

Optical Motion Capture for Performance-based Facial Animation

by

Anuoluwa Iyaniwura

A thesis submitted to the
Faculty of Graduate Studies and Research office
in partial fulfillment of the requirements for the degree of

Master of Applied Science in Electrical Engineering
Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario Canada.

May, 2008

© Copyright 2008 Anuoluwa Iyaniwura



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-40641-0
Our file *Notre référence*
ISBN: 978-0-494-40641-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

ABSTRACTS

Since its successful inception in early 1970s, facial animation has generated a lot of interest in fields including as virtual secretary in e-learning, as a representative in virtual meeting rooms, games industry, medicine and rehabilitation. Even though a lot of improvement has been achieved then, there are still challenges that need to be addressed. The innate ability of humans to detect any unrealistic expressions in facial animation and the complexities of the skin and muscle structures have presented a great challenge to both computer graphics and computer vision communities.

A realistic facial animation will be a useful tool in speech therapy to help those that have speech impairments. In this thesis, we propose method to directly retrieve both facial movements and facial physique with the aid of motion capture cameras and a laser head scanner. An effective facial marker set was modeled for the motion capture cameras to faithfully retrieve the mouth movements from three volunteers. These data were combined with the head scanner data to produce a rapid facial animation tool.

ACKNOWLEDGEMENTS

To my Heaven Father, the Almighty God, the Creator of the universe and Source of wisdom be the Glory and Praise and Adoration for His Strength, Grace, Help for seeing me through this project. Father, You did it again. Thank you very much Lord.

I thank my supervisors Professor A. Cuhadar and Professor C. Joslin for their help, guidance and advice. I am especially grateful for the motion capture sessions and the help with the setting up of the motion capture cameras. I am also grateful for the useful suggestions during the writing up of my thesis.

I also want to express my thanks to the nine volunteers, who willingly gave their time and attention for the motion capturing sessions and the three volunteers who were there for the head scanning.

I thank Professor W. Lee for the permission to use her lab and the Head scanner.

My special thanks goes to my beloved husband, Dr Timothy Iyaniwura, whose unfailing love, understanding, prayers, emotional support kept me through. I am also grateful to our boys, Oluwaloni, Enioluwa and Oluwatoni for their understanding and to my friend Rev. Pauline Scott for her constant prayers for me.

TABLE OF CONTENTS

Abstracts	iv
Acknowledgements	v
List of figures	ix
List of Tables	xii
List of Acronyms	xiii
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
1.3 Proposed Solution	4
1.4 Thesis Contribution	5
1.5 Organization of thesis	6
Chapter 2 Facial Animation Techniques	7
2.1 Model-based animation	7
2.1.1 Key-frame and geometric manipulation	8
2.1.2 Parameterization	9
2.1.3 Physics- Based Muscle modeling	10
2.1.4 Free Form Deformation	11
2.2 Image-based or Sample-based animation	13
2.3 Performance-based or data-driven animation	15

2.3.1 Snakes-----	15
2.3.2 Optical flow tracking -----	16
2.3.3 Motion Capture Systems -----	16
2.4 Evaluation of facial animation techniques -----	17
2.5 The proposed Method -----	18
Chapter 3 The Anatomy of the Head -----	20
3.1 Skull -----	20
3.2 Skin -----	21
3.3 Muscles -----	23
3.4 Muscles of mastication -----	25
3.5 Muscles of facial expression-----	28
Chapter 4 Facial animation process -----	30
4.1 Modeling -----	31
4.2 Rigging-----	35
4.3 Skinning -----	36
4.4 Animation -----	37
4.4.1 Vicon Motion Capture Camera -----	38
4.4.2 Facial markers models -----	39
4.4.3 Compilation of text-----	47
4.4.4 Motion capturing session -----	48
4.5 Rendering -----	56
4.6 Remarks-----	57
Chapter Five Results and Evaluation-----	60

5.1 Results-----	60
5.2 Evaluation-----	77
Chapter Six Conclusions and Future Work -----	78
6.1 Conclusions-----	78
6.2 Future Work -----	79
REFERENCES-----	81
APPENDIX A Anatomical based model-----	89
A.1 Head Structure -----	89
A.2 Skin/tissue -----	90
A.3 Muscles-----	90
A.4 Skull and Jaw -----	92
APPENDIX B Mathematical Models -----	93
B.1 Skin /Tissue Model -----	93
B.2 Muscle Model -----	95
B.3 Zhang et al's muscle model -----	98
B.3.1 Linear muscle model-----	98
B.3.2 Sphincter muscle model -----	99
B.3.3 Sheet muscle model -----	100

LIST OF FIGURES

Figure 2.1 Free form deformation-----	12
Figure 3.1 The basic plan of the skull-----	21
Figure 3.2 Skin layers-----	22
Figure 3.3 Strain/ Stress curve of dermal layer-----	23
Figure 3.4 Muscles structure-----	24
Figure 3.5 Front view of facial muscles-----	26
Figure.3.6 The side view of facial muscles-----	27
Figure 4.1 Facial animation process-----	30
Figure 4.2 Cyberware head and face scanner -----	32
Figure 4.3 Facial scanning process -----	33
Figure 4.4 The polygon mesh and texture -----	35
Figure 4.5 The polygon mesh and character rigging -----	36
Figure 4.6 Vicon motion capture cameras-----	38
Figure 4.7 Thirty facial marker set model-----	39
Figure 4.8 One hundred and twenty-eight facial marker set model-----	40
Figure 4.9 One hundred and twenty-eight facial marker set with segments -----	41
Figure 4.10 Fifty-four facial marker set model-----	42
Figure 4.11 Fifty-four facial marker set model with segments -----	46
Figure 4.12 Set-up of cameras for facial capture. -----	48
Figure 4.13 Arrangements of the six cameras -----	49

Figure 4.14 Facial capture session I -----	51
Figure.4.15 Facial capture session II -----	51
Figure 4.16 Fifty- four facial marker set -----	52
Figure 4.17 Facial animation with motion capture data workflow-----	53
Figure 4.18 Marker Set -----	54
Figure 4.19 Bones/marker set linking-----	56
Figure 5.1 Raw scanned data I-----	61
Figure 5.2 Raw scanned data II-----	62
Figure 5.3 Raw scanned data III-----	63
Figure 5.4 The processed scanned data I-----	64
Figure 5.5 The processed scanned data II-----	65
Figure 5.6 The processed scanned data III-----	66
Figure 5.7 The rigged polygon mesh -----	67
Figure 5.8 The skinned polygon with eyebrow envelope-----	68
Figure 5.9 Selected frames for thirty marker set model I-----	69
Figure 5.10 Selected frames for thirty marker set model II-----	70
Figure 5.11 Selected frames for fifty- four marker set model I-----	71
Figure 5.12 Selected frames for fifty- four marker set model II-----	72
Figure 5.13Movie picture I. -----	74
Figure 5.14 Movie picture II -----	74
Figure 5.15 Movie picture III-----	75
Figure 5.16 Combined video and result.I -----	75
Figure 5.17 Combined video and result II -----	76

Figure 5.18 Combined video and result III -----	76
Figure A.1 Mass-spring system-----	91
Figure A.2 Variety of supported muscles-----	92
Figure B.1 Triangular skin tissue prism element-----	93
Figure B.2 Relaxed Muscle fibre control polygon P -----	96
Figure B.3 Contracted linear muscle polygon q ($c=1/2$)-----	96
Figure B.4 Contracted sphincter muscle polygon q ($c=1/2$)-----	97
Figure.B.5 Linear Muscle model-----	98
Figure B.6 Sphincter muscle model -----	100
Figure B.7 Sheet muscle model-----	100

LIST OF TABLES

Table 2.1 Evaluation of facial animation techniques -----	17
Table 4.1 Comparison of the facial marker sets-----	45
Table 4.2 International Phonetic Alphabet -----	47

LIST OF ACRONYMS

AU Action Units

CMOS Complementary Metal Oxide Semiconductor

CSLU Center for Spoken Language Understanding

EFFD Extended Free Form Deformation

FACS Facial Action Coding System

FFD Free Form Deformation

FIX Feature-based Image Transformation

LED Light Emitting Diodes

RFFD Rational Free Form Deformation

3D Three-Dimensional.

