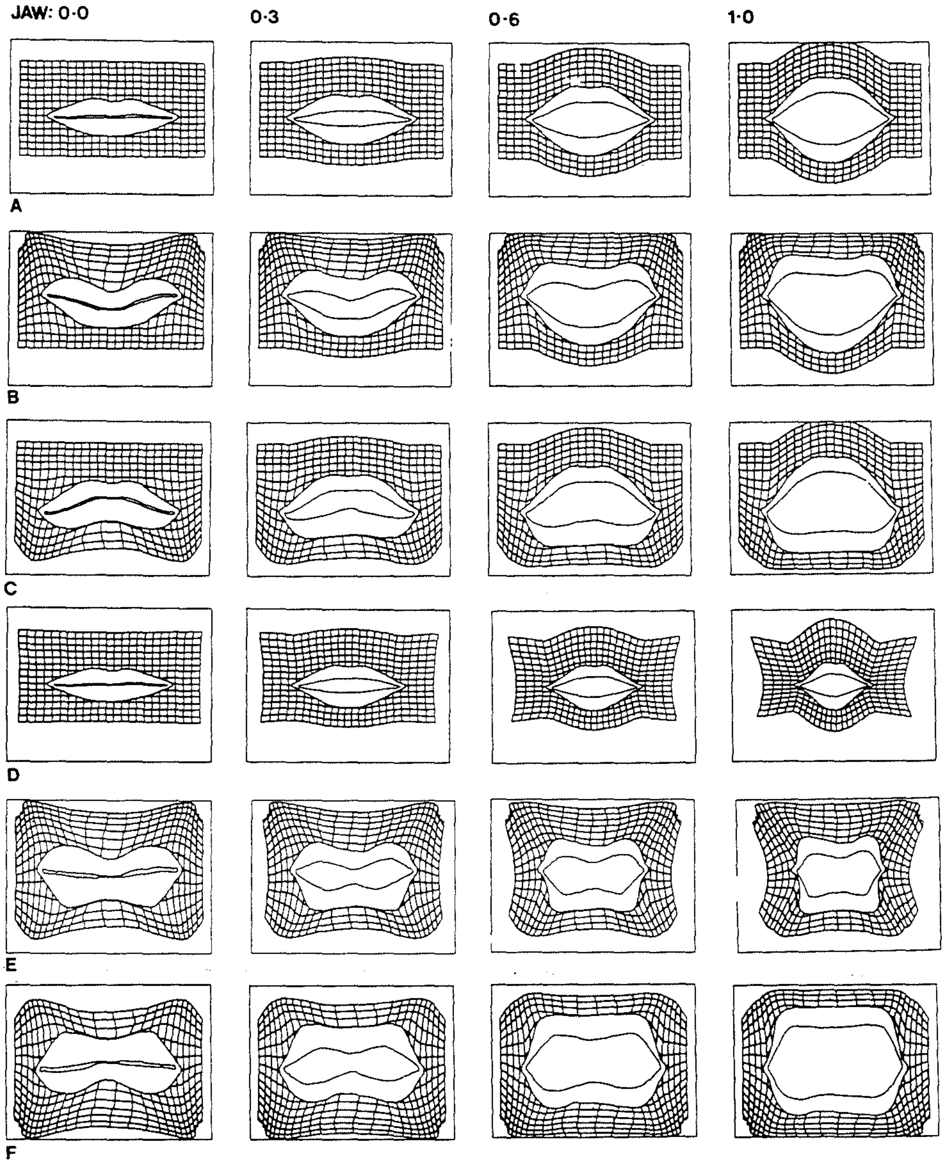


Figure 37

An illustration of the linear and sphincter muscle actions in combination with the jaw opening acting on an arbitrary two-dimensional mesh.

Row A	illustrates (left to right) the jaw opening to a value of 1.0.
Row B	the jaw opening in combination with the zygomatic major contraction.
Row C	the jaw opening in combination with the angular depressors.
Row D	the jaw opening in combination with the sphincter muscle.
Row E	the jaw opening in combination with the sphincter muscle and the zygomatic major and angular depressors.
Row F	the jaw opening in combination with the zygomatic major and angular depressors.

be constructed and this a powerful technique for the creation of the arbitrary mouth shapes found in speech. This method is parametric and incorporates the bio-mechanical approach of muscles, therefore avoiding the complexity of having to store explicit definitions and allowing a multiplicity of mouth shapes like those encountered in section 5.7.2.

Investigations by McGrath et al into the roles of the lips and teeth in lip-reading vowels indicate that the teeth are perceptually important but difficult to measure [25]. The removal of the teeth from the two-dimensional computer image resulted in a decrease in overall score of reading vowels from 56% to 51% [26]. While this is only a small detriment to the scores (within experimental discrepancy), it is evident that they are important in visual perception, therefore the model in Investigation 4 had teeth.

5.9 Investigation 4: Three-Dimensional Manipulations of the Mouth.

The geometric manipulations of the mouth in two dimensions provide good results, therefore Investigation 4 examined the application to three-dimensions.

One of the principal differences between two and three dimensions concerns the protrusion and extension of the lips. The orbicularis oris is the muscle that is responsible for this action, by the constriction and pouting of the lips. Consequently the squeezing action of the lips can be combined with the extension of the lips that is proportional to the width of the mouth. This is consistent with the data recorded by Fromkin and the recordings of the author.

The objective of investigation 4 was to synchronize the speech segment:

> That is the genius of the modern magic lantern,
> Your every wish is its command.
> Input the parameters, modify the algorithm,
> and you can be sure of getting what you like,
> in the style you like it,
> whenever you like it [27].

The first task involved isolating the key mouth shapes in the sequence as in section 5.7.1. The sequence was recorded of the actor speaking onto film rather than video tape as it is an easier and more accurate medium for isolating key actions. The sequence was then played and the extremities of the mouth shapes were traced in respect to the frame time on an exposure sheet. This resulted in seventy-five key positions for three-hundred and seventy-one frames running at twenty-four frames a second. This is a key position, on average, every five frames.

A computer-generated image of the face was then matched to the mouth shape drawn by the mouth opening, the contraction of the zygomatic major, the angular depressor, the sphincter and the risorius muscles. Data files were constructed for the individual muscles for the duration of the sequence such that they could operate independently. To create the intermediate positions of the lips interpolation was employed on the tension parameters of the muscles.

5.9.1 Investigation 4: Results and Discussion.

The video sequence V.9 illustrates the result of the process, and this technique is considered the most successful so far adopted. The success of the synchronization is due to the muscle contractions of the five major muscles being able to create arbitrary mouth shapes. It was found that mouth shapes did repeat, but that the context of these mouth shapes could not isolated.

5.10 Parametric Patches and Muscle Combinations.

The face can be delineated as a collection of curvilinear quadrilateral patches, where each patch can be completely specified by means of a mathematical formula. Bézier patches consist of sixteen control points usually arranged in a four-by-four grid [28]. 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. Figure 38(a) illustrates the patch boundaries and Figure 38(b) illustrates the mouth constructed from the Bézier parametric patches (see also Appendix A.6).

The main advantage of employing patches for the description of the mouth is that the deformation of the object can be more generally managed in comparison with the explicit facet manipulations. In addition to the muscle model process such that:

$$P[k]i,j' \propto f(K,A,R,P[k]i,j)$$

where:

P[k]i,j is the control points such that a set of P[k]i,j's describe the surface generated,

A is the angular displacement activity,

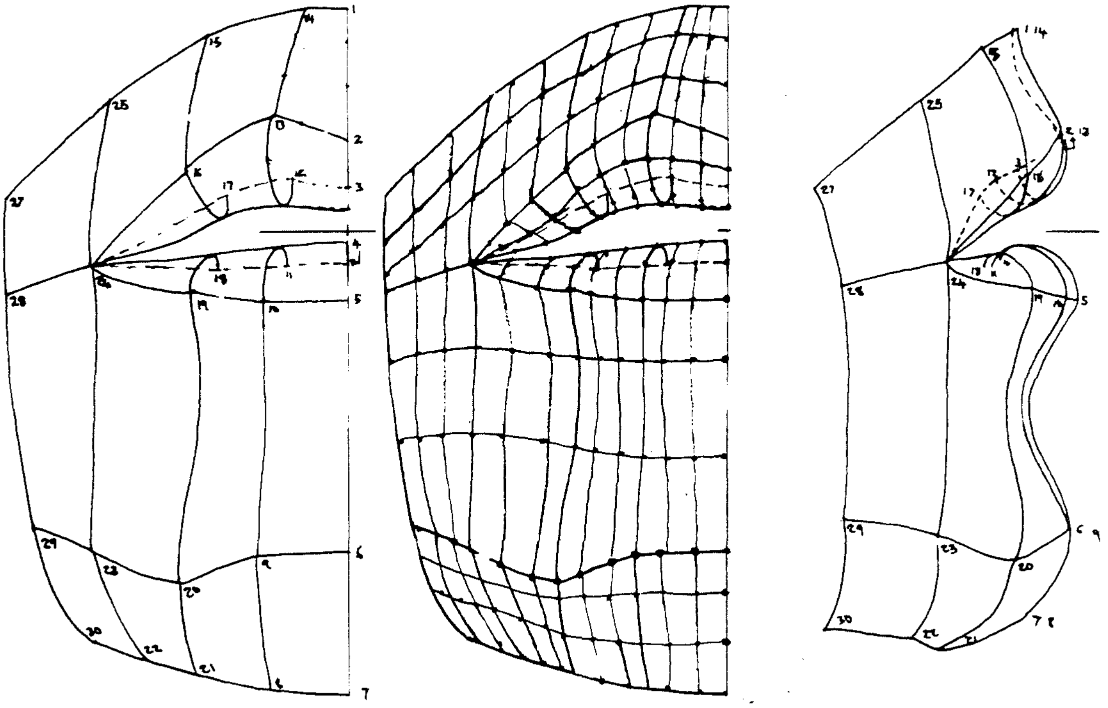


Figure 38 (a)

The Bézier patch boundaries for the mouth.

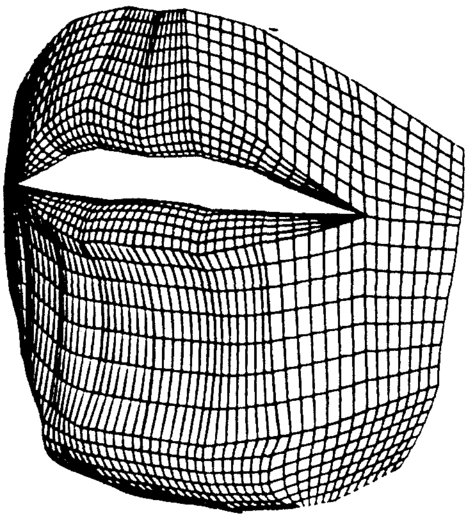


Figure 38 (b)

The three-dimensional facet topology resulting from the Bézier patch definition.

R is the radial displacement activity,

K is the muscle spring constant.

This assumes that each patch is treated independently, whereas patches are joined and the boundary conditions have to be considered. Subsequently the boundary condition is moved once only.

Figure 39 illustrates the interaction of a linear muscle and the Bézier patch. Figure 40 illustrates the deformation that can be achieved with the combination of the muscle model and the patch definition with the following muscle parameters:

Action	Variable	Value Range
Jaw rotation	Lip separation	$0.0 \le \text{value} \le 5.0$
Orbicularis Oris	Muscle squeeze	$0.0 \le \text{value} \le 1.0$
Left Zygomatic	Muscle tension	$0.0 \le \text{value} \le 1.0$
Right Zygomatic	Muscle tension	$0.0 \le \text{value} \le 1.0$
Left Depressor	Muscle tension	$0.0 \le \text{value} \le 1.0$
Right Depressor	Muscle tension	$0.0 \le \text{value} \le 1.0$

Table 7. Muscle parameters.

For these models lip pouting was created from the sphincter muscle that is consistent with the investigation of Fromkin. Using the patch definition for the mouth and controlling it with the muscle model process is a more coherent and

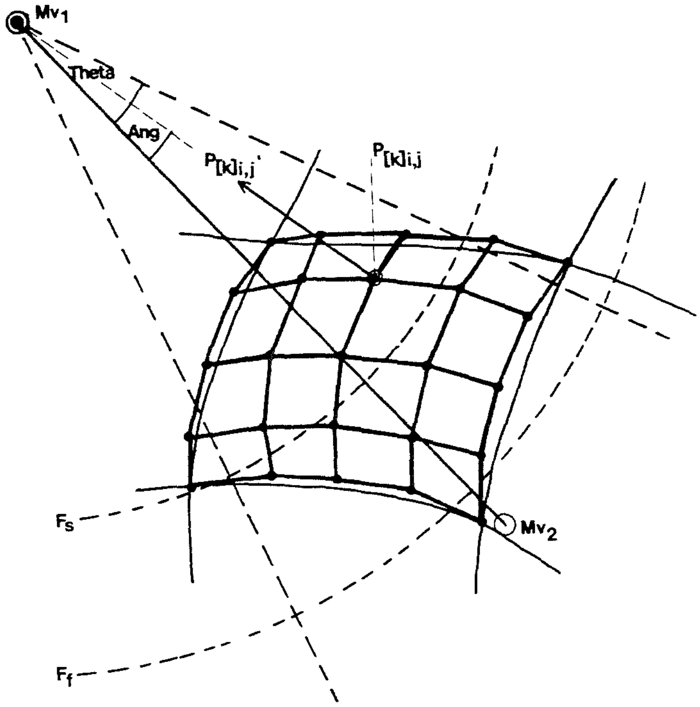


Figure 39

The linear muscle vector interacting with the Bézier patch boundaries.

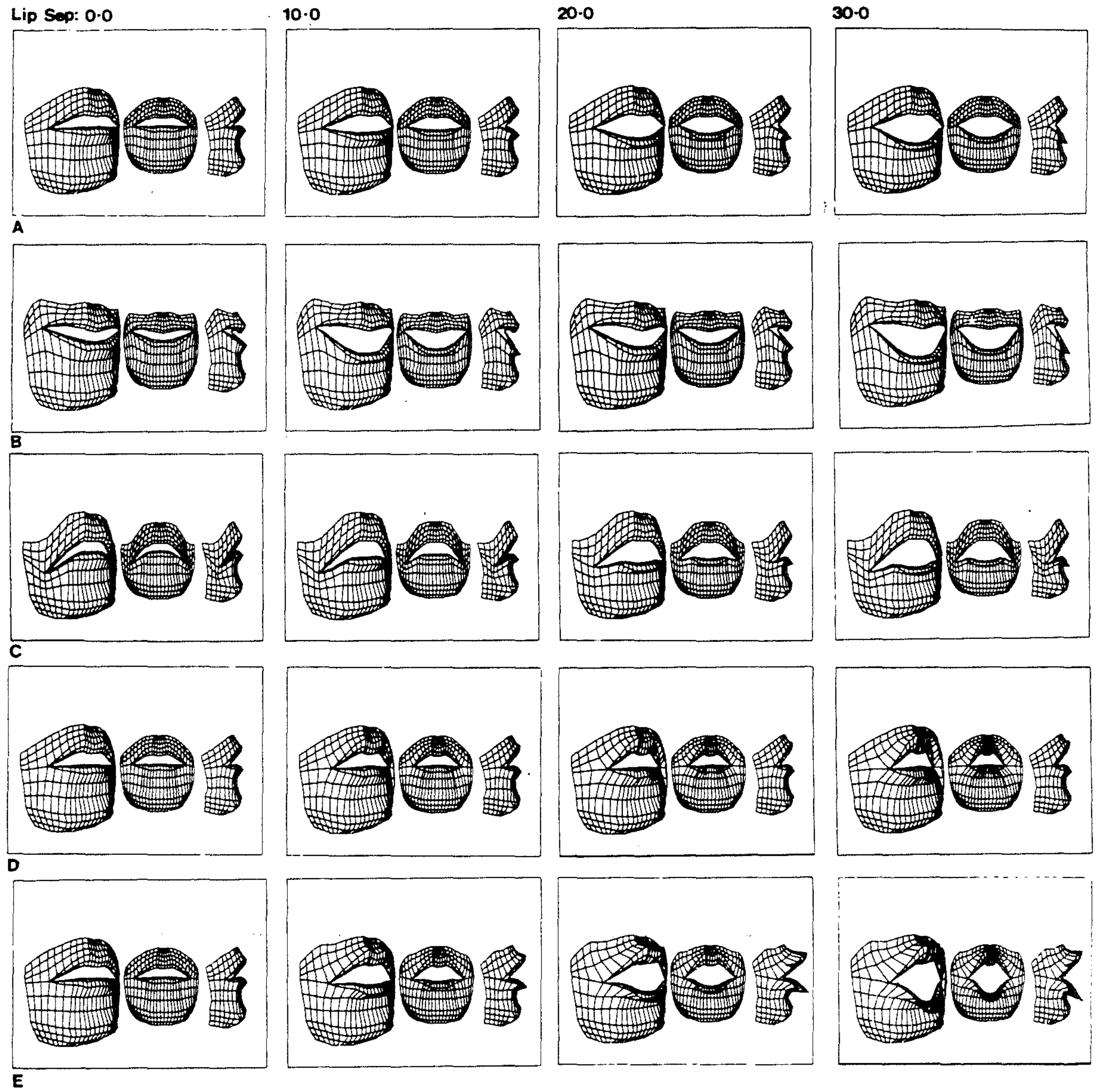


Figure 40

An illustration of the linear and sphincter muscles in combination with the lip separation acting on the Bézier patch mouth.

Row A	illustrates (left to right) the lip separation to a value of 30.0.
Row B	the lip seperation in combination with the zygomatic major contraction.
Row C	the lip seperation in combination with the angular depressors.
Row D	the lip seperation in combination with the sphincter muscle.
Row E	the lip seperation in combination with the sphincter muscle and the lip pouting.

This illustration corresponds to Figure 37, and demonstrates that a wide range of mouth configurations including those found in speech in three-dimensions can be created.

accurate modelling of the flexible surface of the skin. The facet definition blends under the distortion and can generate acurate deformations of the mouth during speech

5.11 Conclusions

Speech is an inherently multi-dimensional process. During speech many articulators are involved to a different degrees and the overlap of motion characteristics is considerable. Consequently, there appears little chance of identifying any fundamental regularity in the motions in terms of group phonetics, even though our perception of the situation leads us to believe differently.

As much as it would be useful to group actions together from phonetics, this can only provide a partial and impoverished solution to synchronization. The reasons for this are twofold. Firstly, phonetics are not created in fluent speech and, in addition, phonetics do not take into account intonation or stress within the speech segment. Secondly, there are perceptual problems due to the fact that as the image fidelity of the face increases, we are more and more pre-disposed to compare the synthetic images with real people speaking and are unable to accept a visual and audio dichotomy.

Another distinctive problem concerns the control of the articulators. Having to pre-record the visual image of someone talking before processing is not an economical process. It is not only tedious but also prone to errors especially for longer sequences. However basing the motion criteria on an understanding of the underlying muscles would in future models allow a more accurate lip synchronization to be produced provided information about individual muscle activity in terms of duration and intensity could be discerned.