2D Two-Dimensional

XML The Extensible Markup Language

CHAPTER ONE

Introduction

1.1 Motivation

The primary motivation for this thesis is the development of a method for realistic facial animation that can be used in speech therapy for people with speech impairments. Verbal communication is essential to social interaction, thus a tool that will enhance this will be of a great benefit to those with some form of speech defects. A combination of speech technology, computer graphics/vision along with virtual environment techniques will be valuable in the development of an appropriate tool in speech therapy, which will aid in alleviating the unique problems (interpersonal communications) of those with speech impairments.

Computer-aided devices for speech training and rehabilitation are available; however, as impairment is unique to each individual, the available device may not be suitable for all and at the same time meet individual needs. The devices also need to address the uniqueness of each disability, which depends on the cause of the disability (e.g. autism, multiple sclerosis) and the age of the patient, thus the available device may not be suitable for all. A computer based speech therapy tool will also be useful for the speech therapist who may have to deal with several patients with varying degrees of speech impairment. Furthermore, some people with speech impairment

have been found to be fascinated by computers [1], thus a computer-based speech therapy tool might increase the patient's motivation to go through the speech therapy session, and this can lead to improvement in therapy. There is also the need to incorporate the state of the art technology into such a system which makes the development of an effective speech therapy tool a most challenging and desirable undertaking.

1.2 Problem Statement.

Facial animation tool could be used effectively for human computer interactions, precisely as conversational agents for speech therapy applications. For example in CSLU Toolkit (Center for Spoken Language Understanding) Toolkit [2], 'Baldi', the talking head interacts with users. It is also made transparent so as to reveal tongue movements during speech production sessions. Turk and Arslan[3] in their Software tools for Speech Therapy and Voice quality monitoring used a talking head as human computer interface to provide a highly interactive and user friendly environment for both patient and the therapist. Thus a computer-aided speech therapy tool like CSLU (Center for Spoken Language Understanding) Toolkit composes of a facial animation system and speech synthesis/analysis /recognition along with suitable software development tools. CSLU Toolkit was designed to support basic research, development and education activities related to spoken language systems and human-computer interfaces. An individualized speech therapy tool needs to be single-purpose and simple to use by the intended users, hence there is a need for an

alternate solution which is the main objective of this thesis. An effective speech therapy tool will require an automated, real-time, and realistic facial animation system that can be adapted to individual faces [4]. A realistic facial animation system will further enhance interpersonal communications by reducing the tedious job of reading texts since it enhances communication medium to verbal level on the computer system. Since the early 70's, after Parke's pioneer work [5] on the development of "*Computer Generated Animation of Faces*" also known as the "*Talking Head*", a great deal of research effort has gone into producing a realistic facial animation system especially in the area of co-articulation, modeling and facial texture mapping. Parke used a parameterization technique to develop the "talking head". Since then, the challenge in this area has been to produce a convincing, realistic talking head, since the facial expressions and movements of a talking person are quite familiar to us, we can easily recognize any defects in the animated model. Moreover, the problems of a realistic talking head are further complicated by the diversity of the human facial features which is due to the unique bone and muscle sizes represented in various individuals. In addition, a talking head that can be incorporated into speech therapy will require the animation of real lip movements for speech articulation. Motivated by these facts, our goal in this study is to produce a computationally efficient, fast and realistic facial animation system that can be employed in speech therapy applications.

1.3 Proposed Solution

In this thesis, we propose to develop fast facial animation method that will exclude the tedious computations and complex algorithm which is associated with facial animation. We intend to use this facial animation for speech therapy applications in the future. We propose to use a combination of person –specific 3D model and data from motion capture camera to develop this facial animation method. The 3D model obtained from laser scanner will provide unique data with accurate facial physique, revealing all the details of the facial curves. In addition, the texture bitmap from the laser scanner is used for accurate texture mapping of the 3D model. Facial expressions and lip movements during speech are produced when the facial muscles which lie underneath the skin contract or relax. The complexity of contraction and relaxation of facial muscles requires the use of motion capture cameras to capture the subtle facial expressions created by facial muscles while the subject reads a predefined text. We intent to compile text, that incorporates the most common phonemes in English language. Thus, we propose to study the anatomy of the head especially the muscles, so that a facial marker set can be modeled in accordance to the location of these facial muscles, for effective capturing of the facial expressions and lip movements. We then investigate the effectiveness of this facial marker set model for animation of actors' faces.

The combination of the laser scan data and the motion capture data is expected to produce realistic facial animation tool for speech therapy applications.

1.4 Thesis Contribution

Facial expressions and speech movements are due to the contraction and relaxation of facial muscles, thus there is need to effectively capture the subtle movements of the facial muscles. Usually for motion capturing process, facial markers are placed at various locations on the face that reflects the locations of the underlying muscles so that the contraction and relaxation of the muscles can be captured. Having a facial marker set that will reflect the locations of all the facial muscles especially those around the mouth will be a step forward towards accurately retrieving the contraction and relaxation of the facial muscles, hence resulting in accurate facial animation. We defined a model for determining the locations of facial markers and evaluated its accuracy. First, we developed a 128 facial-marker model. However, due to the limited number of cameras available in the research laboratory, the processing of the data was tedious. The markers around eyelids and under eyelids swapped position in consecutive frames, a condition known as trajectory crossover. Then, the number of the markers was reduced to fifty-four. The fifty-four marker set was eventually used for motion capturing process. The motion capture data and laser scan data were combined to produce the facial animation tool. The main contributions of this thesis are:

- The modeling of a fifty-four facial marker set which effectively reflects the contraction and relaxation of facial muscles, especially around the mouth area. The effectiveness of the model was demonstrated by combining the motion capture data and the head model data to produce facial animation.

- The modeling of one hundred and twenty-eight facial marker set. Due to the limited number of cameras available, we found that the accuracy of the one hundred and twenty-eight facial marker set could not be evaluated. However, with more cameras, we expect that this experiment can be carried out in the future.
- Production of fast facial animation creation process by laser head scanner and motion capture cameras, which makes mathematical modelling of the talking head, a complex process, unnecessary.

1.5 Organization of the thesis

A brief introduction of the project is given in Chapter 1. Chapter 2 summarizes various facial animation techniques found in the literature. The anatomical structure of the head is discussed in Chapter 3, while Chapter 4 presents the major components of facial animation process that was undertaken in this thesis. The detailed process of facial scanning, motion capturing and the combination of these data in the development of the proposed facial animation system are also discussed in this chapter. Chapter 5 presents the results and Chapter 6 gives the conclusions and future work.