It has been illustrated with the video sequence V.7, that lip synchronization can be achieved with only a small amount of data. However, it can only reach a certain level of credibility so long as the emphasis is upon the matching of lips to an already existing sound track. It is suggested that superior results would be obtained by creating mouth shapes and sound together as one coherent unit. This might take into account the internal organs of the mouth such that lip synchronization is no longer the required and the concentration is on the speech production from the shape and position of the lips, tongue and teeth.

5.12 References

[1] Paul Ekman, Emotion in the Human Face, Cambridge University Press, London, 1972.

[2] Ibid.

[3] Destandes, J., Histoire Comparée du Cinéma, Vol. 1, Castarman, Paris, 1966, pp. 166-167.

[4] Daniel Jones, An Outline of English Phonetics, Heffer, Cambridge & London, 1918.

[5] Matthew McGrawth, 'An Examination of Cues for Visual and Audio-Visual Speech Perception using Natural and Computer-Generated Faces', P.h.D Thesis, University of Nottingham, 1985.

[6] Ibid., pp. 8-9.

[7] D.B.Pisoni, 'Speech Perception: Some new Directions in Research and Theory', Journal of Acoustics, Society of America, Vol. 78, 1974, pp. 381-388.

[8] Frank Thomas and Ollie Johnston Disney Animation: The Illusion of Life Abbeville Press, New York, 1984, pp. 291-322.
Harold Whitaker and John Halas, Timing for Animation, Focal Press, London, 1982, pp. 122-128.

[9] Roy Madsen, Animated Film: Concepts, Methods, Uses, Interland Publishing, New York, 1969, pp. 120-122.

[10] Thomas, Disney, p. 321.

[11] L Hight, 'Lip-Reader Trainer : A Computer Program for the Hearing Impaired Proc of the John Hopkins First National Search for Applications of Personal Computing Aid the Handicapped', IEEE Computer Society, California, 1981, pp. 4-5.

[12] John Moore and Vincent O'Connor, 'Towards an Integrated Computer Package for Speech Therapy Training 'Talking Heads', Microtech Report, Bradford College of Art and Design, 1986.

[13] Phillipe Bergeron, 'Techniques for Animating Characters', Advanced Computer Animation Course Notes, ACM SIGGRAPH, Vol. 10, 1985, pp. 61-79.

[14] Fred I Parke, 'A Model for Human Faces that Allows Speech Synchronized Animation', Computer and Graphics Vol. 1, 1975, pp. 3-4.

[15] Quentin Summerfield, 'Audio-Visual Speech Perception, Lip-Reading and Artificial Stimulation', Hearing Science and Hearing Disorders, 1986, pp. 63-76.

[16] Ibid., p.132.

[17] Victoria Fromkin, 'Lip Positions in American English Vowels', Language and Speech, Vol. 7, 1964, pp. 215-225.

[18] Michael N Brooke and Quentin Summerfield, 'Analysis, Synthesis and Perception of Visible Articulatory Movements', Journal of Phonetics, Vol. 11, 1983, pp. 63-76.

[19] Ibid., p. 67.

[20] J A S Keslo, Eric Vatikiotis-Bateson, Elliot Saltzman and Bruce Kay, 'A Qualitative, Dynamic Analysis of Reitterant Speech Production: Phase Portraits, Kinematics and Dynamic Modelling' Journal of Acoustics, Vol. 77, No. 1, January, 1977, pp. 266-280.

[21] John W Folkins and James H Abbs, 'Lip and Jaw Motor Control during Speech: Response to Resistive Loading of the Jaw', Journal of Speech and Hearing Research, Vol. 18, 1975, pp. 207-220.
David J Ostry and Kevin G Munhall, 'Control Rate and Duration of Speech Movements', Journal of Acoustics, Society of America, Vol. 77, No. 2, February 1985, p. 641.

[22] Fromkin, 'Lip Positions', p. 217.

[23] Fromkin, 'Lip Positions', p. 217.
Brooke, 'Analysis', pp. 63-76.

[24] Leonard Cohen, Selected Poems, Cape, New York, 1969, p. 49.

[25] McGrawth, 'An Examination of Cues', pp. 135-147.

[26] Ibid., p. 140.

[27] British Broadcasting Corporation (BBC2), 'Welcome to my World', 10.00pm, 17th March, 1987.

[28] I D Faux and M J Pratt, 'Computational Geometry for Design and Manufacture', Ellis Horwood, New York, 1979, pp. 129-135.

CHAPTER SIX

Chapter Six

The Implementation of Facial Animation

6.1 Introduction.

This Chapter describes the implementation of facial animation undertaken during this present research. Two techniques are described: the key frame and parametric approach.

The key frame approach entailed compiling a minimum set of features required to achieve the desired results [1]. On the other hand, the parametric approach employed the muscle model process, where the operational goals were more extensive and limited only by the number of muscles implanted into the face.

Apart from the surface topology of the face, additional elements were considered important features to construct. In the upper face the eyes, eyelids, brows and hair, and in the lower face the teeth and tongue. However, not all the features were constructed for every model.

6.2 The Construction of the Face.

For both key framing and parameterization construction of the face models were identical and are independent of the motion attributes. The surface topologies were delineated from a collection of quadrilateral facets constructed from stereophotogrammetry techniques 3.2.4. Figure 41 and Figure 42 are a typical x,y and x,z photographic pair.

Determining the facet topology for the flexible surface of the face was a complex task. General principles were employed to define the most economical topology, and these rationales are summarized in section 3.3. Each stereo view was then digitized in two-dimensions. An algorithm (MIX: see Appendix A.3) then transposed the z-component from the second view into the first, creating a three-dimensional structure of the face. Subsequently, a data file was written holding pointers for each facet definition. This was in the form of N = ((x1, y1, z1) .... (xn, yn, zn)) representing the vertex list, and the polygons defined from the list of pointers in the form P = (1, 3, 8, 12). Once half the face had been constructed, the face was assumed symmetrical and mirrored about the vertical meridian of the model to form a whole facial model.

At this phase the model was rendered to examine the model for artifacts and omissions in the facet list. Importantly it was necessary to observe the rendering characteristics across the face to see if it displayed an accurate

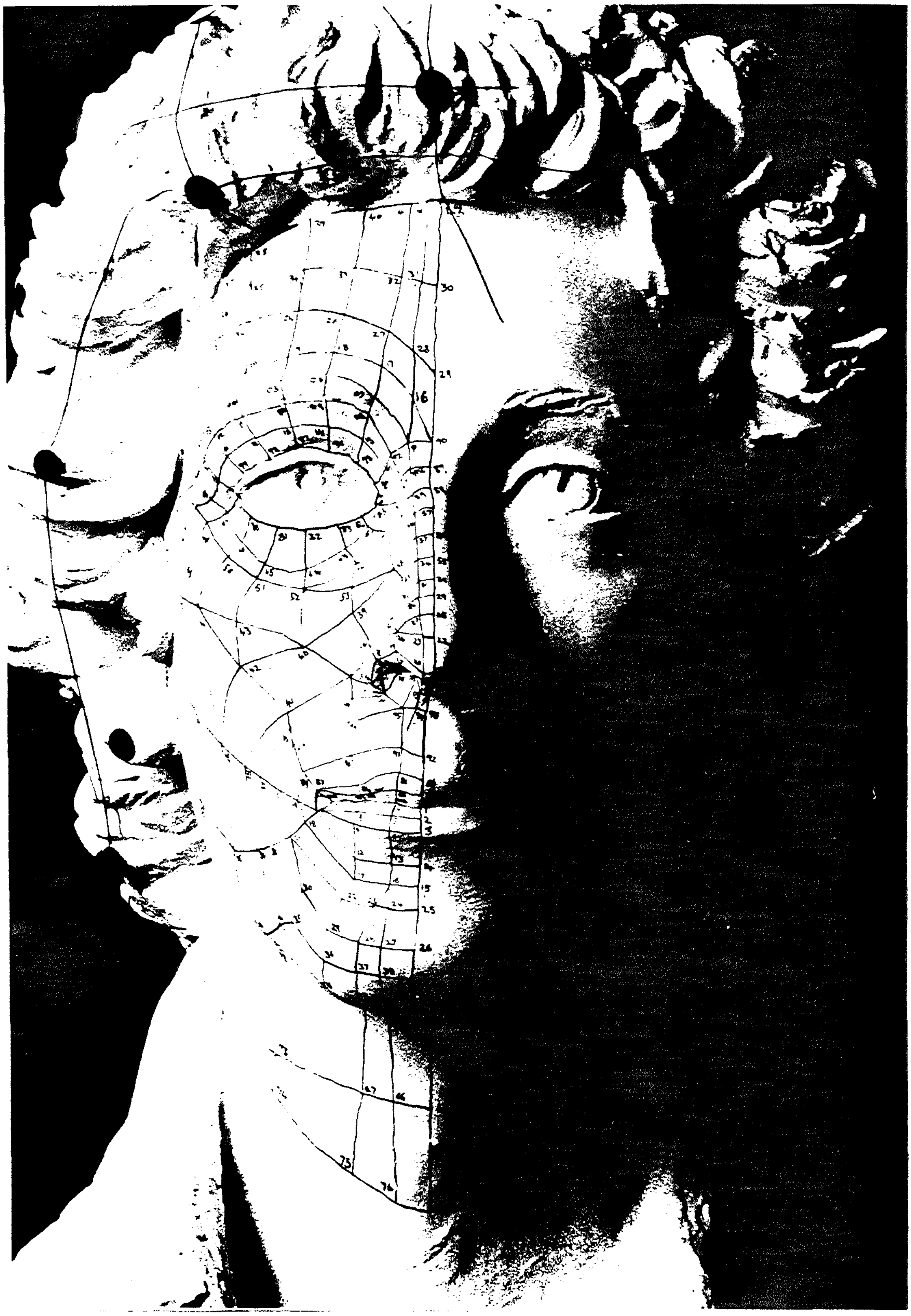


Figure 41

X,Y view of the facial topology.

Figure 42

X,Z view of the facial topology.

resemblance of the face. For example, in the locality of the nostril wing, facet data became complex and often confused, and therefore means by which these vertices could be modified were required. Rather than re-digitizing the face, a rudimentary vertex editor was written to allow the successive adjustment of the original two-dimensional data until the desired facial image was constructed in three-dimensions (see Appendix A.7).

6.2.1 The Upper Face.

The upper face in each model consisted of the eyes, eyelids, eyebrows. The eyeballs were initially developed as data-structures that were manipulated by the program, however this requires additional data to be stored onto disk. Therefore a procedure was developed to construct the eyeball from a faceted hemisphere. The controls for the structure permitted the radius of the eyeball, the radius of the iris ring and the radius of the pupil to be modified. As a result, the eyes and pupils could be dilated in the animation. Positioning the eyes into the head was achieved by measuring the x,y,z displacement from the original head and a focusing function permitted the eyes to track the same point in space so as to avoid the eyes staring outward.

Difficulties were encountered with the construction of the eyelids and their positioning within the face data. Three successive techniques were exploited. Firstly, it was assumed that because the eyeball was a sphere, the eyelids could be wrapped around the spheroid to produce a good fit. In this approach it is evident that the eyelids taper at the corner of the eyes blending into the structure of the face. Consequently this approach was considered to be a poor approximation and was not used. The second technique extracted existing data from the facial topology and duplicated this a number of times to form the upper and lower lids. Unfortunately this was also a poor approximation for eyelids, as they did not necessarily meet when the eyes were closed. Finally the third technique involved the explicit definition of the eyelids into four basic components constructed from Bézier curves; the upper lid consisting of one static curve that is knitted back into the structure of the face, a mid lid position that maintains the shape of the eye during opening and closing by concertinering, and the two curves for the lower part of the top lid that make up the underside of the top lid. This technique was the best approximation to the process involved and was adopted for a majority of the facial models.

The eyebrows were treated as a special case in the key framing approach, such that specified facets were extracted and displaced and modified to create the desired shape.

The hair was not considered in this research to warrant particular investigation, but it was necessary to give the impression of hair to increase the realism of the images. The hair was treated as a solid volume and was constructed as such to give the impression of hair. No complexity was involved to grow hair on the head as this would greatly increase the rendering times involved.

6.2.2 The Lower Face.

The stereophotogrammetric techniques captured the data of the lower face, but it was necessary to describe the action of the jaw independently. The jaw rotation was introduced as a rotation axis from which those vertices of the jaw could be rotated. It was found that square mouth shapes were created by the rotation about this axis, and consequently the corner of the lips were tapered so that they were only displaced by half the angle of the opening.

This resulted in superior results and was consistent with measurements of real faces articulating (see section 5.5).

As the jaw rotates so do the teeth. The teeth were constructed from three Bezier curves that generally follow the curvature of the mandible, and an upper and lower set were positioned within the mouth cavity such that the lower set rotated with the action of the jaw.

6.3 Controlling the Key Frame and Parametric Models.

Both the key frame and parametric models required the control of the time variant, or frame-to-frame, characteristics for the production of animation. The animation sequences were generated by recording a slightly different facial distortion for each frame, such that when they are run together they give the appearance of smooth movement.

6.3.1 The Key Frame Approach.

Shape interpolation techniques are at the heart of the key frame approach for facial animation (see section 1.2.1).

The face, like all other body parts, has inertia when moved, and therefore it is necessary to 'slow in' and 'slow out' all movements in the face. The cosine interpolants (see Appendix A.1) approximate this activity and were employed in the face shape interpolation technique. As a result only the key frames were constructed and the remaining face positions calculated from interpolation. The remaining characteristics of the eyeball focusing, brow separation and raising, and eyelids were controlled from data files. Figure 43 illustrates how the features of the face are drawn together in the system.

The video sequence V.6 is an example of the whole face interpolation, while V.7 and V.8 interpolate only the mouth shape positions in the speech map technique.

6.3.2 The Parametric Approach.

The animation in the parametric approach is a technique of modifying parameters over time. The difficulty lies in determining how the parameters are to change over time in order to create the desired effect. Furthermore, it is usually necessary to vary a large number of parameters simultaneously. Once the data input files have constructed

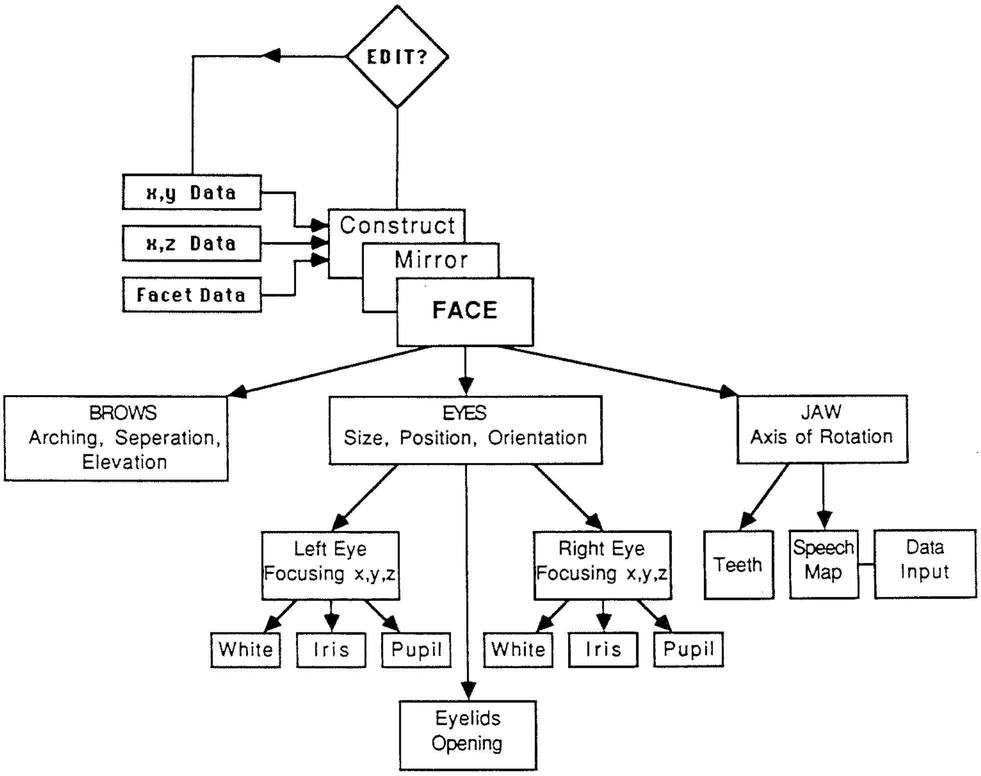


Figure 43

Key frame approach to the modelling and control of the face.

the face, the software sequence begins by calculating the parameter values for each feature for that particular frame. These parameters are then passed onto the muscles that contract in accordance with the parameters supplied to them. When the muscles are contracted, the data-base is updated to its new position. This is subsequently displayed and recorded, and then the process is repeated until the animation is complete, this procedure is illustrated in Figure 44.

To achieve this form of parallelism, algorithms were developed such that for each parameter specification an input data file was created that specified the start frame and finish frame and the value the parameter was to achieve. Figure 45 illustrates the sequence of events where time is displayed as a continuous line on the left. For any particular moment in time any number of parameter files may be active, and each of these possess any number of phases that consist of a start, finish and value. Provided the frame time is known, the current value for that parameter file can be calculated. The result is that many data files can be initiated with information about parameter values and frame times, and however many there are, a set of parameters can be calculated as a snap shot for a particular moment in time.

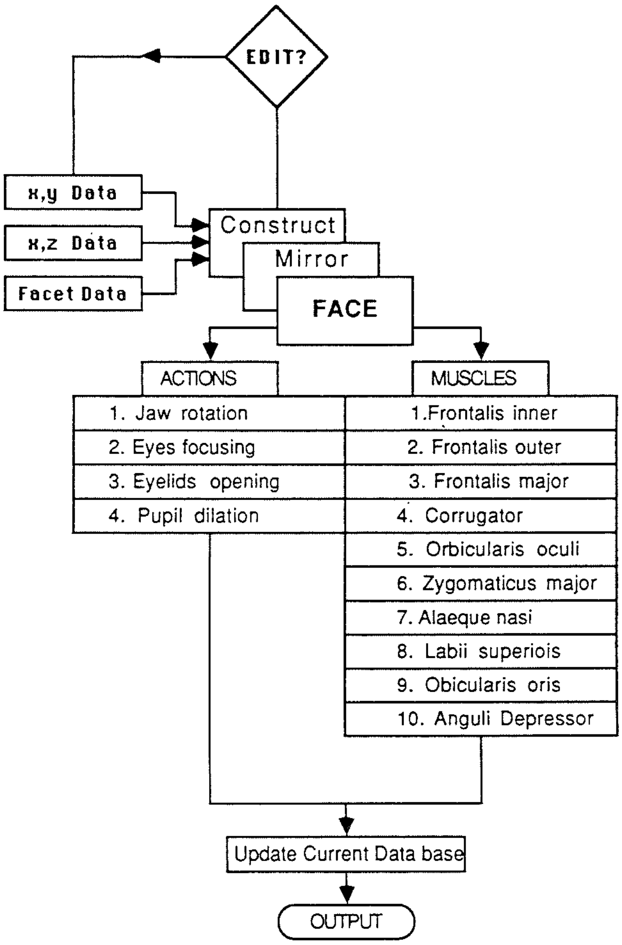


Figure 44

Parameteric approach to the modelling and control of the face with muscles.