CHAPTER TWO

2. FACIAL ANIMATION TECHNIQUES

Numerous efforts have been made to improve on Parke's pioneer work [5] since he published his paper on facial animation in 1972. Its potential application as user interface in e-learning, as virtual secretary, as user assistance in web navigation has attracted a lot of attention from computer graphics and image processing communities [6]. In this chapter we will give a brief review of facial animation techniques existing in the literature. In general, facial modeling and animation techniques can be categorized as:

1. Model-based or
2. Image-based/sample-based
3. Performance based or data driven

In this chapter, we will review some of the approaches in these categories and discuss their advantages and limitations.

2.1 Model-based animation

This involves geometric manipulation with focus on shape; a 3D polygon mesh is first generated and then a human or cartoon-like texture is mapped on to the 3D mesh. Finally, the 3D structure is animated [7][8]. This is accomplished by first

identifying a number of control parameters for the 3D structure. Then, a suitable technique is used to animate the face model.

Range scanner, digitized probe and stereo disparity are capable of measuring the 3D coordinates; they can also be used for modeling 3D geometry of an individual face. The acquired data will require some post-processing so as to make it suitable for facial animation [9][10][11]. First, a prototype or generic animation mesh with all necessary structure and animation information is developed. The generic model is then fitted to the measured specific geometric mesh. Finally, a personalized animation model is created by fitting the generic model to the measured source (*i.e.* specific individual) geometric mesh. The transfer of texture is also facilitated by scanned face data along with the measured mesh. This is usually accomplished by using scattered data interpolation or radial basis function [12][13].

The geometric manipulation techniques include: key-frame and geometric manipulation, parameterization, physics-based muscle modeling and free form deformation.

2.1.1 Key-frame and geometric manipulation

The key-frames (*i.e.* the major frames) of facial positions are stored in a database while the in-between frames of facial positions are realised by using appropriate interpolations. Typically, a linear, cosine or bilinear interpolation function is employed to generate the motion between the key-frames at extreme positions over normalized time intervals [14] [5]. The 2D or 3D positions of face mesh vertices are directly updated for animation purposes. Its main advantage being its fastness and

ease of generating primitive facial animation, however it is limited in creating wide range of realistic facial configuration thus making it difficult to produce a combination of independent face motions (i.e. facial expressions and head movements) .[4]

2.1.2 Parameterization

Due to the limitations of key-framing techniques, the parameterization technique was introduced by Parke in 1972. Typically, combinations of independent parameter values based on conformation (or structure) and the expression (or emotional) content of the face are defined for any possible modeling and animation of the face so that specific facial configuration can be explicitly controlled. For example, a facial feature like the mouth is shaped by specifying the lip opening height and the width of protrusion [5][15].

Facial Action Coding System (FACS) is a system for representing facial expressions in terms of Action Units (AU), which comprises of one or more muscles, and their associated actuation levels *i.e.* the contraction and relaxation of these muscles. FACS was developed by Ekman and Friesen and it was introduced into facial animation systems to widen the range of facial expressions that could be simulated [16]

A wide range of facial expression can be achieved with the combination of parameter values at a relatively reduced computational cost. However, conflicts between these parameter values usually result in unrealistic expressions or configuration. For this reason parametric models are usually employed for specific

facial regions. Finally, the model requires tedious manual tuning to achieve realistic facial animation [8].

2.1.3 Physics- Based Muscle modeling

Physics based muscle modeling take into consideration the various properties, such as structure and behaviour (*i.e.* the movement or transformation) of the bone, the muscles and the various layers of the skin [17][18][19]. In [20], the authors proposed to use the facial muscle and structure of the human face. They used the mass spring elastic laws based on Hooke's spring law (*i.e.* pertaining to elasticity) to model thirty eight regional muscles. Muscle forces propagate through the elastic spring mesh to deform the skin. In [21], the author modeled muscles as vectors with direction and magnitude in both two and three dimensions. The direction is referenced towards the origination of the muscle which is the point of the attachment; the magnitude is minimum (zero at the origin or point of attachment) and maximum at the insertion point. He used ellipsoid parameterization to model the sphincter muscles located around the mouth and successfully animated human expressions like fear, anger, surprise, disgust, joy and happiness. The model required trial and error manual positioning of the muscles which could result in undesirable effect if not properly adjusted.

Waters and Terzopoulos [21][22] modeled the face as layered spring meshes comprising of three layers of deformable mesh the skin, fatty tissue and muscles which originate from the bone to faithfully reflect the anatomical structure and

dynamics of the face. Each layer and node is connected by elastic elements. Facial animation was accomplished by the propagation of muscles forces through the elastic spring network of muscles and the skin. Facial animation result was a big success however the stimulation of volumetric deformation in a three dimensional lattice required extensive computations. Lee *et al.* improved on this model by incorporating the non linear visco-elastic property of the dermal layer to produce a model with greater accuracy [23]. In [24], the authors used the finite element networks to model the visco-elastic properties of the skin. Kahler *et al.* improved on Lee's muscle model by assigning ellipsoidal shape to each muscle segment and inserted muscles into simulated skin tissue [25][26]. The muscle geometry was defined by zone of influence and its deformation was accomplished purely by geometric manipulation. The main advantage of the model-based technique is its provision for greater flexibility in construction of head models in any given view, lighting conditions and deformation. However, it requires more computation due to the level of complexity of its models. Furthermore, its texture mapping is not real in other words, the appearance of the resulting facial animation is not realistic.

2.1.4 Free Form Deformation

In Free Form Deformation (FFD) an object is deformed indirectly by deforming the volumetric object that encloses it. The object to be deformed is embedded in a 3D cubic lattice which is defined as a Bezier volume. *i.e.* volume with a number of control points or vertices. As the control points or vertices of the lattice are being manipulated, the embedded object deforms along with it. Figure 2.1

illustrates the Free Form Deformation. The embedded spherical object is to be deformed, and the second figure shows the deformation. [27]

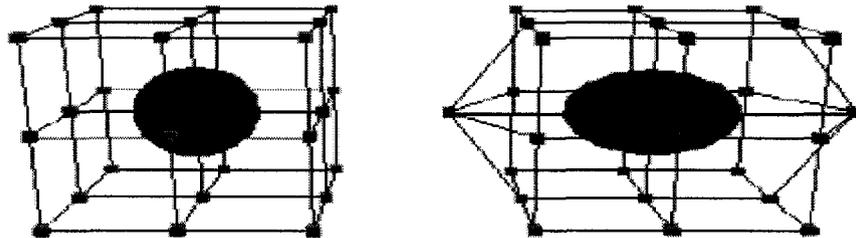


Figure 2.1 Free Form Deformation.[4]

The Extended FFD (EFFD) uses non-cubical lattice providing greater flexibility for object deformation.[28] Rational Free Form Deformation (RFFD) incorporated weight factors to each control point thus increasing the degree of freedom for shape deformation. Thus deformation is accomplished by changing the weight factors instead of control point positions. Kalra *et al.* utilized this method to simulate the facial muscle action on facial skin [29]. They defined the surface region that corresponds to the anatomical description of the muscle actions and parallelepiped (*i.e.* a 3D figure formed by six parallelograms) control volume on the region of interest. Facial muscle actions were stimulated by interactively displacing the control points by changing the weights associated with each control points. The

advantage of free form deformation approaches is that they provide a form of transition that is independent of the surface specifics [4].