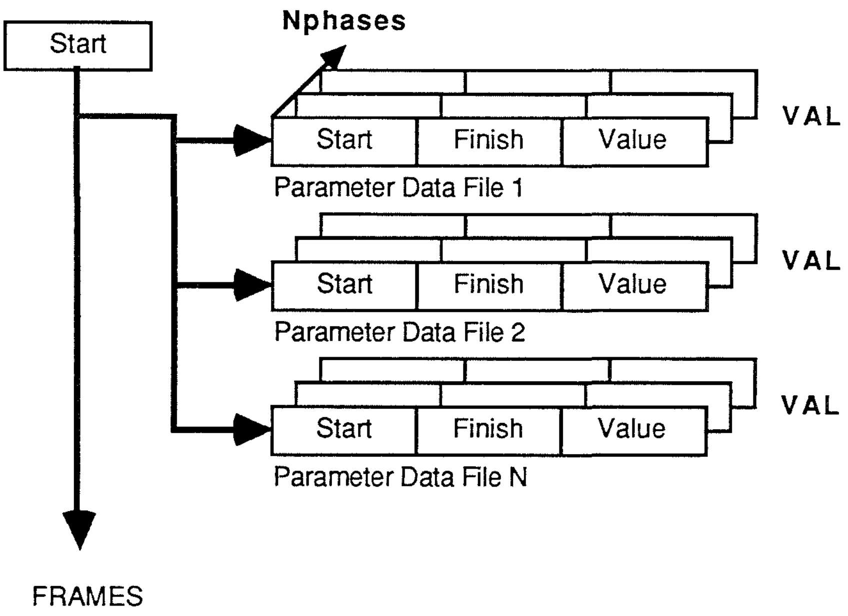


Figure 45

An illustration of parameteric parallelism.

Calculating the parameter values for the intermediate frames can be either linear or non-linear. Many of the parameters varied over short time spans: for example, the blinking of the eyelid can occur within a single frame or the mouth can change shape within two frames. As a result it is possible in these instances to use linear interpolation between the specified parameters. But it was assumed that the physical properties of flesh possess inertia under acceleration or deceleration, and consequently a cosine approximation was used to establish the intermediate parameter values.

The video sequences V.9 and V.10 are examples of a fully parameterized model and illustrate muscle combinations activating to create the speech synchronization and facial expressions. For this model a list of muscles were positioned and mirrored to create seven muscle pairs:

i. Frontalis Major
ii. Frontalis Inner
iii. Frontalis Outer
iv. Corrugator
v. Zygomatic Major
vi. Anguli Depressor
vii. Alaeque nasi Labii superioris

Four additional features were:

i. The jaw rotation angle
ii. The eyelids rotation angle
iii. Pupil dilation
iv. Eye focusing

In the sequence V.9 the muscles were paired through-out the animation, whereas sequence V.10 treated each side of the face independently and created fourteen independent muscles. This greatly enhanced the expression capabilities, as the motion characteristics of the face are not symmetrical.

6.4 Software and Rendering.

For the static frames illustrated in this thesis a Phong shader was utilized from the PRISM system [2]. Although the Phong renderer produces a 'plastic' effect, it was employed because it was possible to create more contrast using a low specular component than Gouraud shading.

For the sequences demonstrating the muscle actions in monochrome, a rendering algorithm was developed to visualize the images as Gouraud shaded images [3]. For this a new data structure had to be employed such that both average surface normals at vertices and an intensity for the incident light sources could be calculated. The data structure was in the form:

Vxtn, Sx, Sy, A, B, C, nF, Fn1 .. Fnn.

Such that a vertex (Vxtn), was stored with its associated screen co-ordinates (Sx, Sy), the vertex normal coefficients A,B,C, the number of facets coincident with the vertex (nF) and finally the list of those facet numbers (Fn1...Fnn).

From this information the grey scale intensity was calculated for each vertex to lie within the range 0 to 255 using Lambert's cosine law of illumination:

$$\text{Vtx\_Value} = L.\text{Cos}(\text{Theta}).C + \text{Ambient}$$

where:

L	is the light source component,
Theta	is the angle created between the light source and the facet normal,

C is the surface coefficients and the ambient is the background.

The corresponding grey scale components for a triangle and the screen co-ordinates were then passed onto the micro coded display algorithms that required the three x,y vertex screen co-ordinates and their associated grey scale level in the range 0-255. A total of forty-two (seven by six) frames could then be stored at quarter full screen size (1024 * 1279) within the eight bit frame store and cycled back at twenty-five frames a second to give the impression of real-time.

6.5 Conclusions.

The techniques employed in this present research provide techniques for the generation and animation of three-dimensional facial expression and speech. The key frame techniques are considered to be by this present research to be a secondary approach to the parametric techniques. However there are areas that require further attention to enhance the animation and visual realism.

i. The control of multiple data files is at best confusing to manage, therefore, a more fundamental control mechanism, that is sympathetic to bio-mechanical structure of the face is required.

i. The generation and display of real-time wire-frame faces for parametric control would allow more accurate definitions of expressions over time. The present research techniques were unable to generate, or modify, the models in real-time, consequently animations had to be clearly defined before the sequence was recorded on video tape frame-by-frame.

iii. The rendering algorithms give a poor illusion of the qualities displayed in real flesh. Further research in texture mapping techniques and procedural texturing could enhance the realism of the facial images.

Despite the short-comings of the rendering algorithms realistic facial images can be generated in three dimensions and accurate facial expressions can be created and speech animated.

6.6 References.

[1] Keith Waters, 'Expressive Three-Dimensional Facial Animation', Proceedings Computer Graphics Online, 1986, pp. 49-56.

[2] John A Vince, The Prism System, Middlesex Polytechnic Campus, 1986.
Bui-Tuong Phong, 'Illumination for Computer Generated Pictures', Communications of the ACM, Vol. 18, No. 6, 1975, pp. 311-317.

[3] H Gouraud, 'Continuous Shading of Curved Surfaces', IEEE Transactions on Computers, C-20, No. 6, 1971, pp. 623-62.

CHAPTER SEVEN

Chapter Seven

The Control of Facial Expressions

7.1 Introduction.

Complex articulations have been performed with the muscle model process, but, manipulating and controlling the volume of parameters required in the animation sequences can be extensive, and the explicit control can be tedious and time-consuming. This chapter therefore, discusses and proposes an implementation that has been developed from the requirement to control computer-generated facial expressions.

Industrial robotics have been developing with various levels of operational control while computer animation techniques have been expanding with concurrent interests. Many of the tasks solved in robotics and three-dimensional computer animation concern the control of articulations and joint configurations of rigid structures, from which a great deal of information can be extracted for computer animation techniques.

The control of synthetic facial expressions is concerned with the articulation of a biological structure, and therefore a brief analysis of biological motor behaviour is required. Some attempts based upon behavioural models, have been made to control biological structures such as human locomotion and flock interaction, but for the most part there has been little evidence of computer animation based on the underlying biological motivators.

7.2 The Potential Pose Space and the Degrees of Freedom Problem.

If an articulated structure is considered as a collection of rigid links, the dilemma of controlling the structure is to generate appropriate values for the joint variables. As the number of links increases for a particular structure, so do the potential configurations for the joints: for a figure with n joints, a n-dimensional pose space can be created, where a coordinate axis is assigned for each of the degrees of freedom (DOF). This in turn produces n potential pose vectors which specify the potential joint configuration for the structure [1].

Quintessentially the DOF are regarded as links manoeuvering relative to one another via one to six translational or rotational latitudes. Considering rigid body segments, a total of six dimensions for the DOF can be described:

1 DOF	Pin Joint or sliding joint

2 DOF	Universal joint or a cylindrical joint
3 DOF	Ball in a socket
4 DOF	Cylinder on a plane
5 DOF	Ball on a plane
6 DOF	Flying object

The volume of data created by only a fundamental set of links with a few DOF can be observed in a simple system with an aggregate of twenty links and a modest pose space every two seconds. This results in the specification of thousands of values for even a few seconds of animation and explains why the management and control of a simple animated character (with five hundred or so joints) is extremely formidable [2].

The flesh of the face can be considered as a flexible mesh of points which move relative to each other behaving like a network of springs distorting under the influence of external or internal forces. Because the points are inexorably linked to one another, each point on the mesh possesses 3 DOF, consequently a surface with Cp control points has 3 * Cp DOF varying over time. This explains why most computer animation deals with rigid body segments [3].

The control of a structure in an animation sequence, or a task for a robot manipulator, requires the selection of particular sets of variables out of the immense potential pose space. As a result, many configurations will be redundant for a particular motion [4]. For example, there are numerous methods of picking an object up from the table, but only a small handful that appear to be correct or physically economic. Zeltzer describes a sequence such as picking up an object as a complete set of poses or a "hyper path" through three-dimensional space [5]. Describing these hyper paths is what motion control in robotics and computer animation has achieved by a number of different means, each tailored to solving a different task.

7.3 Motion Control Systems.

The objective of robotic motion control is to specify, via a series of programmed instructions, the position of an arm, and in particular the end tool tip in three-dimensional space [6]. Likewise, computer animation control is concerned with the articulation of synthetic three-dimensional structures in a three-dimensional space.

In robotic motion control five distinct levels have been outlined: hardware level or computer level, point-to-point, motion level, structured programming level and task-orientated level. For a more comprehensive survey and description of robotic programming languages see Synder [7]. In animation motion control, three levels have been established: interactive or guided level, scripted or animator level and the task or high level control [8].

The computer level and the point-to-point mode in robotic control can be compared with the techniques in animation at an interactive or guided level. The motion level and structured programming level in robotics can be compared with scripted or animator level, and the task-orientated level in robotics is comparable to the techniques at task level in animation.

7.3.1 The Computer Animated Guided Level and the Robotic Computer Level of Control.

The guided level in computer animation was the earliest and remains the most common mode of motion specification. The basic principles involve the explicit definition of positions from which interpolation can then be used to calculate the intermediate positions. Interpolation usually involves a non-linear function such as splining to establish either velocity (first derivative), or acceleration (second derivative). This applies to object interpolation where there is a one-to-one correspondence between the data points, and the intermediate frames are computed for each vertex (see 1.2.1).

Paths created by interpolation or splines can be created interactively by using dials or tablets, whereby each object is assigned a path along which it will move. This technique can involve multi-tracking, where a multitude of tracks can be specified and controlled, as in systems such as MUTAN (Multiple Track ANimator) and Twixt [9]. The advantages are that objects have transitions only at specified key-frames, and consequently each object becomes independent of other objects. As the animation progresses further tracks can be layered to avoid the "key-frame clicking" that occurs when simultaneous extremes coincide [10].

The comparative level of robotic control usually involves the physical structure of the robot. Low-level programming languages implement real-time functions that are often hard-wired for the task, so that one task only will be performed. Alternatively, motions can be described manually by leading the manipulator through the specified motion, such as IBM's Funky that records the joint angles and displacement for the activity [11]. These can then be used as interpolation points through which the manipulator must travel under programmatic control.

It is evident that in the robotic motion control the arm will operate only what it has been wired to do, or the task it has been led through. Similarly in computer animation control, the sequence will be performed exactly as specified, which is reasonable only if the environment is simple and contains no other interactions.

While this technique is useful for certain robotic environments and animation sequences, it becomes very demanding when motions for intricate mechanisms are created. Refining these techniques is an active area of research, but the major disadvantage remains that guiding systems lack any powerful abstraction techniques that allow the adaptive motion of complex environments [12]. Despite the inherent inflexibility of guiding, this system can provide the most economical mode of control.

7.3.2 Scripted and Programming Control Levels.

Scripting is more flexible than the key-frame approach and, in most instances, not as complex to handle as a full programming language [13]. Scripted systems have mechanisms for functional abstraction and structured programming provides an extensive repertoire based upon existing computer languages. The scripts in animation usually describe, both the models and the motion. The notated script can be edited and modified to create the desired effect, which are interpreted and frames are rendered by the animation system, thus isolating the animator from the intricate details of the software.

ANIMA II, developed at Ohio State University, reflect the desire to script animation [14]. Events were specified in scripts by the name of the object, the start and end frames and the position. This allowed the animator to animate multiple objects without being concerned with programmatic flow control while remaining isolated from the computer environment. GRAMPS (GRAphics for the Multi-Picture System) was a similar system to ANIMA II, but it was an interactive user-orientated graphics system [15]. ASAS (Animation with Scripts and Actors) is an interpreted language system and is a full programming language based on LISP, which allows all the programming features such as recussion and typed data structures [16].

Robotic motion scripting control, or the structured programming level, is usually based upon existing languages, such as PL/1 and ALGOL and exhibits the following characteristics: algorithmic branching, subroutines, sensors, parallel activities and coordinate frame descriptions.

In robotics and computer animation control two of the more important characteristics are the inclusion of subroutines, and simple branching. Subroutines allow modules or abstract functions to be built up performing specified tasks and the branching allows simple conditional statements to be performed.

7.3.3 The Task-Orientated or High Level Control.

A number of attempts have been made to supply the user with high level control in animation [17]. The concept of a high level control system means that the animator can outline the motion characteristics, and the system then fills in the details. As a result, the non-expert user may quickly establish complex articulations.

Some techniques used in the high level control include: parameterization, which involves the designating parameters whose values define the configuration or motion of the object modelled; finite state motion control which is appropriate for the description of repetitive or coordinated motion, such as the bio-engineering control of walking [18]; command libraries which store low-level motion descriptions under high-level control; and finally hierarchies, which allow progressively more detailed information to be created, each description feeding to successive lower levels, so that as one action takes place, it may also manipulate the lower leaves in the hierarchy. This is of obvious importance when modelling arms, legs, hands and heads on shoulders.

Robotic task orientated level utilizes low-level aids such as sensors and co-ordinate transforms, but conceals them from the user. AUTOPASS is a language proposal from IBM that has yet to be fully implemented [19]. The aim is to resemble the instructions given to a human assembly worker. Commands such as:

> Place Object One On Object Two.

The executing of this command requires finding Object 1 and Object 2, determining how to move towards Object 1 and how to manipulate it, remembering the position of Object 2, and finally placing Object 1 on Object 2. The execution of these simple commands requires a sophisticated world modelling system to keep track of all the objects. This involves both visual location and tactile sensors for help in the exact positioning of the manipulator in three-dimensional space.

The primary limitation of the task level control systems is that while they allow the user to specify general activities, they lack the capability of specific control, and separate levels of motor programs are required for each motion used. Consequently, only a small handful of systems have been developed.

7.4 Discussion.

The above-mentioned control mechanisms deal with the control of rigid articulated bodies with fixed joints. The face, and other flexible biological structures, do not comprise rigid segments, but a collection of interconnected nodes that flex and stretch under the influence of internal or external forces. Despite the differences between the motion characteristics of the rigid and the flexible structures, two levels of motion control have been employed in this present research for the face; firstly, the guided level that is the simplest mode of control and secondly, the abstract parameters that are scripted muscle contractions in terms of muscle tensions and muscle spring constants.

What is required for further development is a more appropriate mode of muscle control that stems from an understanding of biological behaviour, rather than a series of data files defining the motions as described in this present research.

7.5 Biological Reflex Control and Functional Synergies.

The human biological motor system has evolved into an extremely complex structure capable of manoeuvering multi-linked forms with many degrees of freedom, through dynamic environments. The conscious decision of motion has been described as being built upon a base of reflex processes, or low-level motor programs regulated by high level controllers [20]. These low level motor programs, or Local Motor Programs (LMP), have been identified to be common principles for the control of walking in animals and may be the only economical method of controlling structures with multiple degrees of freedom [21]. Zeltzer successfully based a system upon the creation of low level motor programs for the kinematic control of walking figures under high level control, such that it was possible to describe gaits from the command "walk" [22]. This has since been developed by Maciejewski and Girard with the PODA system [23]. Here the animator is supplied with the computational model which facilitates the direct control of the functional dependence between different parts of the figure. The strategy allows the interactive design of the motion at various levels, with a focused attention on the complex motor skills such as those found in dance and gymnastics.

Motor programs can be observed in a variety of human activities such as walking, grasping and smiling [24]. The prehensile activity of the hand is a good example of motor activity, where various grips can be described that have very distinctive characteristics [25]:

i. The power grip, where the fingers are flexed around the object, with a counter pressure from the thumb. The purpose of the grip is to hold an object firmly, either to be wielded as a tool or weapon by the whole hand, or to be worked upon by the other hand.

ii. The precision grip where, as the description indicates, the activity is precise, but the main source of activity is the small movement of the digits. The object is usually gripped between the tips of the fingers and the thumb, like holding a pen.

iii. The hook grip, which is used to suspend or pull on objects or to elevate the body in climbing. The fingers are looped around such things as handles, straps, cords, branches etc. The thumb may, or may not, be involved, as it is a grip for the transmission of forces and not for skilled manipulation.

The three grips described for the hand are the general details of how they can be formed. These are not precise definitions of the individual mechanisms, but the group action of the fingers and thumb orchestrated to perform particular acts. Controlling all the individual joint activities of the hand requires an enormous amount of information, and is considered by Evarts to be outside conscious control [26]. Consequently the human biological motor system is unlikely to store explicit definitions for these motions. It has been suggested that the human biological system maintains a library of Local Motor Programs that can be invoked in a variety of combinations dependent upon the situation and the activities, such as the gripping activity of the hand, are biologically "pre-programmed" so that once the activity is understood, variations on a theme can be created.

Lower motor programs (LMP) are directly analogous to computer functions that carry out pre-defined and explicit operations, and consequently an investigation in this present research was carried out into the algorithmic solutions to the prehensile activity of the hand to examine how the control of a biological structure could be specified [27]. The activities were decomposed into discrete independent lower motions of the digits and the events were considered skills performing tasks, such as;

>Grasp Object 'B' at x,y,z with grasp.h.

Figure 46 illustrates two gripping activates of the hand,

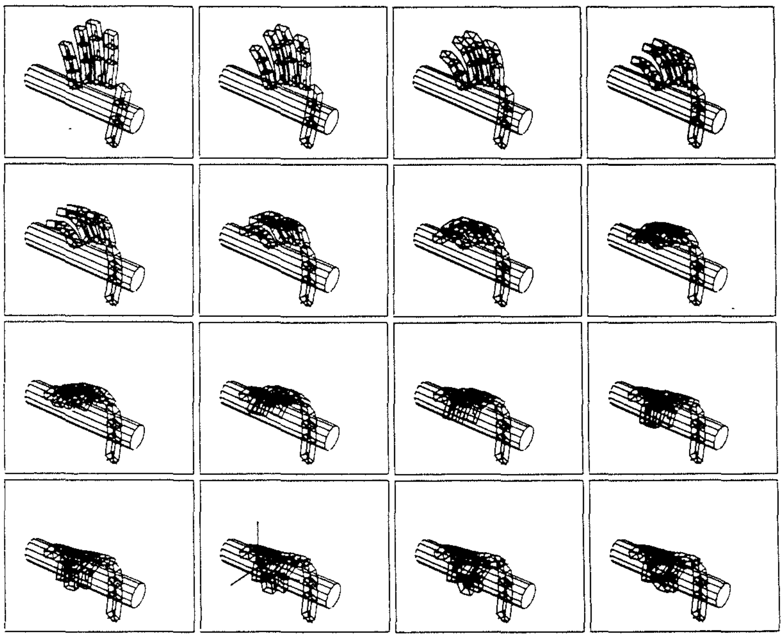


Figure 46(a)

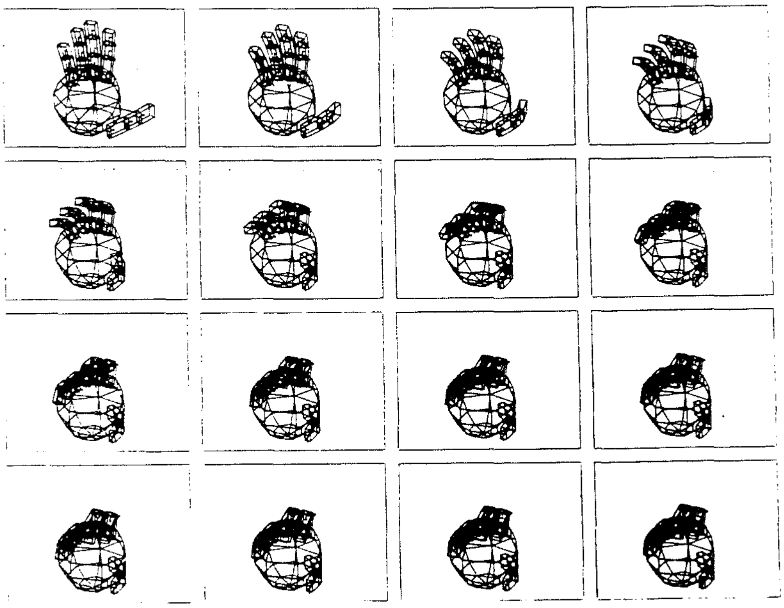


Figure 46(b)

Two examples of the gripping activity as functional synergies: the hook grasp and the power grasp. The hand comprise five fingers with three distal joints per finger. Collision testing prevents the hand from passing through the objects and the motional attributes for the fingers are setup in data files.

the hook and power grip, from the investigation. The conclusions indicate that the existing modes of motion specification are not appropriate for biological structures. Defining key positions in guiding system does not allow motion adaptation and cannot perform any other task. Parameter data files, while being appropriate for computer scripting motion specification of geometric objects, do not provide the animator with a comprehensible user interface for biological structures.

7.5.1 Functional Synergies for the Face.

The muscle model process with a limited number of muscles can generate recognizable facial expressions (see 4.6). The expressions of happiness, anger, fear, surprise and sadness and disgust therefore, are muscle contractions that could be clustered as functional synergies for computer synthesis. The aim would be to organize muscle subsystems that effect a particular class of motions recognizable as expressions, such that it would be possible to specify events as:

```
>Be sad.3.0s
>Be happy.0.5s
```

```
>Be really happy.1.0s
>Laugh.5.0s.
```

Figure 47 illustrates the proposed animation programmatic flow where a skill or task is invoked at a high level, this is then passed to the task manager (Animation Parser) that determines if it is a legal command, a physically possible command or a contradictory command. From such commands the correct motion motor descriptions are selected and pointers to the data base selected and manipulated in the modelling section. This is then dynamically tested and the data base updated and rendered.

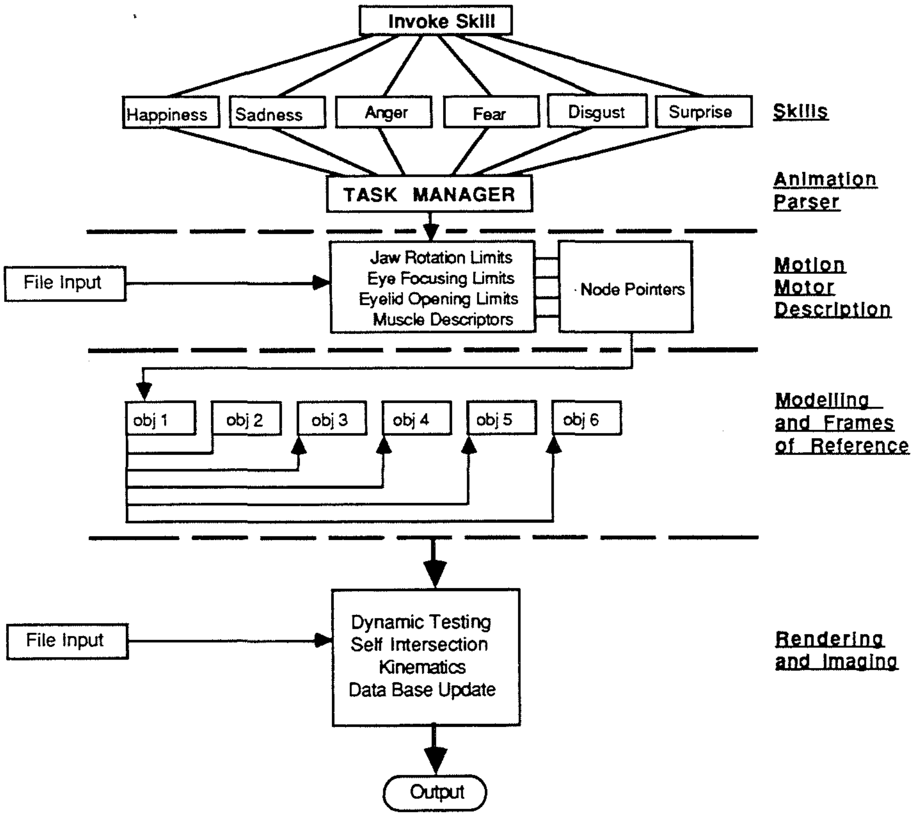


Figure 47

A illustration of the proposed animation program flow.

7.8 Conclusions.

Robotic motion control systems are analogous to the techniques developed in three-dimensional computer animation. They provide a large range of motion control modes and the techniques of guiding and scripting have been used in this present research as the means of facial motion specification.

Biological structures however, cannot be clearly and comprehensively encapsulated by these control modes. Biological structures do not comprise of discrete rigid links and are not easily understood by data files that control joint rotations, scaling and displacements. Decomposing the activity of a biological structure into discrete units such as local motor programs for the hand or function synergies for the face, provide a more comprehensive control mechanism for the animator and a description that can be understood by the animation system.

Describing expressions as skills is a powerful abstraction to the complex articulations because the low-level details of the execution can be suppressed. Parameterization has been demonstrated to be the most comprehensive and versatile mode of facial control. This has to be based upon a basic understanding of the cause motivators of actions. However, the suggested control environment for the face has three critical issues that need to be addressed, expression subtlety, blend and range;

i. To be able to create the subtlety of expressions, parameter values for the skills as synergies need to be specified. As a result 'verb specificity' becomes paramount in the description of the expression; what is the quantitative difference between 'happy' 'really happy' or 'very happy'?

ii Provided some defined criteria from psychology can be established for the 'verb specifity', the description of happy is as unique as sad. Many expressions are blends between two extremes or subtle variations. A possible solution allowing the creation of these blend parameters could be as follows:
NewBlend_Expression = Expression_5 + 33% of Expression_4

These definitions these would have to be carried out as a separate component in the animated system.

iii. The range of the expressions could be extended by the blends of other expressions, or new expressions formulated by the user. As with the blends they would have to be created as a separate component in the animation system.

Finally, decomposing and distributing complex problems is of a strategic advantage, that have mainly been born out of artifical inteligence research [28]. For the control of facial expressions it is also critical to break-down the facial actions in to manageable and discrete units that can be quantified and controlled.

7.9 References.

[1] David Zeltzer, 'Representation and Control of Three-Dimensional Computer Animated Figures', PhD Thesis, Ohio State University, 1984, p. 5.

[2] H Siegel, 'An Overview of Computer Animation and Modelling', Proceedings Computer Graphics, London, 1987, p. 35.

[3] Jane Wilhelms, 'Using Dynamic Analysis for Realistic Animation of Articulated Bodies', IEEE Computer Graphics and Applications, Vol. 7, No. 6, 1987, pp. 12-27.

[4] J Korien, 'Techniques for Generating the Goal-Directed Motion of Articulated Structures', IEEE Computer Graphics and Applications, Vol. 4, No. 9, November 1982, pp. 71-81.

[5] Zeltzer, 'Representation and Control', p. 6.

[6] P Paul, Robot Manipulators: Mathematics, Programming, and Control, MIT Press, London, 1981.

[7] W Synder, Industrial Robots: Computer Interfacing and Control, Prentice-Hall, International Editions, London, 1985.

[8] Jane Wilhelms, 'Towards Automatic Motion Control', IEEE Computer Graphics and Applications, Vol. 7, No. 4, 1987, pp. 11-22.

[9] D Fortin, J F Lamy and D Thalmann, 'A multiple Track Animation System', Proceedings ACM SIGGRAPH, 1983, pp. 180-186.
Julian Gomez, 'Twixt: A 3D Animation System', Eurographics, 1984, pp. 121-133.

[10] Frank Thomas and Ollie Johnston Disney Animation: The Illusion of Life, Abbeville Press, New York, 1984, pp. 103-145.

[11] Synder, Industrial Robots, pp. 301-303.

[12] Doris Kockanek, 'Interpolating Splines with Local Tension Continuity and Bias Control', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 18, No. 3, 1984, pp. 24-41.

Alvy Smith, 'Spline Tutorial Notes', *Technical Memo*, No.77, Lucas Film, 1983.
N Steketee, 'Parametric Keyframe Interpolation Incorporating Kinematic Adjustment and Phase Control', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 19, No.3, 1985, pp. 255-262.
B Basky, 'Local Control of Bias and Tension in Beta-Splines', ACM Transactions on Graphics, Vol. 2, No. 2, 1983.
K Shelley, 'Path Specification and Path Coherence', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 16, No. 3, 1982, pp. 157-166.

[13] Craig Reynolds, 'Computer Animation with Scripts and Actors', Proceedings ACM SIGGRAPH, Vol. 16, No. 3, 1982, pp. 289-296.
R Chaug, '3D-Shaded Computer Animation Step-by-Step', *IEEE Computer Graphics and Applications*, Vol. 3, 1983, pp. 18-25.

[14] R Hackathorn, 'ANIMA II A 3D Colour Animation System', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 2, No. 11, 1977, pp. 54-64.

[15] T O'Donnell, 'Gramps - A Graphical Language Interpreter for Real-Time, Interactive Three-Dimensional Picture Editing and Animation', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 15, No. 3, 1981, pp. 133-142.

[16] Craig Reynolds, 'Flocks, Herds and Schools: A Distributed Behavioral Model', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 21, No. 4, 1982, pp. 25-34.
B Patridge, 'The Structure and Function of Fish Schools', *Scientific American*, Vol. 246, No.6 June 1982, pp. 114-123.
M P Hassell, *The Dynamics of Competition and Predation*, Edward Arnold, London, 1976.
R A Hide, *Animal Behaviour a Synthesis of Ethology and Comparative Psychology*, McGraw-Hill, London, 1966.

[17] Norman Badler, 'Articulated Figure Positioning by Multiple Constraints', *IEEE Computer Graphics and Applications*, June 1987, pp. 28-38.
Zeltzer, 'Representation', 1984.

[18] R Tomovic, 'A Finite State Approach to the Synthesis

of Bioengineering Control System', IEEE Transactions on Human Factors in Electronics, Vol. 7, No. 2, 1966, pp. 65-69.

[19] Synder, Industrial Robots, p. 312. Lieberman L and Wesley M, 'AUTOPASS: An automatic programming System for Computer Controlled Mechanical Assembly', IBM Journal of Research and Development, July, 1977.

[20] V Evarts, 'Brain Mechanisms of Movement', Scientific American, Vol. 241, No. 3, September 1979, pp. 146-154.

[21] K Pearson, 'The Control of Walking', Scientific American, December 1976, pp. 72-86.

[22] David Zeltzer, 'Toward an Integrated View of 3D Computer Animation' The Visual Computer, Vol. 1, No. 4, 1984, pp. 249-259.

[23] Micheal Girard, 'Interactive Design of 3D Computer-Animated Legged Animal Motion', IEEE Computer Graphics and Applications, Vol. 7, No. 4, 1987, pp. 39-51.

[24] Evarts, 'Brain Mechanisms', p. 154.

[25] R Warwick and P L Williams, Gray's Anatomy, 35th edition Longman, London, 1973.

[26] Evarts, 'Brain Mechanisms', p. 154.

[27] Keith Waters, 'Towards Autonomous Control for Facial Animation', BCS Displays Group, 1987, pp. 1-10.

[28] B Chandrasekaran, 'Natural and Social System Metaphors for Distributed Problem Solving' IEEE Transactions on Systems, Man, and Cybernetics, Vol. 1, SMC-11, 1981, pp. 1-5.

CHAPTER EIGHT

Chapter Eight

Extensions to Current Work and Conclusions

8.1 Modelling a Flexible Topology for the Face.

Four data acquisition techniques for the face were described in section 3.2 - lofting, light beam or lasers, mechanical digitizers and stereophotogrammetery - each partly determining the representation of the raw data captured, and the subsequent free-form surface generated. The techniques described and employed in this present research result in a polygonal data structures and are the fundamental building block on which further manipulations can be applied.

Geometric distortions and soft moulding algorithms have been demonstrated to be powerful tools for the manipulation of the face once it has been constructed, but if the full potential of these techniques is to be exploited, they must be based upon real processes that occur in the face. A good example is the cardioid strain manipulation that is a technique to simulate the growth sequence from infancy to adulthood (see section 3.4). On the other hand the manipulations of free-form deformation, while providing powerful moulding techniques, are not based upon any physical constraint, and as a result their use is limited.

Patches are an additional technique for the modelling of facial forms and are based on a fundamental polygonal structure. In Chapter 5.10 patches were utilized to demonstrate that mouth shapes can be accurately modelled. The advantages of patches are twofold. Firstly, the polygonal resolution can be selected for the model and, secondly, the control of the face can be more precisely managed by the guiding polygons. The guiding polygons can be controlled with the muscle model process, and are a closer approximation to skin than polygonal descriptions [1].

Modelling the skin on the underlying fundamental structure of the face requires further development. Describing the features of the face features from the skull is pertinent area of research, especially in reconstructive surgery and facial identification [2]. Information concerning skin depth and tissue type at key nodes on the face could provide the necessary information for the surface skin to be patched together. These problems are soluble from enough quantitative measurements of tissue depth. The most difficult regions of the face to discern employing this technique, are the soft features such as the nose and mouth where no indication of their shape is given from the underlying bone structures. This is critical as the eyes, nose and mouth regions that are vital cues for facial recognition. A solution would be to provide a comprehensive three-dimensional parameterized model.

8.2 Parameterization.

Parameterization is a coherent and desirable method for the generation and control of articulating structures. In this present research parameterization of the physical attributes of the head was not employed because clearly defined, or bounded sets of parameters, have not been described to encapsulate all the permutations of skull, muscle and skin. The face therefore, in this present research, was considered to be an independent topology. This is not to suggest that parameters cannot be defined for the physical facial attributes. Rule-based or heuristic schemes could be developed [3]. Initially this would require the measurement of hundreds of three-dimensional nodes from different faces of different race, age, sex groupings before a coherent strategy could be developed for the parameterization.

The parameterization of the muscle model process describing the three muscle types - linear, sphincter and sheet - has demonstrated that synthetic three-dimensional facial expressions can be defined in relation to the muscle contractions defined by Facial Action Coding System. FACS is not concerned with the motion attributes involved in expressions, but, because the muscle model process closely resembles the bio-mechanics of muscle, three-dimensional animated facial expressions can be generated, thereby extending the range of FACS.

8.3 The Biological Structure of Skin and Muscles.

The process described in 4.3 provides three muscle types - linear, sphincter and sheet - for the animation of the facial form. This type of synthesis is the first step towards a more comprehensive modelling of facial musculature. The muscle model process makes it possible to produce expressions quickly and effectively based upon an anatomical derivation, and is an approximation of the bio-mechanical muscle activity which results in the displacement of the skin. Two additional characteristics could be employed to improve the realism of the current muscle model behaviour, creasing and buckling of skin and a simulation of the visco-elastic properties of facial tissue.

8.3.1 The Creasing and Buckling of Skin.

Most buckling occurs at right angles to the direction of contraction. By calculating the contraction length of the muscle fibre and comparing it to the elasticity of the flesh, could provide information on the extent of the creasing. Also, the rendering algorithm could be involved by the perturbation of the surface normals to give the appearance of wrinkles in the surface [4]. Normal perturbation however, would only provide a partial solution, as the physical geometry would not be deformed.

8.3.2 Visco-Elastic Properties of the Face.

The models described in this research do not react to the underlying bone structure of the skull. This is a complex problem to achieve effectively. One solution would be to create a vector field around the underlying bone, consequently, as the muscle contracts the flow of skin would then be determined by the local potential. This requires an exact and precise model of the skull and mandible before the vectors are generated.

8.3.3 Tension Nets.

Skin behaves as an elastic mesh deforming under the influences of internal and external forces and it would improve the existing muscle model process with the exact simulations of the visco-elastic properties of flesh.

The skin can be considered as an interconnected network such that every node directly influences its immediate neighbour. This is the principle involved in tension nets [5]. Fundamentally, it is the extension of Hooke's Law, that states that the force F exerted by a spring at either end is linearly proportional to the displacement x of the length from some equilibrium length dependant upon the spring constant K:

$$F = -Kx$$

The force applied at this node is then applied through adjacent arcs to adjoining nodes, the process is then repeated until the force is dissipated through the mesh and a level of resolution is reached. Forces along arcs are proportional to their displacements, the spring constant and the vector direction of the contraction:

$$F(p2) = [F(p1) * (\mathrm{Arc}\ p(1), p(2)] * U(p(1),p(2))$$

where:

$F(p2)$	is the force at the next node,
$F(p1)$	is the force at current node,
$\mathrm{Arc}(p(1), p(2))$	is the arc between the nodes,
$U(p(1),p(2))$	is the unit vector between nodes.

The mesh node tension varies over time, consequently it is necessary to sum all the elasticities affecting that node:

$$1/K = \sum (1/k(p(1),p(2))$$

Then the final displacement could be calculated for each individual x,y,z component:

$$x' = x + F(p)/K$$

where:

$F(p)$	is the force applied to the node,
K	the sum of the spring elasticities,
x	the initial node position,
x'	the resultant displacement for node.

If this process is iterated outward across adjacent nodes on the face, eventually the force will become lower and lower, at which point the propagation could be terminated.

Controlling the tension network with the muscle vectors would provide an accurate, versatile and manageable means of creating synthetic facial expressions.

8.4 Animation Control.

The animation sequences V.4 - V.8 were controlled with parameters that are multi-tracked to form expressions and this is considered to be a low-level motion control system. The existing modes of motion control, in both robotics and animation, do not clearly encapsulate biological structures, and it has been demonstrated with the hand that a more suitable control can be achieved by grouping low-level motor programs together.

Facial expressions in this research are based on muscle actions, and grouping the activity of muscle as functional synergies that form facial expressions would be a more direct and manageable interpretation of complex activities of the face, as direct goals could be specified which are also comprehensible to the user and to the system.

Motor programs appear to be inherent in a great deal more human activity than just the hand and the face. This indicates that further investigation is required into the nature of how biological motor behaviours are controlled and how these descriptions can be employed in computer graphics.