2.2 Image-based or sample-based animation

Image-based animation systems focus on texture and utilize image processing techniques for producing a photo realistic talking head [30][31][11]. Some transformations (rotation, scaling and translation) are applied to multiple-view model of the head so as to create a desired view. In this technique, to achieve facial animation, video images of facial parts are recorded and then rendered directly at image level. The system captures both the articulation of mouth and its transient complex behaviour in time. The appropriate model is obtained directly from the captured samples of data which comprises of both video and audio of an actor during performance (e.g. articulation of mouth while talking) [32][33][34]. The methods [6][35] for accomplishing this:

1. Recording of a limited set of visemes (a viseme is a visual equivalent of a phoneme) , usually less than one hundred, as key image samples while the in-between transitions are synthesized by employing image processing techniques. Ezzat *et al.* [36] used optical flow algorithm to automatically compute correspondence from every viseme to every other viseme. A smooth transition between the viseme images was generated by morphing along this corresponding. Morphing is the process of changing from one form of image to another until the desired form is realised. The model was improved on by utilising a variant of

Multidimensional Morphable Model (MMM) [31] to produce entirely different sentence based on what was learnt from the recorded image samples. The system also a trajectory synthesis module which synthesizes mouth trajectories with correct motion, smoothness, dynamics and co-articulation effects. Input streams of phonemes and their respective frame durations are mapped to a trajectory of MMM shape-appearance parameters. Finally this trajectory is fed into the MMM for the synthesis of the final visual stream that represent the talking face.

2. Alternatively, all possible visemes are recorded and the required visemes are automatically retrieved from the large database of visemes for the synthesis of the required facial animation [37]. Video Rewrite used computer-vision technique for tracking points on the speaker's mouth in the training footage. These mouth gestures were then combined into the final video sequence by employing morphing techniques.
3. For FIX: Feature-based Image Transformations [38] approach, used normal 2D images to created realistic personalized face animation. Facial features and their corresponding translations for each face activity were used to reduce the required data, which made scaling and the processing of the new character possible. Different visual effects (*i.e.* talking, various facial expressions and head movements) in a given image are created from the learned set of transformations.

Image-based systems usually require a large database of 2D images, except the feature-based Image transformation approach which used minimum input data to create realistic images; however the output is a photo realistic talking head [6]. It is

not flexible for the generation of pose variations as this will require visemes of different poses to generate multi-view facial animations and a very large image corpus database.

2.3 Performance-based or data-driven animation

Performance based or data driven animation involves the recordings of motion capture data, or video sequences. The actors/actresses' speech movements are captured by motion capture and or video cameras. This was introduced in order to overcome the animation control problem such that the actor's action can be directly employed in the animation of the model by generating and transferring the motion data into the model [39]. This is usually accomplished by the following methods.

2.3.1 Snakes

A snake is an energy minimizing spline that deforms to fit local minima given the initial location. Snakes are used in tracking motion of facial features (like mouth, or eyes) in video sequences primarily based on color samples and edge detection [40]. Dynamic muscle parameters can also be estimated from snake state variables [22][23]. Main drawback of this approach is the accumulation of tracking error over large range of image sequences may result in a situation where the contour to be tracked is lost.

2.3.2 Optical flow tracking

“Optical flow is a concept which approximates the motion of objects within a visual representation” [41]. Optical flow tracking is employed in tracking facial expression or for speech recognition from video sequences [39][42][11]. Usually coloured markers are glued on the actor’s face while speaking or performing some expressions [39][43]. Its limitations include high computational costs and its sensitivity to head movements.

2.3.3 Motion Capture Systems

Specialized cameras known as motion capture cameras are used to record the facial movements of actors or individuals while speaking or displaying various facial expressions. More accurate animation is achieved by this process [10] Types of motion capture systems include

- **Mechanical:**

A skeletal-structure gadgets known as exoskeleton is worn by the subject so that the body joints angles are directly tracked while he/she performs an act.

- **Electromagnetic (magnetic):**

The system uses electromagnetic waves to measure both the position and orientation of the subject within a tracking volume.

- **Optical:**

Retro-reflective markers are attached to the subject's body at well defined positions while the system captures the movements of the markers as the subjects perform some acts. Some facial markers are also attached to the actor's face while performing.

2.4 Evaluation of facial animation techniques

Table 2.1 shows the comparison of both Model-based and Image-based facial animation techniques in tabular form.

Characteristics	Model-based	Image based
Automatism	High	Low
Flexibility	High	Low
Realism	Medium	high

Table 2.1 Evaluation of facial animation techniques

Automatism reflects model creation and synthesis of animation. Flexibility encompasses characteristics such as synthesis speed, face model rendering from arbitrary views and user defined texture. Finally realism reflects its closeness to real human face [6].

Other systems are of hybrid nature [38]. They combine both model based and image-based techniques to improve on the limitations of each technique while combining their advantages to generate realistic face animation with increased flexibility. As suggested by Noh and Neumann, a perfect facial animation must be automated, real-time, realistic and adaptable to individual faces [4].

2.5 The proposed method

Considering the complexity of facial muscles contraction and relaxation beneath the skin layer, direct capturing of such movements can produce more accurate facial expressions lip movements during speech. With the uniqueness of individual facial physique, a person-specific model has the ability to produce a more real head model taking into consideration the specific facial features. In view of these reasons, we propose to use a laser scanner for person-specific model along with motion capture data to produce a computationally efficient and realistic facial animation. In order to capture comprehensive facial expressions, we ask the subject to read out a text that includes the forty-four phonemes in English language. Facial motion capturing will require an effective facial marker set to reflect the location of the facial muscle, so that the movements can be captured while the subjects are reading the text aloud. Steps to accomplish this goal include the following:

1. Develop a model of an effective facial marker set.
2. Compile a text that will incorporate the forty-four major English phonemes.
3. Perform laser scanning of the subjects' faces

4. Use six motion capture cameras to capture facial expressions for the three subjects (two females and a male) as they read aloud the compiled paragraph of sentences.
5. Simultaneously use video camera to record the facial motion capture process.
6. Analyze the data from the motion capture cameras, the facial scanner and combine the two data together for facial animation.
7. Compare the resulting facial animation with the original video.

CHAPTER THREE

3. The Anatomy of the Head

Facial expression and lip movements during speech are due to the relaxation and contraction of muscles. A sound understanding of the anatomy of the head and the mechanism of the skull, muscles and skin is required to be able to produce facial animation. Since the facial marker set is to be modeled according to the location of the muscles, we will give a brief review of the anatomy of the head in this chapter. Basically, the head comprises of the skull, skin and the muscles which lie between the skull and the skin. Muscles are known to originate from the skull (*i.e.* firmly attached to the skull) and insert into the skin. Therefore, for physics-based animation, the skull edge of the muscle is fixed while the skin edge is deformed (shortened or lengthened) depending on the contraction of the muscles [44].

3.1 Skull

The skull comprises of various sized 28 bones [45][46] jigsaw together but generally assumed as comprising of three main skeletons. This is illustrated in Figure

3.1 The three main skeletons are:

- *Cranium:* Domed shaped, rigid top that encloses the brain and its membranes.

- *Facial Skeleton*: This is attached to the front part of the cranium's base. It comprises of the nasal bone, upper jaw (also known as the maxilla) and others.
- *Movable Mandible (also known as lower jaw)*: The lower jaw is attached to the rear side of the cranium's base [45][46].