8.5 Towards Automatic Speech Synchronization.

Speech is the most complex problem to synthesise effectively. The mouth, lips, tongue and teeth coalesce in a multitude of shapes as the utterance is made. The research has demonstrated that the phonetic break-down only provides an approximation to lip synchronization as fluent speech is more than the creation of static mouth shapes to the audio. The muscle model process has demonstrated that a wide variety of mouth shapes can be generated in two and three dimensions and it is the most accurate mode of speech synchronization. Speech is inherently a multi-modal means of communication, with the whole face involved, and this present research has demonstrated that both characteristics can be achieved concurrently with the muscle model process.

Spectrograms and analogue synthesizers can provide quantitative information of speech. This has recently been used by Hill et al for the three-dimensional computer synthesis of the speech utterances 'Hi there, how are you', and 'Speak to me now, bad kangaroo' [6]. The techniques were used in combination with Parke's original model and nine parameters to control the jaw and lips. This present research has demonstrated that the generation of mouth shapes from a biomechanical structure of muscle and skin is a clearer definition of the activity, and therefore the techniques of Hill et al could be more coherently applied.

There remains however, the primary problem of speech synchronization. Computer animation, so long as it tries to synchronize to an existing sound track, will only produce approximations to what already exists. A more comprehensive approach is to be involved with speech production rather that speech synchronization. This would involve the modelling of the internal organs of the mouth and the analysis of speech resonance in the vocal track.

8.6 Surface Rendering.

The images created in this research have not considered rendering algorithms, only basic Gouraud and Phong shading have been employed, and have resulted in plastic looking faces [7]. To enhance the rendered realism of facial images a number of techniques can be employed.

8.6.1 Texture Mapping.

Texture mapping is a technique employed to map surface details onto the three-dimensional polygonal surfaces [8]. The surface of the face could be constructed in three dimensions and a whole two-dimensional face could be mapped onto this topology. Recent research by Yau and Duffy have exploited this technique and have created animation sequences by interpolating individual sub-images [9]. The sub-image interpolation approach has the same inherent problems as key frame interpolation (see section 1.2.1) and could be greatly improved by the application of the techniques described by this present research. Texture mapping could also be applied to surfaces such as a stubble texture for the chin or hair texture for the brows.

8.6.2 Procedural Texture.

Procedural texture has been demonstrated to create realistic images of marble texture and other material types [10]. Skin colour is not uniform over the face, therefore a procedural algorithm could imitate stubble on the chin or the the hair of the brows, the subtle redness of the cheeks and lips, the darkening under the eye lids and slight paleness of the bridge of the nose and forehead. Procedurally driven colour for the face reduces the data storage required for the texture map and also the explicit model definition for each model. It is therefore a very powerful technique capable of parameterization.

8.7 Conclusions.

Previous research had two critical and connected constraints preventing precise, accurate and universal facial animation. The construction of computer-generated faces was predominately involved with unique models resulting in predefined computer algorithms being coordinated with the construction of the face. Consequently, each face exhibited a narrow range of facial expressions applicable only to that model and no other.

This present research has developed effective techniques for the modelling, animation and control of a broad range of computer-generated faces. These are demonstrated with static illustrations and animated in video sequences. The techniques surmount the two significant restrictions of previous research by isolating the construction of the face from the motion attributes, thereby allowing a free range of facial topology to be constructed.

The muscle model process was the most significant development. This is based on a comprehension of the underlying facial structure that is non-specific to a particular facial topology. The technique provides a new framework for the construction of static and animated facial expressions based on biological motivators rather than arbitrary selected parameters and has therefore substaintially extended our knowledge in this area.

Static facial expressions, employing the muscle model process, have been demonstrated to correspond accurately to real facial expressions in accordance with a facial coding system developed by psychologists. The video animation sequences of the modular muscle actions have also demonstrated the potential extension of the facial coding system by the addition of time dependence.

Lip synchronization has been successfully controlled with the muscle model process. Static mouth shapes can be created from the synthetic muscle contractions and animated to synchronize with the sound track. The static mouth shapes during speech have also been effectively encapsulated with Bezier patch definition and the muscle model process. The advantages are twofold. Firstly, a more precise definition of the intricate curved surfaces of the lips can be generated at a specified resolution, and secondly, a smooth blending of the facets is produced as only the control points are deformed with the muscle contraction.

The parameterization of the eyes, eyelids, jaw and muscles was developed as an efficient technique by this present research for the control of the facial attributes. The primary synthetic muscle parameters - muscle type, muscle tension and zone of influence - are the most appropriate parameters for synthetic facial muscle simulation. Facial expressions can be generated from these low-level actions negating the requirement for individually selected parameters.

The development of a musculature basis for synthetic facial expression led the research to analyse biological motor control. Functional muscle synergies can be developed for computer activities such as the grasping activity of the hand. This provides a more comprehensive definition of the task for the user and the system. The research therefore concludes that biological control has to be the major new direction of future research, not only with faces, but also with a wide spectrum of computer-controlled animal structures.

The computer graphic algorithms developed in this present research allows precise, accurate and universal facial animation. The research also opens avenues for further synthetic facial motion studies in reconstructive surgery applications, as well as the psychological study of expression in non-verbal communication and speech therapy training.

8.8 References.

[1] Koji Komatsu, 'Human Skin Model Capable of Natural Shape Variation', The Visual Computer, Vol. 3, No. 4, 1988, pp. 265-271.

[2] J S Rhine and C E Moore, 'Facial Reproduction Tables of Facial Tissue Thickness of American Caucasoids in Forensic Anthropology', Maxwell Museum Technical Series 1, Alberque, New Mexico, 1982.

[3] Mark Gillenson, 'The Interactive Generation of Facial Images on A CRT using a Heuristic Strategy', PhD Thesis, Ohio State University, 1974.

[4] Jim F Blinn, 'Simulation of Wrinkled Surfaces', Proceedings ACM SIGGRAPH, Vol. 12, No. 3, 1978, pp. 286-292.

[5] David Haumann, 'Modelling the Behaviour of Flexible Objects', Course Notes No. 17, ACM SIGGRAPH, 1987, pp. 1-13.

[6] David R Hill, Andrew Pearce and Brian Wyvill, 'Animated speech: An Automated Approach using Speech Synthesized by Rules', The Visual Computer, Vol. 3, No. 4, 1988, pp. 277-289.

[7] H Gouraud, 'Continuous Shading of Curved Surfaces', IEEE Transactions on Computers, C-20, No. 6, 1971, pp. 623-628.
Bui-Tuong Phong, 'Illumination for Computer Generated Pictures', Communications of the ACM, Vol. 18, No. 6, 1975, pp. 311-317.

[8] J D Foley and A Van Dam, Fundamentals of Interactive Computer Graphics, Addison-Wesley Publishing, Reading, Massachusetts, 1981, pp. 588-590.

[9] John F S Yau and Neil D Duffy, 'Texture Mapping Approach to 3-D Facial Image Synthesis', Proceedings Eurographics, Sussex, 1988, pp. 17-30.

[10] Ken Perlin, 'An Image Synthesizer', Proceedings ACM SIGGRAPH, Vol. 19, No. 3, pp. 287-296.

APPENDICES

Appendices

A.1 Algorithmic Interpolation.

The following pseudo-code ´C´ describes how non-linear interpolation can be achieved to ease in and ease out of motions.

```
/***************************************************

This algorithm interpolates between the values of
0.0 and 24.0 using the selected ease-in or ease-out
over 12 frames.
Four selections can be made:
1 = constant velocity
2 = acceleration ease in
3 = deceleration ease out
4 = acceleration at deceleration ease in and ease out

*************************************************/
£Define Totalframes 12
£Define Pi          3.141592654
£Define Start_Val   0.0
£Define Fin_Val     24.0
```

```
START ;

for (Loop_C = 1; Loop_C < Totalframes; Loop_C++)

    {

Frame_Step = Loop_C / Totalframes ;

switch (Operation) {

      case 1 ;
        Val = Fin_Val / Frame_Step ;
      break ;
      case 2 ;
        Val = Fin_Val * ((1 - cos (Pi * Frame_Step * 0.5) ;
      break ;
      case 3 ;
        Val = Fin_Val * Sin (Pi * Frame_Step * 0.5) ;
      break ;
      case 4 ;
        Val = Fin_Val * (1-cos(Pi * Frame_Step)) * 0.5 ;
      break ;

                    }
       }
END ;
```

A.2 Action Unit Definitions.

The Action Units defined consists of approximately fifty independant facial actions (some are created by muscle combination blends). Additional facial actions are marked as AD. The list that follows are considered the most important action units. For the use with the muscle model process the muscle types are also associated with the action unit.

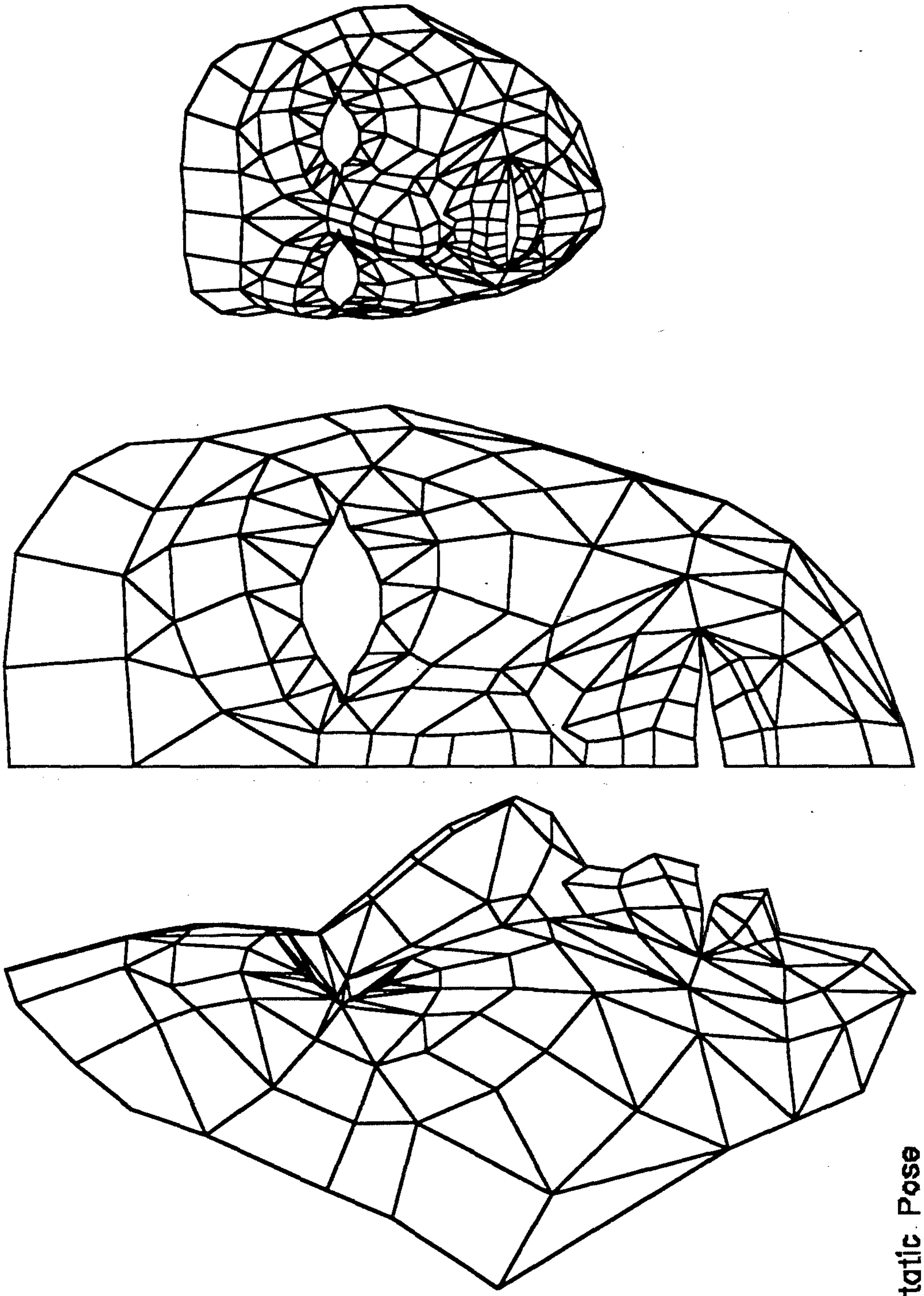
AU	FACS Name	Musclature Basis
AU1	Inner Brow Raiser	Medial portion of the Frontalis
AU2	Outer Brow Raiser	Laterial portion of the Frontalis
AU4	Brow Lowerer	Corrugator, depressor glabella and/or depressor supercilli
AU5	Upper Lid Raiser	Levator palpebrae superioris
AU6	Cheek Raiser and Lid Compressor	Orbicularis Oculi
AU7	Lid Tightener	Inner portion of the Orbicularis Oculi
AU8	Lips Towards each other	Orbicularis Oris
AU9	Nose Wrinkler	Levator Labii Superioris alaeque nasi
AU10	Upper Lip Raiser	Levator Labii Superioris Caput infraorbitalis
AU11	Nasolabial Furrow Deepener	Zygomatic minor
AU12	Lip-Corner Puller	Zygomatic major
AU13	Cheek puffer	Caninus
AU14	Dimpler	Buccinator
AU15	Lip Corner Depressor	Triangularis

AU16	Lower Lip Depressor	Depressor labii Inferioris
AU17	Chin Raiser	Mentalis
AU18	Lip Puckerer	Incisivii labii superioris
AD19	Tongue Show	(Miscellaneous)
AU20	Lip Stretcher	Risosious
AU21	Neck Tightener	(Miscellaneous) Platysma
AU22	Lip Funneler	Obicularis oris
AU23	Lip Tightener	Oricularis oris
AU24	Lip Presser	Obicularis oris
AU25	Lips Part	Depressor labii
AU26	Jaw Drop	Masseter
AU27	Mouth Stretch	Pterygoids
AU28	Lips Suck	Oribularis oris
AD30	Jaw Sideways	
AD31	Jaw Clencher	
AD32	Bite	
AD33	Blow	
AD34	Puff	
AD35	Suck	
AD36	Bulge	
AD37	Lip Wipe	
AU38	Nostril Dilator	Nasalis
AU39	Nostril Compressor	Nasalis
AU41	Lid Droop	Levator palpebrae
AU42	Slit	
AU43	Eyes Close	Levator palpebrae
AU44	Squint	Orbicularis oculi
AU45	Blink	Lavator palpebrae
AU46	Wink	Orbicularis oris
AU51	Head Turn Left	
AU52	Head Turn Right	
AU53	Head Up	
AU54	Head Down	
AU55	Head Tilt Left	
AU56	Head Tilt Right	
AU57	Head Forward	
AU58	Head Back	
AU61	Eyes Turn Left	
AU62	Eyes Turn Right	
AU63	Eyes Up	
AU64	Eyes Down	
AU65	Walleye	
AU66	Cross-Eye	

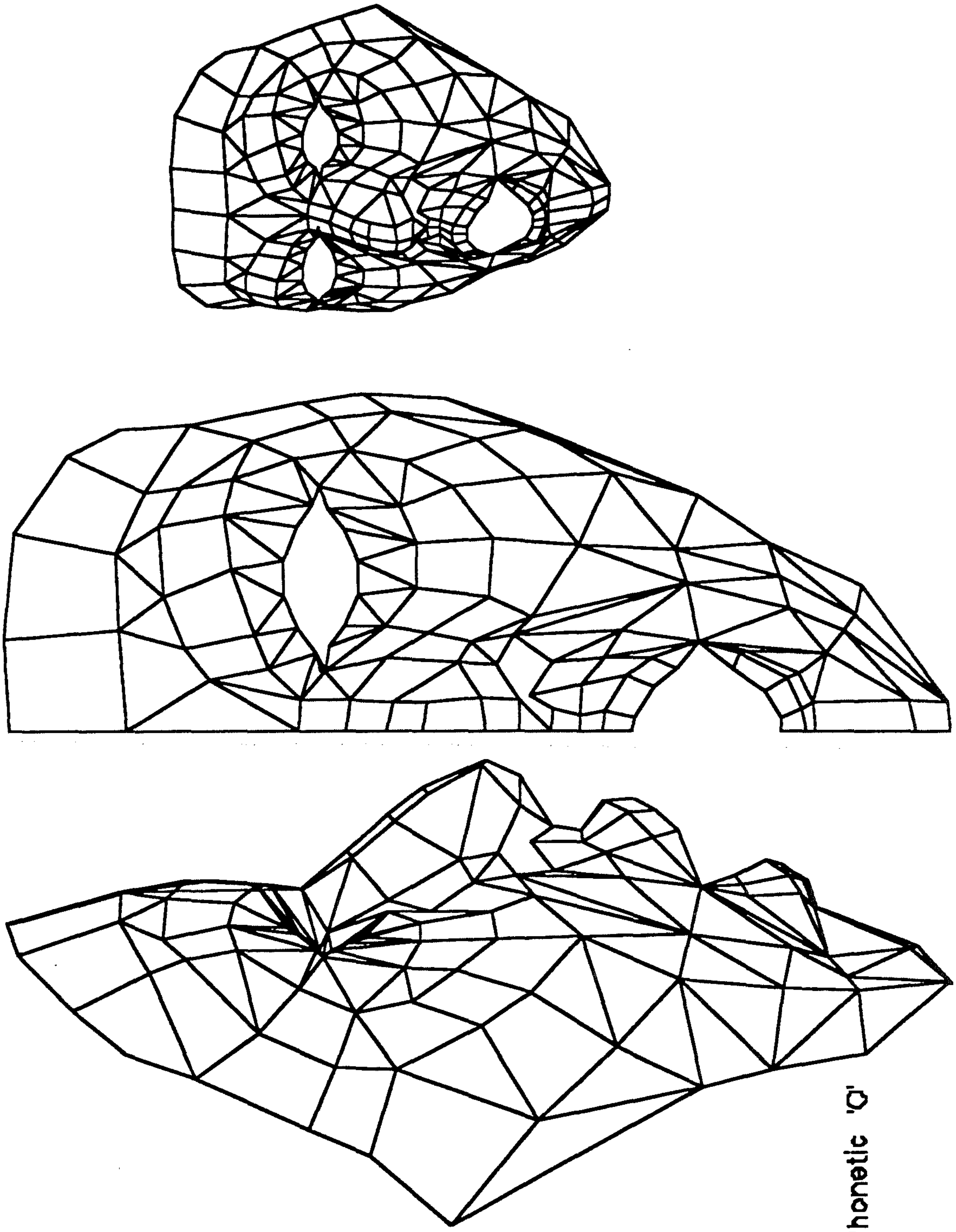
A.3 Appendix of Phonetic Guides.

The static phonetic guides for Investigation 1. the divisions are:

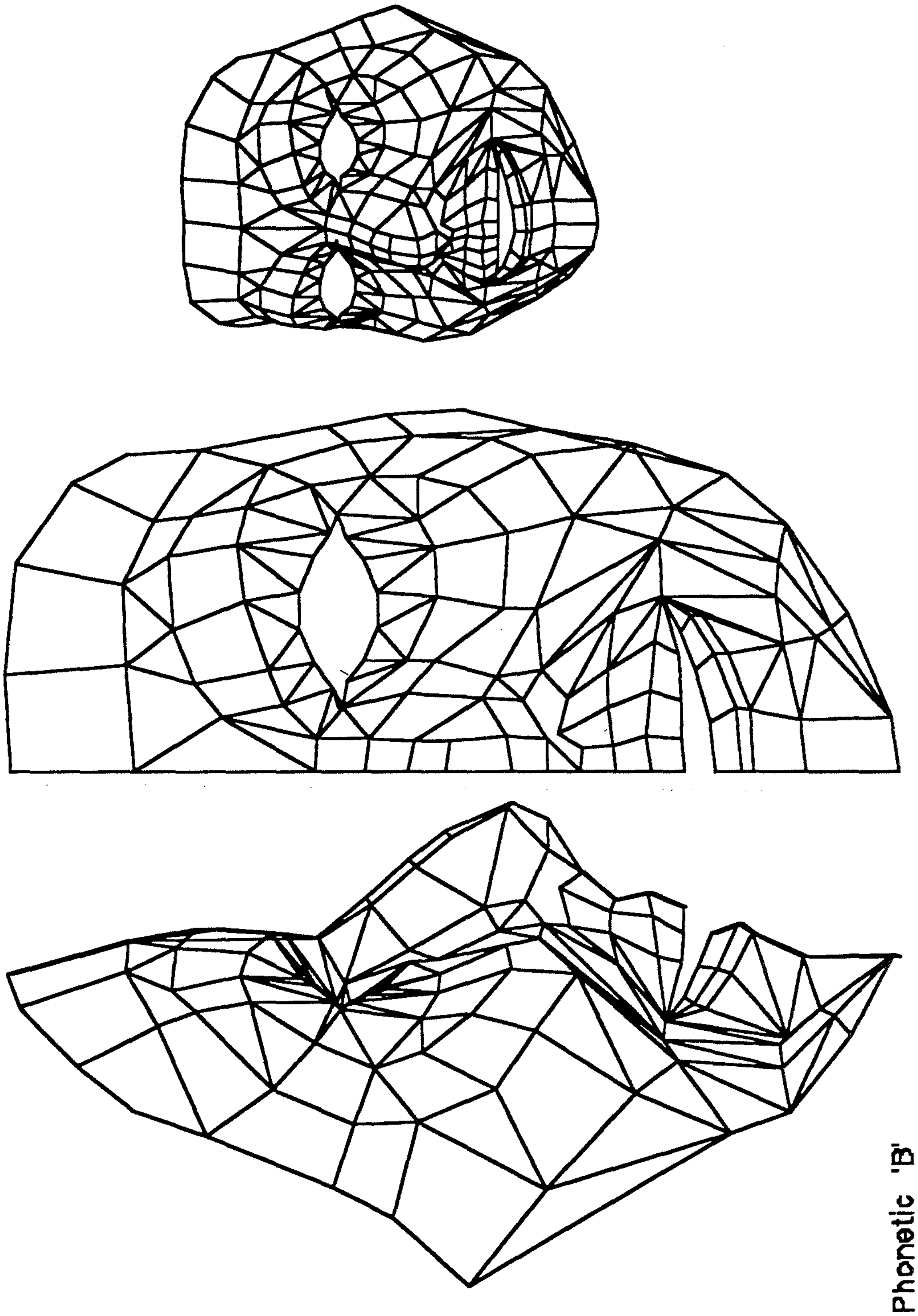
i. Static
i. /b, m, d, p/
ii. /o, u, w/
iii. /f, v/
iv. /a, e, i/
v. /s, t, z/



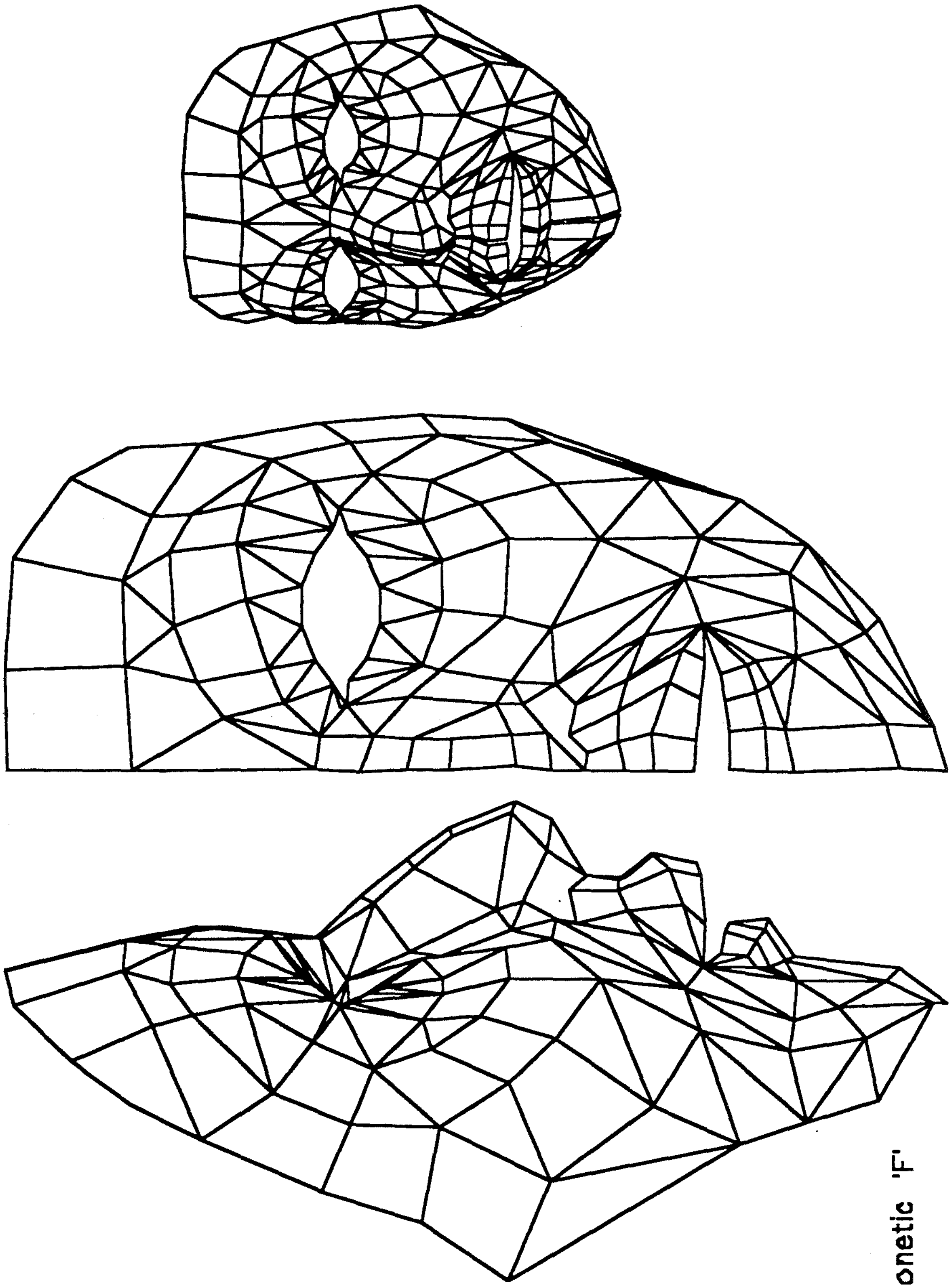
Static Pose



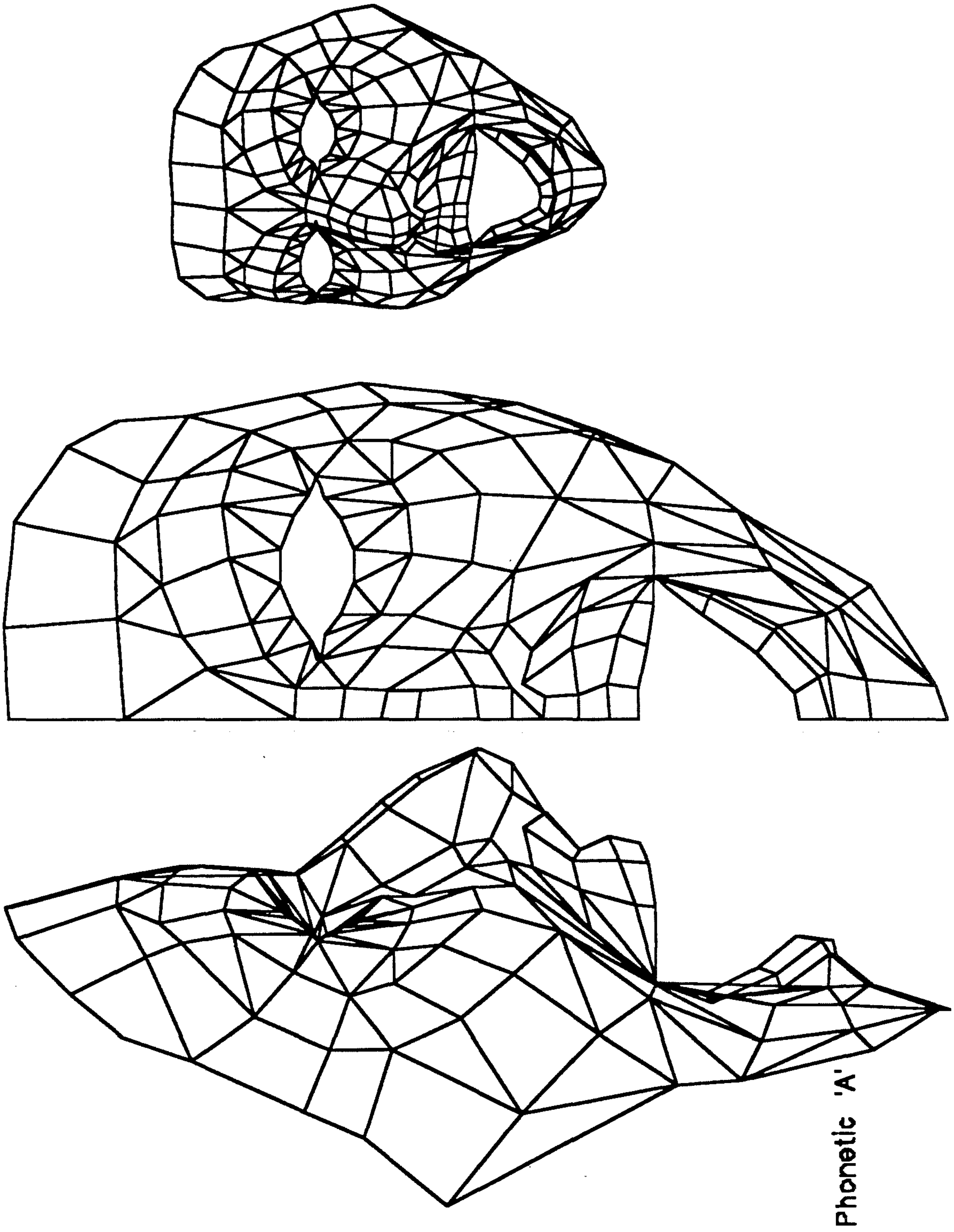
Phonetic 'O'



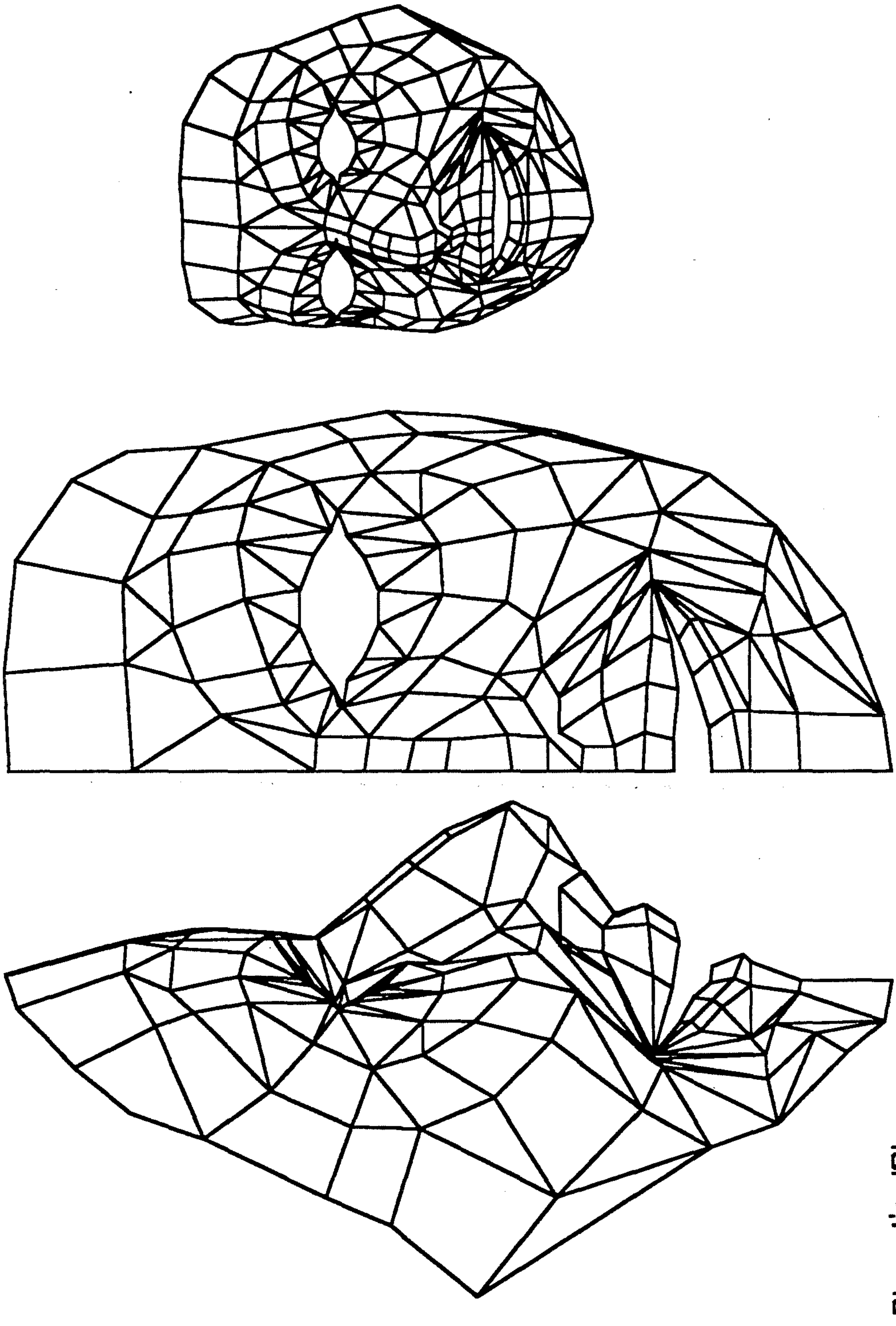
Phonetic 'B'



Phonetic 'F'



Phonetic 'A'



Phonetic 'S'

A.4 Graphics Library Routine Headers.

This Appendix lists the routine headers developed and employed for the construction and animation of the sequences. It is split into a general library of graphics routines that manipulates the data structures, and a muscle library where specialized muscle actions can be described.

All the algorithms were written in portable ´C´ both under VAX VMS on a 11/785 with a 24 bit frame store, and also in ´C´ under Unix with an eight bit frame store. The ´C´ library of routines was constructed to be modular in usage and could be cross compiled and linked with an existing Fortran library of general graphics and rendering routines [4]. Under the Unix environment the ´C´ library routines passed parameters to the low level display algorithms written in micro code.

A.4.1 General Library of Graphic Routines.

Function : ANIMAT
Operation : Returns the Value from a File Depending upon the Ith Frame
Arguments : (Val, Ith, Start, Finish, Value)
Val : Returned Value
Ith : Current Frame
Start : Array holding the Start Values

Finish : Array holding the Finish Values
Value : Array holding the Values

Function : CALLSPEC
Operation : Calls in from Disc
Arguments : (FileName, Ara1, Ara2, Ara3)
FileName : Input Character String
Ara1 : First Loaded Array
Ara2 : Second Loaded Array
Ara3 : Third Loaded Array

Function : CONVERT
Operation : Converts the data base into eye coordinates
Arguments : (Ara)
Ara : Array to be converted

Function : BEZIER
Operation : Creates a Bezier Curve from a Guiding Polygon
Arguments : (Guide, Npts, Curve)
Guide : Guiding Polygon
Npts : Number of Points on the Curve
Curve : Return Spline

Function : COPY
Operation : Makes a Copy of Ara into Arb
Arguments : (Ara, Arb)
Ara : Input array
Arb : Output array

Function : EXTCON3
Operation : Extracts a Contour
Arguments : (Ara, N, Arb)
Ara : Array Holding Contour
N : Contour Pointer
Arb : Output Contour

Function : FACLIST
Operation : Calls in Three-Dimensional Data from Disk and Creates a Data Structure for Gouraud Rendering.
Arguments : (Ara, FileName)
Ara : Globally stored Data Structure
FileName : Input Character String

Function : MIX
Operation : Create a Three-Dimensional File from two Input Files
Arguments : (Ara, Arb, Arc)

Ara : First Input Array x,y
Arb : Second Input Array x,z
Arc : Output Three-Dimensional Array

Function : NORMS
Operation : Calculates the Surface Normals A, B and C and stores them in the shading data base
Arguments : (Ara)
Ara : Array on which to operate

Function : PCEYE
Operation : Calculates the necessary transforms for perspective drawing commands. The information is stored globally to the programs
Arguments : (Xf, Yf, Zf, Distance, Azmith, Elevation)
Xf,Yf,Zf : Point Under Observation
Distance : Distance from the object
Azmith : Real degrees of Rotation
Elevation : Real Degrees of Elevation

Function : INARA2D
Operation : Input a Two-Dimensional Array
Arguments : (Ara, FileName)
Ara : Array to be Loaded
FileName : Input Character String

Function : INBET
Operation : Interpolation of three-Dimensional Data using Cosine
Arguments : (Ara, Arb, Arc, Step, Total)
Ara : First Array to be interpolated
Arb : Second Array to be Interpolated
Step : Current step in interpolation
Total : Total number of Frames to interpolate

Function : IN3ARA
Operation : Input a Three-Dimensional Array
Arguments : (Ara, FileName)
Ara : Array to be Loaded
FileName : Input to be Loaded

Function : MIRROR3D
Operation : Three-Dimensional Mirror
Arguments : (Ara, Arb)
Ara : Input array
Arb : Output array

Function : CAGE

Operation : Swept Revolution Horizontally from pointers into the Two-Dimensional Contour.
Arguments : (Arb, Nsteps, Vtxl, Vtx2, Arb)
Arb : Input Two Dimensional Profile
Nsteps : Number of Revolutions in the Profile
Vtxl : Pointer to the start vertex
Vtx2 : Pointer to the finish vertex
Arb : Output Array

Function : JOINPR
Operation : Join two arrays Together
Arguments : (Ara, Arb)
Ara : Array to be loaded
Arb : Array receiving Ara

Function : LINK
Operation : Links Across Two Three-Dimensional Contours
Arguments : (Ara, Arb, Arc)
Ara : First Contour
Arb : Second Contour
Arc : Array of Contours

Function : ROTX
Operation : Rotates Object about the origin about X-axis.
Arguments : (Ara, Angle)
Ara : Array to be Manipulated
Angle : Angle in Degrees of Rotation

Function : ROTY
Operation : Rotates Object about the origin about Y-axis.
Arguments : (Ara, Angle)
Ara : Array to be Manipulated
Angle : Angle in Degrees of Rotation

Function : ROTZ
Operation : Rotates Object about the origin about Z-axis.
ArgumentS : (Ara, Angle)
Ara : Array to be Manipulated
Angle : Angle in Degrees of Rotation

Function : SCALE3
Operation : Scales an Object in Three-Dimensions
Arguments : (Ara, Size)
Ara : Array to be Scaled
Size : Scaling Parameter

Function : SHIFT3
Operation : Three-Dimensional Shift in x,y and z
Arguments : (Ara, Shx, Shy, Shz)
Ara : Array to be shifted
Shx : Displacement in x
Shy : Displacement in y
Shz : Displacement in z

Function : SPEECH
Operation : To Specify Mouth Shapes from Disk for Interpolation
Arguments : (FileName, Ith, Prevarray, Retarray, Iptb)
FileName : Input Filename holding the order list of mouth shapes and interpolation intervals
Ith : Current Frame
Prevarray : Previous Mouth Shape Array
Retarray : Return Mouth Shape Array
Iptb : Updated Pointer for the Interpolation Interval

Function : TO3D
Operation : Adds a Zero z component to a two-dimensional array to construct a three-dimensional file.
Arguments : (Ara, Arb)
Ara : Input Two-Dimensional Array
Arb : Output Three-Dimensional Array

Function : TRICUT
Operation : Triangulator
Arguments : (Ara, Arb)
Ara : Input Array
Arb : Output Array

Function : VTXCUL
Operation : Culls Co-Linear Vertices From an Array
Arguments : (Ara)
Ara : Input Array

A.4.2 Muscle and Face Library Functions.