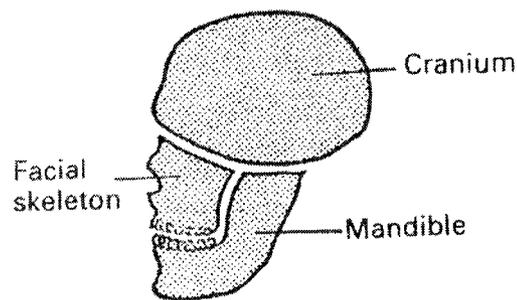


Figure 3.1 The basic anatomy of the skull [45]

3.2 Skin

Figure 3.2 shows the cross-section of the skin. The skin comprises of two main layers which are

- *Epidermis*: Comprises of dead cells and it is as thick as a tenth of dermis layer's thickness. It is the outside layer of the skin.
- *Dermis*: It has three types of tissues which are elastin, collagen and reticular fibers. *Elastin fibers* constitute 4% of dry weight of the dermal tissue and have elastic properties. While *collagen fibers* constitute 70% of dry weight of the dermal tissue is made up of tough insoluble protein. [47]

Beneath the epidermis and dermis layers are the subcutaneous fat layer (also known as the hypodermis layer) and the connective layers which comprises of larger blood vessels and nerves.

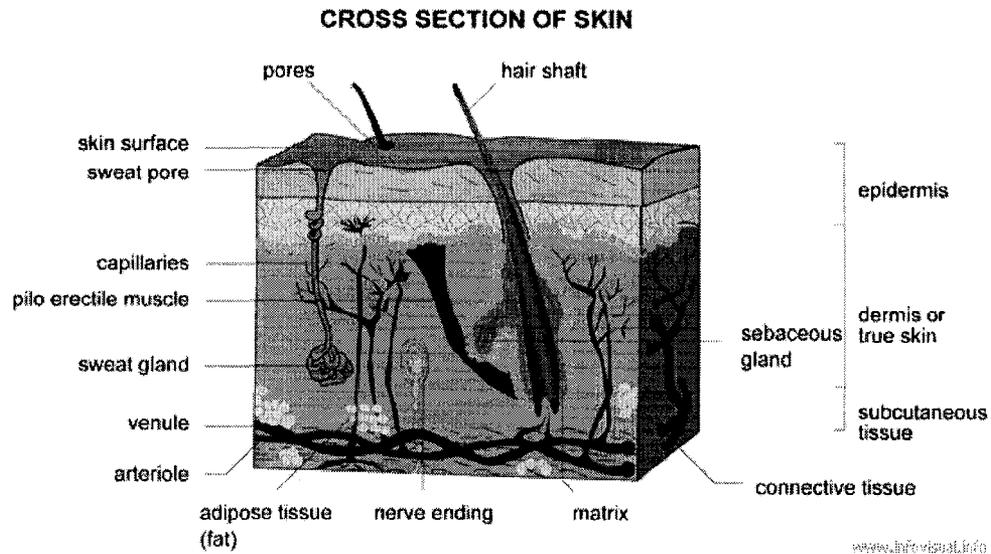


Figure 3.2 Skin layers [48]

The biomechanical properties of the skin is mainly influenced by the dermal layer due to the viscoelastic behaviour of the both the elastin and collagen fibers. From the strain/stress curve of Fig.3.3, at low stress, the collagen fibers uncoil in the direction of the strain with low resistance to stretch, however at higher stress, the collagen fibers are fully uncoiled presenting a higher resistance to stretch. This produces an approximate biphasic stress-strain curve [49]. At zero load the elastin fibers with its elastic properties return the collagen fibers to their original coiled condition [22][50]. Thus with these observed properties, the skin is concluded as

being anisotropic with non linear stress/strain properties and it displays viscoelastic properties or time-dependent behaviour. [23][44]

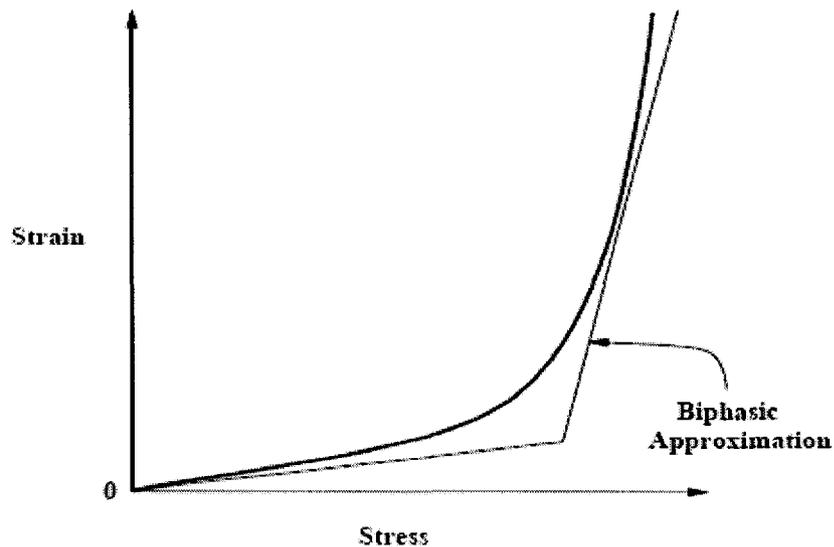


Figure 3.3 Strain/ Stress curve of dermal layer [22]

3.3 Muscles

As mentioned earlier on muscle layers lie between the bone structure and the skin. Facial muscles comprise of cylindrical shaped muscle fibers or cells; about 15cm by 10-100 μ m in size. Each muscle fiber is made up of a thread-like *myofibrils* (1-2 μ m wide), which in turn comprises of *sarcomeres* , the basic functional units with the cell. The fibers are connected together by connective tissues which link to nerves

and blood tissues as shown in Figure.3.4.[45][46] Facial muscles are classified into two groups: muscles of mastication and muscles of expression according to their functions either for mastication or expressions, while some muscles are involved in both functions. Muscles of mastication are responsible for chewing, and grinding of food while muscles of expression are involved in verbal and non-verbal communications. Muscles originate from (that is firmly attached to) the bone structure and insert into the skin [51]. Thus, when muscles contract the bone edge of the muscle is fixed or static while the skin edge moves depending on the type of the muscles. Facial muscle types include rectangular, triangular, sheet, linear and sphincter [17][21][18]

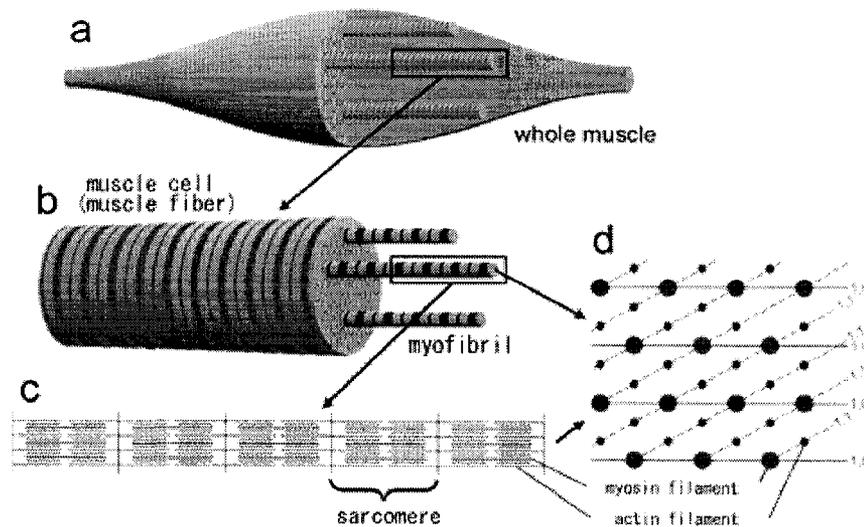


Figure 3.4 Muscles structure [52]

3.4. Muscles of mastication

Muscles of mastication include

- The *Masseter* which originates from the Zygomatic arch and inserts in to the mandible.
- *Temporalis*, originates from the temporal bone and inserts in to the mandible.