Function : SHEETMUSCLE
Operation : Emulates Sheet Muscle
Arguments : (Ara, Tx, Ty, Tz, Hx, Hy, Hz, FallDist, Tension)

Ara : Array to be Distorted
Tx,Ty,Tz : Tail of Vector
Hx,Hy,Hz : Head of Vector
FallDist : Falloff Distance
Tension : Muscle Tension

Function : MAKEYE
Operation : To Construct the Eyeball
Arguments : (Ball, Radius, PupilSize, IrisSize)
Ball : Returned Array Holding the Eyeball
Radius : Radius of the Eyeball
PupilSize : Size of the Pupil
IrisSize : Size of the Iris

Function : M3DFACE
Operation : Construct the Face from Input Data Files
Arguments : (Ara, Arb, Nfac, Constin)
Ara : Input List of Vertices
Arb : Output Facet Data
Nfac : Total Number of Facets
Constin : Input File Holding the Facet Pointer Data

Function : PULL3D
Operation : Linear Muscle Pull
Arguments : (Ara, Tx, Ty, Tz, Hx, Hy, Hz, Fallstart, Fallfin, Theta, Tension)
Ara : Ara to be distorted
Tx,Ty,Tz : Tail Position of Muscle Vector
Hx,Hy,Hz : Head Position of the Muscle Vector
Fallstart : Distance along Muscle Vector start of the Fall-off
Fallfin : Distance along Muscle Vector to the Fall-off Finish
Theta : Half Angle for the Zone of Influence
Tension : Tension in Muscle

Function : ROTJAW
Operation : Rotate the Jaw of the Face
Arguments : (Ara, Vtx1, Vtx2, Ang)
Ara : Array Holding the Face Data
Vtx1 : Start Pointer for the Jaw
Vtx2 : Finish Pointer for the Jaw

Function : SETBAL
Operation : To Focus the Eyes on Objects
Arguments : (Profile White, Iris, Pupil, IrisVtx, PupilVtx, OBx, OBy, OBz, Pox, Poy, Poz)
Profile : Input Profile of the Eye Ball in Two-Dimensions

White : Returned Array Holding the White of the Eye
Iris : Returned Array Holding the Iris of the Eye
Pupil : Returned Array Holding the Pupil of the Eye
IrisVtx : Pointer for the Start of the Iris
PupilVtx : Pointer for the Start of the Pupil
OBx,OBy,OBz : Input the Object x,y,z position
Pox,Poy,Poz : Input the eyeball x,y,z position

Function : SPHINCTER
Operation : Emulates the Sphincter Muscle of the Face
Arguments : (Ara, Cx, Cy, Tension)
Ara : Array to be squeezed
Cx,Cy : Longitudinal and Latitudinal Axis of the Ellipse
Tension : Tension of the Muscle

A.5 Three-Dimensional Displacements of Speech.

The data in this appendix is thre tree-dimensional unscaled displacement of the bottom lip, the top lip and the right corner of the lip from a static rest position for the sentence: ´I wonder how many people in this city,´.

	Bottom Lip			**Mid Lip**			**Top Lip**		
I	-10.0	-80.0	-23.0	-5.0,	0.0	0.0	-17.0	-42.0	-16.0
Won	-1.0	-22.0	-5.0	0.0	-4.0	2.0	0.0	0.0	0.0
Der	-18.0	-18.0	-12.0	-12.0	2.0	2.0	-3.0	-12.0	-5.0
H	-15.0	-64.0	-17.0	-11.0	5.0	4.0	-5.0	-21.0	-65.0
Ow	2.0	-21.0	-5.0	0.0	-12.0	0.0	-3.0	-10.0	-7.0
M	1.0	2.0	0.0	0.0	-10.0	0.0	10.0	0.0	-5.0
Any	-20.0	33.0	-15.0	0.0	-13.0	-10.0	10.0	-10.0	-20.0
P	0.0	-10.0	-2.0	4.0	-13.0	-8.0	0.0	0.0	-9.0
Eo	-20.0	32.0	-12.0	0.0	-12.0	-10.0	8.0	-10.0	-18.0
P	0.2	-11.0	-3.0	4.0	-13.0	-8.0	0.0	0.0	-10.0
Le	4.0	-40.0	-10.0	5.0	-3.0	0.0	10.0	-10.0	-18.0
In	-30.0	-42.0	-10.0	-21.0	-5.0	0.0	-8.0	-15.0	-15.0
Thi	0.0	-18.0	-12.0	0.0	-3.0	-7.0	0.0	-4.0	-12.0
Ci	-5.0	-21.0	-5.0	-5.0	5.0	0.0	0.0	0.0	5.0
Ity	7.0	31.0	-1.0	0.0	2.0	0.0	3.0	-2.0	-6.0

A.6 Parameteric Bezier Patch Definition.

```
Patch_Array_Type Patches =
{
   0,1,2,3,            4,5,6,7,            8,9,10,11,          12,13,14,15,
   12,13,14,15,        16,17,18,19,        20,21,22,23,        24,25,26,27,
   28,29,30,0,         31,32,33,4,         34,35,36,8,         37,38,39,12,
   37,38,39,12,        40,41,42,16,        43,44,45,20,        46,47,48,24,
   49,50,51,28,        52,53,54,31,        55,56,57,34,        58,59,60,37,
   58,59,60,37,        61,64,67,40,        62,65,68,43,        63,66,69,46,
   70,71,72,49,        73,74,75,52,        76,77,78,55,        79,80,81,58,
   58,61,62,63,        82,83,84,85,        86,87,88,89,        90,91,92,93,
   90,91,92,93,        94,95,96,97,        98,99,100,101,      102,103,104,105,
   102,103,104,105,    106,107,108,109,    110,111,112,113,    114,115,116,117,
   102,106,110,114,    118,119,120,121,    122,123,124,125,    126,127,128,129,
   90,94,98,102,       130,131,132,118,    133,134,135,122,    136,137,138,126,
   58,82,86,90,        139,140,141,130,    142,143,144,133,    145,146,147,136,
   79,80,81,58,        148,149,150,139,    151,152,153,142,    154,155,156,145,
   126,127,128,129,    157,158,159,160,    161,162,163,164,    165,166,167,168,
   136,137,138,126,    169,170,171,157,    172,173,174,161,    175,176,177,165,
   145,146,147,136,    178,179,180,169,    181,182,183,172,    184,185,186,175,
   154,155,156,145,    187,188,189,178,    190,191,192,181,    193,194,195,184,
};
```

```
Duck_Array_Type Ducks =
    {
      -11.0,    59.0,    74.0,
      -8.0,     58.5,    73.0,
      -3.0,     58.0,    73.0,
       0.0,     57.5,    72.0,

      -14.0,    51.0,    78.0,
      -8.5,     50.5,    78.5,
      -3.5,     50.3,    78.0,
       0.0,     50.0,    77.5,

      -16.5,    40.0,    83.0,
      -12.0,    39.0,    84.0,
      -5.0,     37.5,    83.5,
       0.0,     36.5,    85.0,

      -19.0,    30.0,    90.5,
      -14.0,    29.0,    89.0,
      -7.0,     26.0,    86.5,
       0.0,     24.0,    87.5,

/*--   Patch 2  --*/

        -21.0, 20.0,  90.0,
        -15.0, 19.0,  89.0,
        -7.5,  18.0,  89.0,
         0.0,  16.0,  88.0,

        -18.0, 6.5,  82.0,
        -13.0, 6.5,  82.0,
        -7.0,  6.0,  82.0,
         0.0,  6.0,  82.0,

        -18.0, 13.0,  76.0,
        -13.5, 12.0,  76.0,
        -8.0,  11.5,  76.0,
         0.0,  11.0,  76.0,

/*-- Patch 3 --*/

      -37.0,    52.0,  68.0,
      -30.0,    55.0,  70.0,
      -21.0,    57.0,  74.0,

      -40.0,  42.0,  71.5,
      -32.0,  45.0,  74.0,
      -22.0,  49.0,  78.0,

      -42.0,  31.0,  77.0,
      -35.0,  35.0,  81.0,
      -26.0,  39.0,  85.0,

      -43.0,  16.5, 79.0,
      -36.0,  22.5, 82.5,
      -27.0,  27.0, 86.0,

/*--Patch 4 --*/

     -40.0, 11.0, 80.0,
     -34.0, 15.0, 83.5,
     -27.0, 18.0, 86.0,

     -34.0, 3.0, 74.0,
     -30.0, 4.0, 79.0,
     -23.0, 6.0, 82.0,

     -34.0, 9.0, 71.0,
     -30.0, 10.0, 73.5,
     -23.0, 12.0,  75.5,

/*--Patch 5 --*/

      -63.0,  34.0,  51.5,
      -57.0,  39.0,  57.0,
      -49.0,  44.5,  63.0,

      -66.0,  21.5,  54.0,
      -58.0,  28.0,  60.0,
      -50.0,  35.0,  68.0,

      -67.0,  8.0,   58.0,
      -60.0,  14.5,  65.0,
      -51.0,  22.5,  74.0,

      -69.0,  -9.0,  59.0,
      -60.0,   0.0,  68.0,
      -53.0,   7.0,  75.0,

/*-- Patch 6 --*/

        -69.0,  -9.0,  59.0,
        -69.0,  -9.0,  58.0,
        -69.0,  -9.0,  57.0,

        -59.0,  -2.0,  67.0,
        -55.0,  -6.0,  66.0,
        -56.0,  -3.0,  62.0,

        -51.5,  4.0,  74.0,
        -44.0,  -2.0,  73.0,
        -46.0,  3.0,  65.0,

/*-- Patch 7 --*/

        -90.0,  9.0,  23.0,
        -85.0,  15.0,  33.0,
        -75.0,  25.0,  44.0,

        -91.0,  -4.0,  26.0,
        -84.0,  5.0,  36.0,
        -75.0,  12.0,  47.0,

        -92.0,  -11.0,  29.0,
        -84.0,  -5.0,  38.0,
        -76.0,  2.0,  50.0,

        -91.0,  -16.0,  30.0,
        -84.0,  -14.0,  39.0,
        -75.0,  -12.0,  51.0,

/*--Patch 8 --*/

         -61.0,  -12.0,  65.0,
         -69.0,  -9.0,  66.0,
         -58.0,  -7.0,  64.0,
         -57.0,  -9.0,  62.0,

         -51.0,  -15.0,  72.0,
         -49.0,  -10.0,  73.0,
         -48.0,   -6.0,  69.0,
         -47.0,  -10.0,  66.0,

         -41.0,  -16.0,  81.0,
         -40.0,  -10.0,  78.0,
         -39.0,  -5.0,  72.0,
         -38.0,  -10.0,  67.0,

/*--Patch 9 --*/

        -37.0,  -17.0,  83.0,
        -34.0,    -10.0,  81.0,
        -32.0,  -4.0,  75.0,
        -30.0,  -10.0,  69.0,

        -27.0,  -19.0,   87.0,
        -26.0,  -10.0,   85.0,
        -25.0,  -1.0,  74.0,
        -23.0,  -11.0,  70.0,

        -22.0,  -20.0,   89.0,
        -22.0,  -11.0,   88.0,
        -21.0,  -2.0,    77.0,
        -17.0,  -10.0,  70.0,

/*-- Patch 10 --*/

        -16.0,  -19.0,  90.0,
        -15.5,  -9.0,  89.0,
        -13.0,  -3.0,  79.0,
        -11.0,  -7.0,  72.0,

        -8.5,  -19.0,  92.0,
        -8.0,  -8.0,  89.0,
        -6.0,  -2.0,  78.0,
        -5.0,  -8.0,  72.0,

         0.0,   -18.0,  93.0,
         0.0,   -7.0,   89.0,
         0.0,   -3.0,   78.0,
         0.0,   -8.0,   72.0,

 /*-- Patch 11 --*/

         -22.0,  -34.0,  73.0,
         -16.0,  -34.0,  74.0,
         -8.0,   -34.0,  75.0,
          0.0,   -34.0,  76.0,

         -21.0,  -60.0,  71.0,
         -16.0,  -60.0,  72.0,
         -8.5,  -60.0,   73.0,
          0.0,  -60.0,   74.0,

         -24.0,  -85.0,  90.0,
         -18.0,  -85.0,  90.0,
         -8.5,  -85.0,  90.0,
          0.0,  -85.0,  90.0,

 /*-- Patch 12 --*/

         -40.0,  -34.0,  69.0,
         -38.0,  -34.0,  72.0,
         -28.0,  -34.0,  73.0,

         -40.0,  -57.0,  75.0,
         -35.0,  -58.0,  78.0,
         -28.0,  -59.0,  80.0,

         -45.0,  -83.0,  76.0,
         -38.0,  -83.5,  82.0,
         -32.0,  -84.0,  87.0,

 /*-- Patch 13 --*/

         -66.0,  -31.0,  53.0,
         -58.0,  -32.0,  61.0,
         -47.0,  -33.0,  67.0,

         -67.0,  -53.0,  53.0,
         -60.0,  -53.0,  60.0,
         -49.0,  -55.0,  65.0,

         -68.0,  -77.0,  55.0,
         -61.0,  -80.0,  66.0,
         -53.0,  -81.0,  73.0,

/*-- Patch 14 --*/

        -90.0,  -35.0,  32.0,
        -82.05, -33.0,  41.0,
        -75.0,  -32.0,  53.0,

        -87.0,  -55.0,  40.0,
        -82,    -55.0,  40.0,
        -75.0,  -54.0,  52.0,

        -84.0,  -73.0,  30.0,
        -79.0,  -74.0,  39.0,
        -75.0,  -75.0,  49.0,

/*-- Patch 15 --*/

        -23.0,  -96.0,  92.0,
        -16.0,  -96.0,  92.0,
        -8.0,   -96.0,  92.0,
         0.0,   -96.0,  92.0,

        -22.0,  -110.0,  80.0,
        -15.0,  -110.0,  80.0,
        -7.5,   -110.0,  80.0,
         0.0,   -110.0,  80.0,

        -20.5,  -121.0,  55.0,
        -13.0,  -121.0,  55.0,
        -6.0,   -121.0,  55.0,
         0.0,   -121.0,  55.0,

/*--Patch 16 --*/

        -44.0,  -95.0,  91.0,
        -38.0,  -97.0,  91.0,
        -30.0,  -95.5,  91.0,

        -43.0,  -106.0,  79.0,
        -36.0,  -108.0,  79.0,
        -28.0,  -109.0,  79.0,

       -41.0,  -116.0,  54.0,
       -34.0,  -118.0,  54.0,
       -27.0,  -119.0,  54.0,

/*-- Patch 17 --*/

       -65.0,  -90.0,  55.0,
       -58.0,  -92.0,  63.0,
       -52.0,  -93.0,  69.0,

       -60.0,  -101.0,  53.0,
       -55.0,  -103.0,  59.0,
       -48.0,  -105.0,  64.0,

       -54.0,  -113.0,  49.0,
       -49.0,  -114.0,  52.5,
       -44.0,  -115.0,  54.5,

/*-- Patch 18 --*/

       -81.0,  -84.0,  31.0,
       -76.0,  -86.0,  37.0,
       -71.5,  -87.5,  49.0,

       -76.0,  -96.0,  30.0,
       -71.0,  -97.0,  37.0,
       -67.0,  -99.0,  47.0,

       -68.0,  -107,   25.0,
       -65.0,  -108,   35.0,
       -60.0,  -110.0,  45.0
};
```

A.7 Vertex Editor.

The vertex editor for the face consists of a number of elementary commands allowing the modification of the two-dimensional data until the desired three-dimensional data was produced.

The commands were as follows:

VIEW : Views the three-dimensional data constructed from the input files.

MOVEP : Move a Selected Vertex

FINDPOS : Print the Position of the cursor

MARKER : Mark the position of the cursor

INPUTF : Input data file name

NEXTV : Move to the next vertex in the data

PREV : Move to previous vertex in data

CLEARS : Clear the Screen

UPDATE : Update the data

The editor was designed to be interactive. It requires the basic two-dimensional files to be already digitized. The program is an executable file that enquires as to the file to edit and responds with the drawing of the image on the screen. Abreviations to the commands allow the selection of the cross-hairs to be displayed on the terminal that can be moved to specified verticies and selected, modified or deleted as required.

Biblography.

Addleman, D., and Addleman, L., 'Rapid 3D Digitizing, Computer Graphics World, November 1985, pp. 41-44.

Angell, I., A Practical Introduction to Computer Graphics, Macmillan Press, London, 1981.

Angell, I., and Griffith, G., High-Resolution Computer Graphics Using Fortran 77, Macmillan Press, London, 1987.

Arridge, S., Moss, J P., Linney, D A., James, D R., 'Three-Dimensional Digitization of the Face and Skull', Journal of Surgery, Vol. 13, 1985, pp. 136-143.

Asada, H., Slotine, J E., Robot Analysis and Control, John Wiley, New York, 1986.

Badler, N I., 'A System for Computer Simulation of the Human Face', MSc Thesis, University of Pennsylvania, 1980.

Badler, N I., 'Animating Facial Expression, Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 15, No. 3, 1981, pp. 245-252.

Badler, N I., 'Articulated Figure Positioning by Multiple Constraints', IEEE Computer Graphics and Applications, June 1987, pp. 28-38.

Baerle, S V., Computer-Generated Three-Dimensional Character Animation, MA Thesis, Ohio State University, 1985.

Baker-Shenk, C., 'Nonmanual Behaviour in Sign Languages: Methodological Concerns and Recent Findings', Nonmanual Behaviours, 1983, pp. 175-184.

Basky, B., 'Local Control of Bias and Tension in Beta-Splines', ACM Transactions on Graphics, Vol. 2, No. 2, 1983.

Bergeron, P.,'Techniques for Animating Characters', Advanced Computer Animation Course Notes, ACM SIGGRAPH, Vol. 10, 1985, pp. 61-80.