Both muscles are used for biting and closing of the jaw [45][46]. Figure 3.5 shows the front view while Figure 3.6 shows the side view of facial muscles.

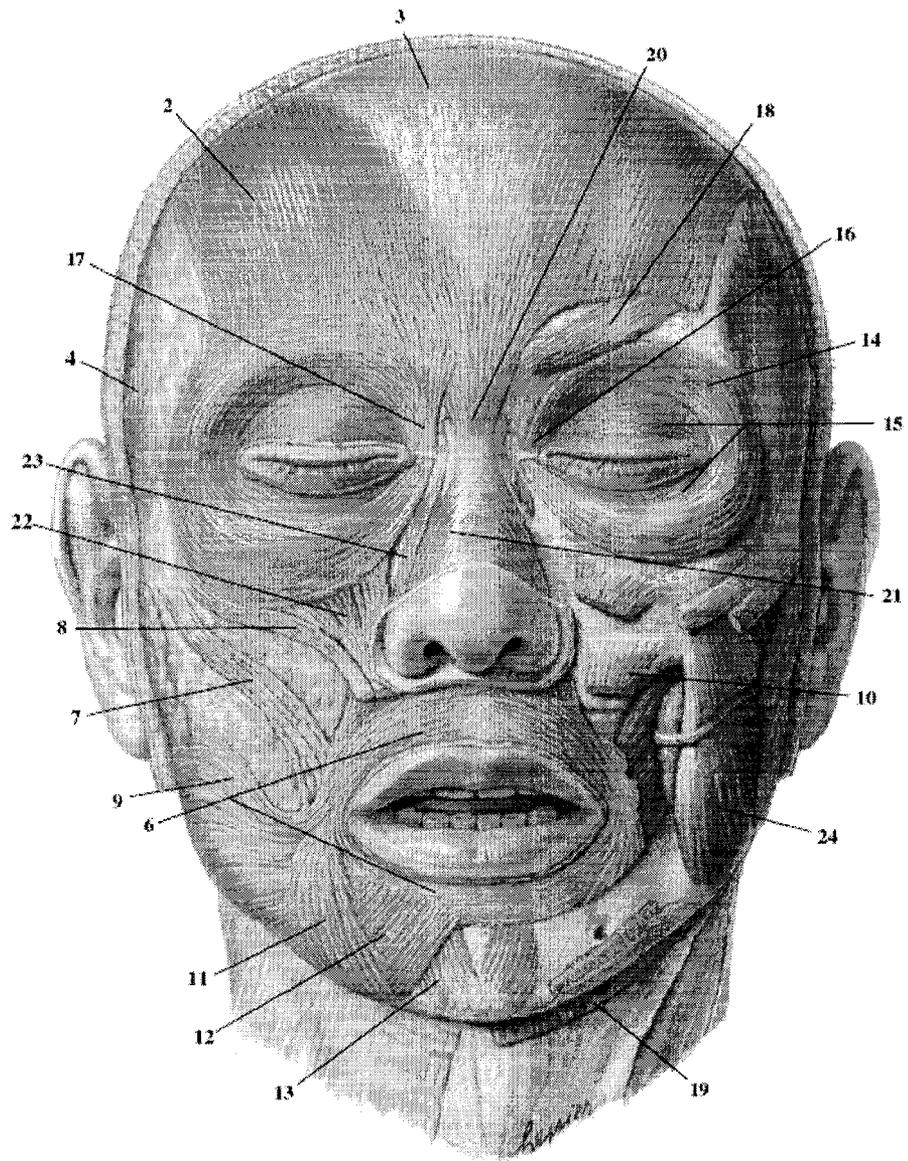
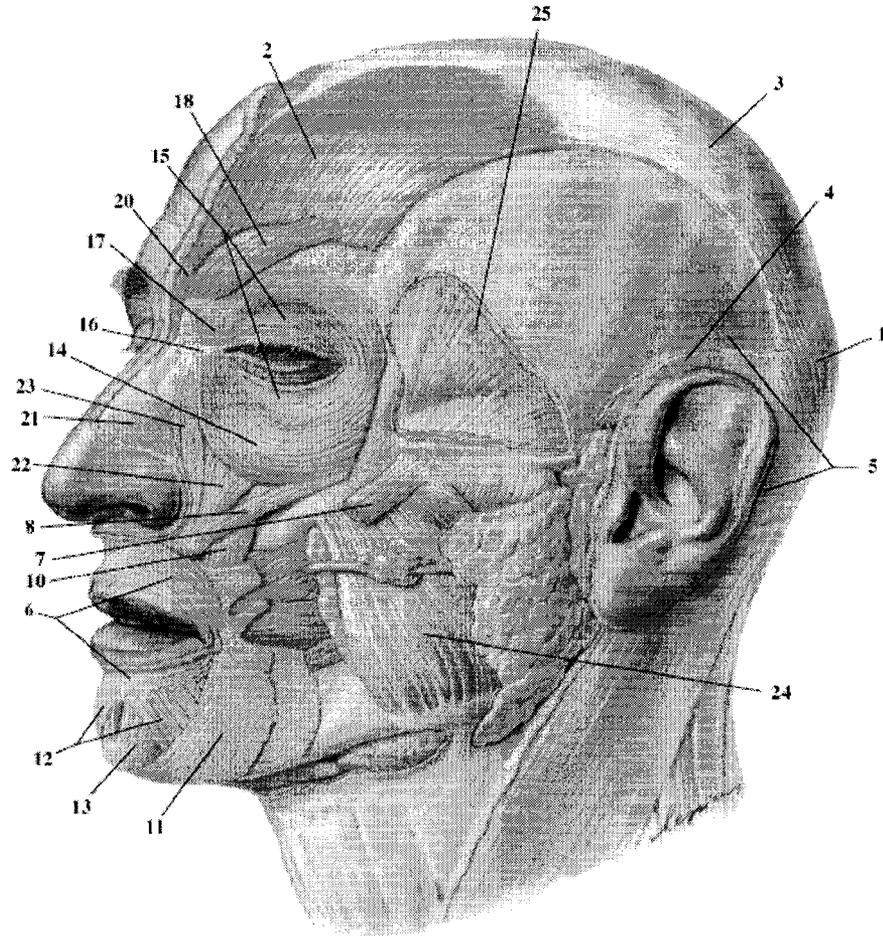


Figure 3.5 Front view of facial muscles [53]



<i>scalp</i>			<i>mouth</i>			<i>eyelid</i>		
1	occipitalis	6	orbicularis oris	14	o. oculi, p. orbitalis			
2	frontalis	7	zygomaticus major	15	o. oculi, p. palpebralis			
3	epicranial aponeurosis	8	zygomaticus minor	16	palpebral ligament			
4	temporoparietalis	9	risorius	17	depressor supercilii			
5	auricularis	10	levator anguli oris	18	corrugator supercilii			
		11	triangularis					
		12	depressor labii inferioris					
		13	mentalis					
<i>neck</i>			<i>nose</i>			<i>m. of mastication</i>		
19	platysma	20	procerus	24	masseter			
		21	nasalis	25	temporalis			
		22	levator labii superioris					
		23	l. labii sup. alaeque nasi					

Figure 3.6 The side view of facial muscles. [53]

3.5 Muscles of facial expression

Facial expression muscles are categorized according to their region of influence as shown in Figure 3.5. and Figure 3.6 into scalp, neck, mouth, nose, eyelid and muscles of mastication.

- i) Scalp: The *frontalis* and the *occipitalis* combines together as *occipitofrontalis*. It originates from the occipital and inserts in to the eyebrow tissues involved in raising eyebrows and the horizontal forehead wrinkles. [54][55]
- ii) Mouth: For this we have *oribicularis oris* which encircles the mouth with no bone attachment. This lies between the skin and the lip mucous membrane and aids in the closure and puckering of the lip.
 - *Buccinator* originates from maxillae and inserts into the skin of the sides of mouth; it lies in cheek walls and aids in chewing by holding food in contact with teeth and facilitates smiling
 - *Zygomaticus major and Zygomaticus minor* originate from the zygomatic bone and inserts into the mouth corners; it is used for laughing by elevating the angle of the mouth. Also used for laughing is the *risoris* which pulls the mouth corner backward horizontally; it originates from the fascia of masseter muscle.
 - The *levator anguli oris* inserts into to the *oribicularis oris* at the mouth's corner and is responsible for raising the upper corner of the lips.
 - *Mentalis* pulls the chin down, originating from just below the teeth and inserts into the chin ball skin. [55]

- iii) Eyelid: The *orbicularis oculi* encircles the eyelid and comprises of the thick *pars orbitalis* which runs around the orbit and the *pars palpebralis* which lies beneath the upper and the lower eyelids. [54][46].
- iv) Neck: Platysma originates from the fascia of pectoral and deltoid regions and inserts in to the mandible and several muscles. It is responsible for the lowering of the lower jaw, lip and the movement around the neck. [56][45]
- v) Nose: Nasalis has two parts known as the Compressor naris and Dilator naris. Compressor naris originates from the upper jaw near the canine tooth and inserts into the nasal cartilage on the nose bridge while the Dilator naris originates from the upper jaw and nose cartilage and inserts into the nostril skin.

The technique for anatomical based models also known as the physics based models is discussed in appendix A while its derived associated mathematical models for facial animation is discussed in appendix. B

Understanding of the anatomy of the head, especially the knowledge about the location of facial muscles is essential for the modeling of facial marker set. In the next chapter, we will describe the various steps that we followed in the development of the proposed facial animation tool.

CHAPTER FOUR

4 FACIAL ANIMATION PROCESS

Our facial animation process was accomplished by incorporating the anatomical structure of the face. As mentioned earlier, the anatomical head comprises of the skull,

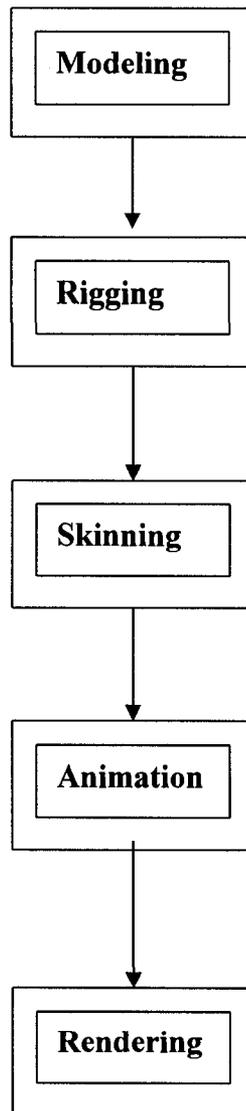


Figure 4.1 Facial animation process.

bones, muscles and the skin. Based on this structure, the procedure we used for the development of the facial animation tool includes the following stages: modeling, rigging, skinning, animation, and rendering as illustrated in Figure 4.1. The objective of modeling process is to produce the shape of the skull. During the rigging process we attempted to form the muscles on the skull and finally we placed the skin on as the top layer. The detailed process of accomplishing this is presented in the following discussions.

4.1 Modeling

The initial stage in facial animation is the modeling of the head to reflect the general structure of the head with all the facial features (that is the eyes, mouth, nose, eyelids and eyebrows.). The base layer of the head anatomy is the skull; in technical terms this should be the polygon mesh. The head model is a person-specific 3D model developed for each actor. In this study we used three volunteers: two females and a male. Cyberware 3030 Color Scanhead is employed to scan the volunteers' faces.

Cyberware 3030 Color ScanHead comprises of a low-intensity laser light source, aluminized, dielectric-coated first-surface mirror assembly, CMOS sensing optics and color sensors and supporting electronics. Figure 4.2 shows a photo of the Cyberware 3030 Color ScanHead.



Figure 4.2 Cyberware Head and Face Scanner.

During the scanning process, lighted profiles are created as low-intensity laser and shone on the subject as shown in Figure 4.3. These profiles are captured by the video sensors which are then digitized by the system to produce the complete shape of the head of the subject while the color sensors register the RGB color information. These data are processed by the attached graphics workstation to produce the range and texture data. The whole process takes 16 seconds for each scan. The output data is further processed to remove noise and it is then smoothed and converted into a suitable format (object files). The head scanner has workstation attached to it to process these data. These routines are capable of removing noise, smoothing the polygon mesh and capable of converting the mesh to a suitable format. The choice of object files was due to its popularity among many graphics software. For each volunteer, six different poses (neutral, happy, angry, sad, smile and surprised) are scanned. Each subject had to maintain these poses in turn for the entire scanning

period. Figure 4.3 shows a volunteer during the scanning process for smile pose. The scanned data (i.e. the polygon mesh and the texture bitmap) were imported into 3DS MAX software for further processing.

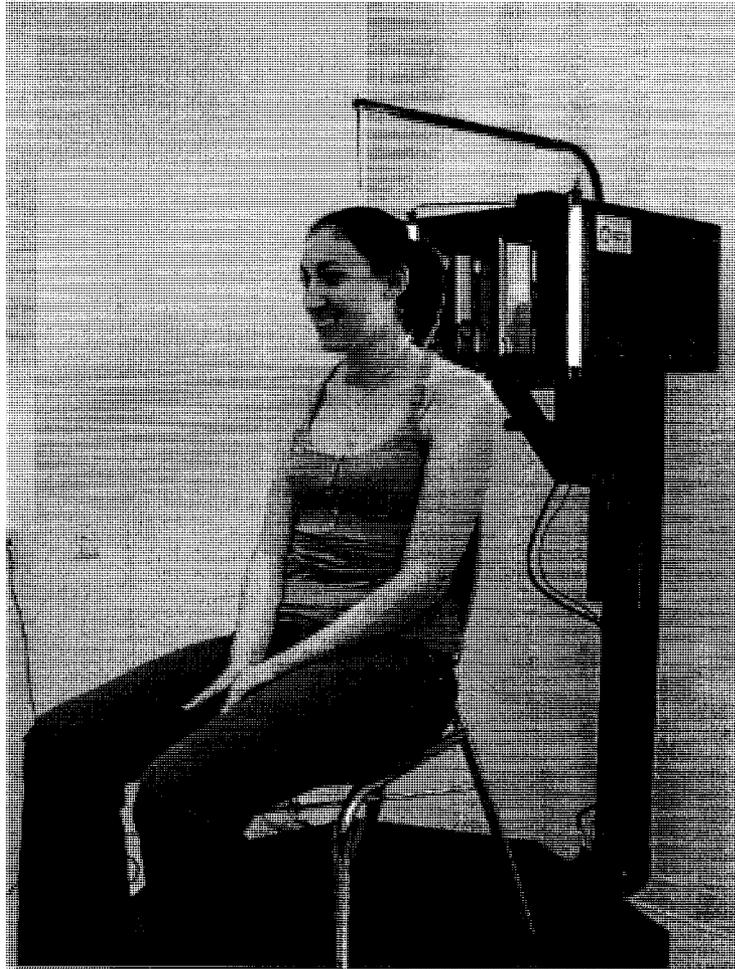


Figure 4.3 Facial scanning process.