Bertillon, A., Signaletic Instructions, The Werner Company, Chicago, 1893.

Bezier, P., 'Mathematical and Practical Possibilities of Unisurf', Computer Aided Geometric Design, Academic Press, New York, 1974, pp. 127-151.

Birdwhistell, L., Kinesics and Context, University of Pennsylvania Press, Philadelphia PA, 1970.

Blinn, J, F., 'Simulation of Wrinkled Surfaces',Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 12, No. 3, 1978, pp. 286-292.

Bowyer, A., 'Symbolic Ray Tracing', Proceedings Computer Animation, Online Publications, 1987, pp. 127-133.

Bragdon, C., The Beautiful Necessity, Reprinted Quest, London, 1939.

Braid, I., 'Six Systems for Shape Design and Representation', University of Cambridge, CAD Group, Document No. 87, 1975.

Bright, S., 'Shading of Solid Voxel Models', Proceedings Eurographics UK, 1986.

Brooke, M., and Summerfield, Q., 'Analysis, Synthesis and Perception of Visible Articulatory Movements', Journal of Phonetics, Vol. 11, 1982, pp. 63-76.

Bruce, V., Young A., 'Understanding Facial Recognition', British Journal of Psychology, Vol. 77, 1986, pp. 305-327.

Bruce, V., and Valentine, T., 'When a Nod's as Good as a Wink The Role of Dynamic Information in Facial Recognition', Practical Aspects of Memory: Current Research and Issues, Vol. 1, Wiley, Chichester, 1988, pp. 169-174.

Calvert, T., Chapman, J., 'Aspects of the Kinematic Simulation of Human Movement', IEEE Computer Graphics and Applications, Vol. 2, No. 9, November, 1982, pp. 41-49.

Carico, C C., and Drooyan, I., Analytical Geometry, John Wiley and Sons, New York, 1980.

Catmull, E., 'The Problems of Computer-Assisted Animation', Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 12, No. 3, 1978, pp. 348-353.

Chandrasekaran, B., 'Natural and Social System Metaphors for Distributed Problem Solving' IEEE Transactions on Systems, Man and Cybernetics, Vol. 1, No. 11, 1981, pp. 1-5.

Chaug, R., '3D-Shaded Computer Animation Step-by-Step', IEEE Computer Graphics and Applications, Vol. 3, 1983, pp. 18-25.

Cohen, L., Selected Poems, Johnathan Cape, New York, 1969.

Crow, F., 'The Origins of the Teapot', IEEE Computer Graphics and Applications, Vol. 7, No. 1, 1987, pp. 8-19.

Cyberware Laboratory Inc., 'Computerized Three-Dimensional Reproductions', Sales Pamphlet, 2062 Sunset Drive, Pacific Grove, CA., 93950, 1986.

Destandes, J., Histoire Comparée du Cinéma, Vol. 1, Castarman, Paris, 1966.

Dreyfuss, H., The Measure of Man Human Factors in Design, Second Edition, Whitney Publishing, New York, 1967.

Duncan, P., 'Anatomical Definition and Modelling', Engineering in Medicine, Vol. 15, No. 3, 1986, pp. 109-116.

Ellis, H., 'Recognizing Faces', British Journal of Psychology, Vol. 66, 1975, pp. 409-425.

Ekman, P., Unmasking the Human Face, Prentice Hall, New York, 1971.

Ekman, P., Emotion in the Human Face, Cambridge University Press, London, 1972.

Ekman, P., Darwin and Facial Expression a Century of Research in Review, Academic Press, London, 1973.

Ekman, P., 'About Brows: Emotional and Conversational Signals in Human Ethology', Journal of Phonetics, No. 11, 1983, pp. 169-202.

Ekman, P., Manual for the Facial Action Coding System, Palo Alto Press, California, 1977.

Evarts, V., 'Brain Mechanisms of Movement', Scientific American, Vol. 241, No. 3, September 1979, pp. 146-154.

Faux, I D., and Pratt, M A., 'Computational Geometry for Design and Manufacture', Ellis Horwood, New York, 1979.

Ferner, H., Atlas of Human Anatomy, Urban and Schwarzenberg Munich, 1983.

Foley, J D., and Van Dam, A., Fundamentals of Interactive Computer Graphics, Addison-Wesley Publishing, Reading, Massachusetts, 1981.

Folkins, J W., and Abbs, J H., 'Lip and Jaw Motor Control during Speech: Response to Resistive Loading of the Jaw', Journal of Speech and Hearing Research, Vol. 18, 1975, pp. 207-220.

Fortin, D., Lamy, J F., and Thalmann, D., 'A multiple Track Animation System', Proceedings ACM SIGGRAPH, 1983, pp. 180-186.

Fuchs, H., 'Optimal Surface Reconstruction from Planar Contours', Communications of the ACM 20, pp. 693-702, 1977.

Fromkin, V., 'Lip positions in American English Vowels', Journal of Language and Speech, No. 7, 1964, pp. 215-225.

Gillenson, M L., The Interactive Generation of Facial Images on A CRT using a Heuristic Strategy, PhD Thesis, Ohio State University, 1974.

Gillenson, M L., 'A Heuristic Strategy for Developing Human Facial Images on a CRT', Pattern Recognition, Vol. 7, 1975, pp. 197-196.

Girard, M., Maciejewski, A., 'Computational Modelling for the Computer Animation of Legged Figures', Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 19, No. 3, 1975, pp. 263-270.

Girard, M., 'Interactive Design of 3D Computer-Animated Legged Animal Motion', *IEEE Computer Graphics and Applications*, Vol. 1, No. 4, 1987, pp. 39-51.

Godfrey, B., Jackson, A., *The Do-it-Yourself Animation*, British Broadcasting Corporation Publications, London, 1974.

Goldstien, N., *Figure Drawing: The Structure, Anatomy, and Expressive Design of Human Form*, Prentice-Hall, New Jersey, 1976.

Gomez, J., 'Twixt: A 3D Animation System', Eurographics, Elsevier Science Publishers, North-Holland, 1984, pp. 121-133.

Gouraud, H., 'Continuous Shading of Curved Surfaces', *IEEE Transactions on Computers,* C-20, No. 6, 1971, pp. 623-628.

Hackathorn., 'ANIMA II A 3D Colour Animation System', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 2, No. 11, 1977, pp. 54-64.

Harmon, L., 'The Recognition of Faces', *Scientific American*, Vol. 229, No. 5, November 1973, pp. 101-112.

Harper, R G., Wiens, A N., and Matarazzo, J D., *Non Verbal Communication "State of The Art,* John Wiley, New York, 1978.

Hassell, M P., *The Dynamics of Competition and Predation*, Edward Arnold, London, 1976.

Hide, R A., *Animal Behaviour a Synthesis of Ethology and Comparative Psychology,* McGraw-Hill, London, 1966.

Isaacs, P M., 'Controlling Dynamic Simulations with Kinematic Constraints, Behaviour Functions and Inverse Dynamics', Computer Graphics Proceedings, ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 215-224.

Maison, K, E., *Daumier: Drawings*, Sagamore Press, New York, 1960.

Muybridge, E., *The Human Figure in Motion*, Reprinted Dover Publications, NewYork, 1955.

Johansson, G., 'Visual Motion Perception', *Scientific American*, Vol. 232, No. 6, 1974, pp. 76-88.

Jones, D., *An Outline of English Phonetics*, First Edition, W Heffer, London, 1918.

Keslo, J A S., Vatikiotis-Bateson, E., Saltzman E., and Kay, B., 'A Qualitative, Dynamic Analysis of Reitterant Speech Production: Phase Portraits, Kinematics and Dynamic Modelling' *Journal of Acoustics*, Vol. 77, No. 1, January, 1977, pp. 266-280.

Kingsbury, D., 'Computer Generated Three-Dimensional Character Animation: A Report and Analysis', MA Thesis, Ohio State University, 1985.

Kockanek, D., 'Interpolating Splines with Local Tension Continuity and Bias', Computer Graphics Proceedings, ACM SIGGRAPH, Vol. 18, No. 3, 1984, pp. 34-41.

Komatsu, K., 'Human Skin Model Capable of Natural Shape Variation', *The Visual Computer*, No. 3, 1988, pp. 265-271.

Korien, J., 'Techniques for Generating the Goal-Directed Motion of Articulated Structures', *IEEE Computer Graphics and Applications*, Vol. 2, No. 9, November 1982, pp. 71-81.

Landau, B R., *Essential Human Anatomy and Physiology*, Scott Foresman and Company, Illinois, 1923, pp. 156-166.

Laughery, R., 'Recognition of Hmuan Faces: Effects of Target Exposure Time, Target Position, Pose Position, and Type of Photograph', *Journal of Psychology*, Vol. 55, 1971, pp. 477-483.

Lasseter, J., Principles of Traditional Animation Applied to Three-Dimensional Computer Animation, Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 35-44.

Latham, W., 'Visualization of Sculpture', Proceedings BCS Displays Group, December, 1987.

Lenihan, J., *Human Engineering*, Weidenfeld and Nicolson, London, 1974.

Lorensen, W E., and Cline, H E., 'Marching Cubes: A High Resolution 3D Surface Construction Algorithm', Proceedings ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 163-169.

Lord, M., 1987 'Curve and Surface Representation by Iterative B-Spline Fit to a Data Point Set', *Engineering in Medicine*, Vol. 16, No. 1, 1987, pp. 29-35.

Liggett, J., *The Human Face*, Constable and Company, London, 1974.

Lundin, R V., 'Motion Simulation', Computer Graphics Proceedings, NICOGRAPH, 1974, pp. 2-10.

McGovern, T., 'The Use of Live-Action Footage', Course Notes, ACM SIGGRAPH, No. 5, 1987.

McGrath, M., Summerfield, Q., and Brooke, M., 'Roles of Lips and Teeth in Lip-Reading Vowels', *Proceedings of the Institue of Acoustics*, 1984.

McGrawth, M., An Examination of Cues for Visual and Audio-Visual Speech Perception using Natural and Computer-Generated Faces, P.h.D Thesis, University of Nottingham, 1985.

Madsen, R., *Animated Film:Concepts, Methods and Uses*, Interland Publishing, New York, 1969.

Magnenat-Thalmann, N., Thalmann, D., *Computer Animation Theory and Practice*, Springer-Verlag, Tokyo, 1985.

McQueen., 'Applying Classical Techniques to Computer Animation', Course Notes, ACM SIGGRAPH, Vol. 10, 1987, pp. 217-230.

Menosky, J., 'Recording for the Future', *Science*, 1982, pp. 60-63.

Moore, J., 'Towards an Integrated Computer Package for Speech Therapy Training Talking Heads, MicroTech Solutions Report', Bradford College of Art and Design, 1986.

Moss, J P., Linney, A D., Gindrod, S R., Arridge, S., Clifton, J S., 'Three-Dimensional Visualization of the Face and Skull using Computerized Tomography and

Laser Scanning Techniques', European Journal of Orthodontics, Vol. 9, 1987, pp. 247-253.

Newman, W N., Sproull, R F., 'Principles of Interactive Computer Graphics', McGraw-Hill, London, 1979.

Noton, D., and Stark, L., 'Eye Movements and Visual Perception', Scientific American, June 1971, pp. 114-121.

O'Donnell, T., 'Gramps - A Graphical Language Interpreter for Real-Time, Interactive Three-Dimensional Picture Editing and Animation', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 15, No. 3, 1981, pp. 133-142.

Oppenheimmer, P., 'Real-Time Design and Animation of Fractal Plants and Trees', Proceedings ACM SIGGRAPH, Vol. 20, No. 4, 1986, pp. 55-64.

Ostry, D J., and Munhall, K J., 'Control Rate and Duration of Speech Movements', Journal of Acoustics, Society of America, Vol. 77, No. 2, February 1985, pp.640-648.

Patridge, B., 'The Structure and Function of Fish Schools', Scientific American, Vol. 246, No.6, June 1982, pp. 114-123.

Parke, F I., 'Computer Generated Animated Faces', Proceedings ACM National Conference, Boston, Vol. 1, 1972, pp. 451-457.

Parke, F I., 'A Parameterized Model for Facial Animation', IEEE Computer Graphics and Applications, Vol. 2, No. 9, November 1982, pp. 61-68.

Parke, F I., 'A Parametric Model for the Human Face', Technical Report UTECH-CSc-72-120, Dept. of Computer Science, University of Utah, 1974.

Parke, F I., 'Measuring Three-Dimensional Surfaces with a Two-Dimensional Data Tablet', Computer Graphics, Vol. 1, 1975, pp. 5-7.

Parke, F I., 'A Model for Human Faces that allows Speech Syncronized Animation', Computer Graphics, Vol. 1, 1975, pp. 1-5.

Paul, P., Robot Manipulators: Mathematics, Programming, and

Control, MIT Press, London, 1981.

Peck, S., *Atlas of Facial Expression*, Oxford University Press, New York, 1987.

Pearson, K., 'The Control of Walking', *Scientific American*, Vol. 10, No. 70, December 1976, pp. 72-86.

Perlin, K., 'An Image Synthesizer', Proceedings ACM SIGGRAPH, Vol. 19, No. 3, pp. 287-296.

Phong, B., 'Illumination for Computer Generated Pictures', Communications of the ACM, Vol. 18, No. 6, 1975, pp. 311-317.

Pisoni, D B., 'Speech Perception: Some new Directions in Research and Theory', *Journal of Acoustics, Society of America,* Vol. 78, No. 3, 1974.

Platt, S., A System for Computer Simulation of the Human Face, MSc Thesis, University of Pennsylvania, 1980.

Platt, S M., and Badler, N I., 'Animating Facial Expression', Proceedings Computer Graphics, ACM SIGGRAPH Vol. 15, No. 3, 1980, pp. 245-252.

Preston-Dunlop, V., *Practical Kinetography Laban*, MacDonald and Evans, New York, 1969.

Reynolds, C., 'Computer Anmimation with Scripts and Actors', Proceeding ACM SIGRAPH, Vol. 16, No. 3, 1982, pp. 289-296.

Reynolds, C., 'Flocks, Herds and Schools: A Distributed Behavioral Model', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 21, No. 4, 1987, pp. 25-34.

Sabin, M., 'A Survey of Contouring Methods', *Computer Graphics Forum*, Vol. 5, 1986, pp. 325-340.

Sederberg, T W., and Parry, S R., 'Free-From Deformation of Solid Geometric Models', ACM SIGGRAPH Proceedings, Vol. 20, No. 4, 1986, pp. 151-160.

Schmitt, F., Maitre, H., Clainchard, A., and Lopez-Krane, J., 'Acquisition and Representation of Object Surface Data', Biostereometrics SPIE, Vol. 602, 1985, pp. 42-49.

Shelley, K., 'Path Specification and Path Coherence', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 16, No. 3, 1982, pp. 157-166.

Siegel, H., 'An Overview of Computer Animation and Modelling', Proceedings Computer Graphics, London, 1987, pp. 27-37.

Smith, A., 'Spline Tutorial Notes', Technical Memo, No. 77, Lucas Film, 1983.

Synder, W., Industrial Robots: Computer Interfacing and Control, Prentice-Hall, International Editions, New Jersey, 1985.

Steketee, N., 'Parametric Keyframe Interpolation Incorporating Kinematic Adjustment and Phase Control', Computer Graphics Proceedings ACM SIGGRAPH, Vol. 19, No. 3, 1985, pp. 255-262.

Studec, 3D Videolaser, Sales Pamphlet, Licence ENST, Paris Cedex 11, 1987.

Sturman, D., 'A Discussion on the Development of a Motion Control System', Computer Graphics Proceedings ACM SIGGRAPH, Course Notes, No. 10, Computer Animation: 3D Motion Specification and Control, 1987, pp. 3-15.

Strauss, L., The Human Figure by Albert Dürer: The Complete Dresden Sketchbook, Dover Publications, New York, 1972.

Summerfield, Q., 'Analysis, Perception and Analysis of Visible Articulatory Movements', Journal of Phonetics, Vol. 11, 1974, pp. 63-76.

Summerfield, Q., 'Audio-Visual Speech Perception, Lip-Reading and Artificial Stimulation', Hearing Science and Hearing Disorders, 1986, pp.131-182.

Terzopoulos, D., Platt, J., Barr, and Fleischer, K., 'Elastically Deformable Models', Proceedings Computer Graphics and Applications, ACM SIGGRAPH, Vol. 15, No. 3, 1987, pp. 205-214.

Thomas, F., and Johnston, O., Disney Animation: The Illusion of Life, Abbeville Press, New York, 1984.

Thomas, F., 'Can Classic Disney Animation be Duplicated on the Computer?', Computer Pictures, July/August, 1984.

Thompson, D W., On Growth and Form, Abridged Version, Cambridge University Press, London, 1961.

Todd, J., Mark, L S., Shaw, R E., and Pittenger, J B., 'The Perception of Human Growth', Scientific American, Vol. 242, February 1982, pp. 106-114.

Tomovic, R., and McGhee, R B., 'A Finite State Approach to the Synthesis of Bioengineering Control Systems', IEEE Transactions on Human Factors in Electronics, Vol. 7, No. 2, 1966, pp. 65-69.

Vince, J, A., Computer Graphics for Designers, Pinter, London, 1985.

Vince, J, A., The PRISM System, Middlesex Polytechnic Campus, London, 1986.

Walker, G., and Kowalski, C., 'A Two-Dimensional coordinate Model for the Qualification, Description, Analysis, Prediction and Simulation of Craniofacial Growth', Growth, Vol. 35, 1971, pp. 191-211.

Waters, K., 'Expressive Three-Dimensional Facial Animation', Proceedings Computer Graphics Online, 1986, pp. 49-56.

Waters, K., 'Towards More Autonomous Control for Facial Animation', Proceedings British Computing Society, 1987.

Waters, K., 'A Muscle Model for Animating Three-Dimensional Facial Images', Proceeding Computer Graphics ACM SIGGRAPH, Vol. 22, No. 4, 1987, pp 17-24.

Warwick, R. and Williams P L., Gray's Anatomy, 35th Edition, Longman, London, 1974.

Weil, J., 'The Synthesis of Cloth Objects', Proceedings Computer Graphics, ACM SIGGRAPH, Vol. 20, No. 4, 1986, pp. 49-54.

Whitaker, H., Halas, J., Timing for Animation, Focal Press, London, 1982.

Wilhelms, J., 'Using Dynamic Analysis to Animate

Articulated Bodies such as Humans and Robots', Graphic Interface, 1985, pp. 107-114.

Wilhelms, J., 'Towards Automatic Motion Control', IEEE Computer Graphics and Applications, Vol.7, No. 4, 1987, pp. 11-22.

Wilhelms, J., 'Using Dynamic Analysis for Realistic Animation of Articulated Bodies', IEEE Computer Graphics and Applications, Vol. 7, No.6, 1987, pp. 12-27.

Yau, J, S., and Duffy, N., 'Texture Mapping Approach to 3-D Facial Image Synthesis', Proceedings Eurographics UK, pp. 17-30.

Zeltzer, D., 'Toward an Integrated View of 3D Computer Animation' The Visual Computer, Vol. 1, No. 4, 1984.

Zeltzer, D., 'Representation and Control of Three-Dimensional Computer Animated Figures', PhD Thesis Ohio State University, 1984.