3DS MAX is a computer graphics tool for modeling, animation and rendering.

The process of using 3DS MAX involves the following stages:

- First, the 3D head model (i.e. the scanned data) is imported
- The Optimize Modifier routine of 3DS MAX was then applied to the high polygon 3D head model to reduce the number of faces and vertices to ten percent (10%) of the original polygon mesh in order to simplify the geometry of the head model and speed up the rendering process while maintaining an acceptable texture. The texture of the reduced polygon mesh is still comparable with the original texture of the polygon mesh. A low polygon model is especially desirable for animation since the model has little geometry which can be animated in real-time.
- With the Material Editor, the texture bitmap file was imported on to the scene and applied to the head model. Material Editor is an interface within the 3DS MAX for defining, creating and applying texture to the model

Figure 4.4 shows the screen shot of the reduced imported 3D model with the texture applied.

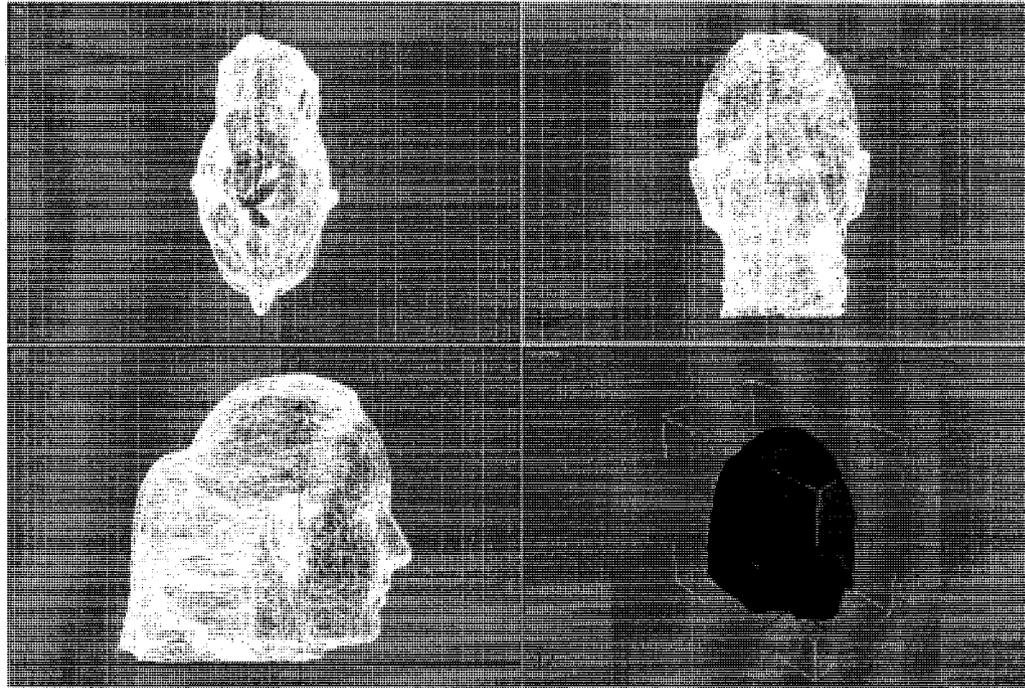


Figure 4.4 The polygon mesh and texture.

4.2 Rigging.

The next stage is the skeleton rigging during which the facial bones/muscles were created. Figure 4.5 shows a rigged face. As mentioned previously, the contraction and relaxation of facial muscles produces facial expressions and speech movements. Altogether twenty six bones were created in the following areas:

- Sixteen bones around the mouth to mimic the *orbicularis oris* muscles for the movements of the mouth.
- Four bones at the eyebrow to mimic *oculi*. *Orbitalis* and *depressor supercilii* muscles.

- Four bones at the cheek to mimic *zygomaticus major* and *minor* muscles.
- Two bones at the chin to mimic the *depressor labii inferioris* muscles.
- Finally two bones around the nose area to mimic the *levator labii superioris* muscles.

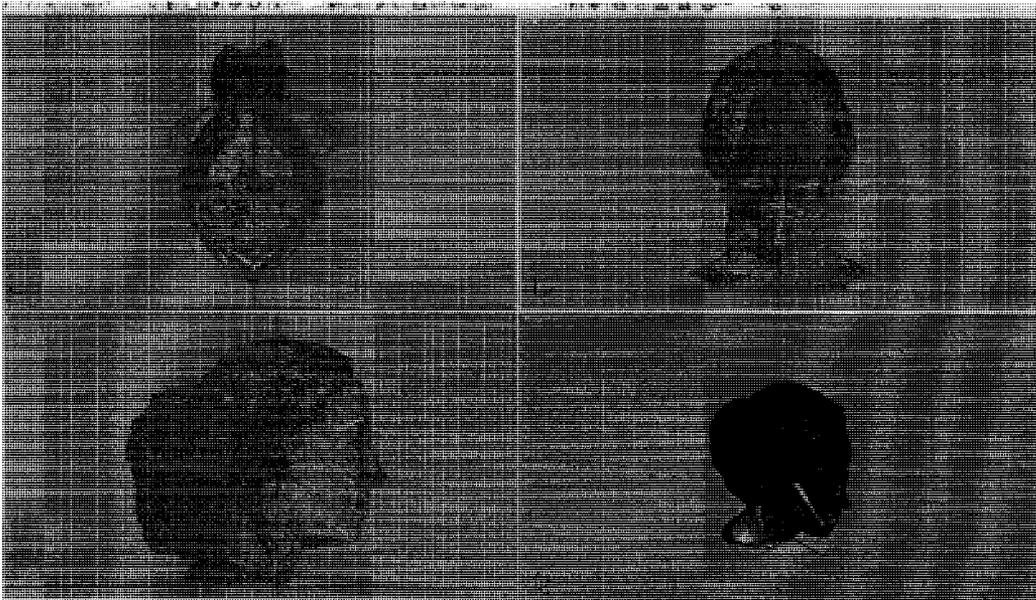


Figure 4.5 The polygon mesh and character rigging

4.3 Skinning

Anatomically, the skin movement is as a result of facial expressions and speech movements which are mainly due to the contraction/relaxation of facial muscles. Similarly, the head polygon mesh is deformed to produce facial expression and speech movements in accordance to the rotation and position of the muscle objects. Thus during this process, the movements of the muscles within the head

mesh is defined in such a way that the mesh moves as the muscles relax/contract. The head mesh is connected to the muscles and it is allowed to deform as the muscles are animated. The Skin Modifier, a routine in 3DS MAX, attempts to associate (i.e. bind) the vertices of the head mesh to one or more muscles so that the facial expression is controlled by the contraction/relaxation of the muscles. The muscles are arranged in a hierarchy depending on their area of influence.

4.4 Animation

Developing a realistic facial animation is a challenging problem. This is due to the fact that facial physique is unique to each person and depends on the muscles size and bones, gender, race and age. Nevertheless, the advent of motion capture camera, which is capable of retrieving subtle facial expressions have attracted the attention of many in computer graphics community for solving this problem[43][10]. In order to produce realistic facial animation, we have to emulate real-life motion by simulating muscles contraction/relaxation. This is accomplished by using the motion capture cameras to capture the actual physical movements of the facial muscles while the subject speaks. The motion capture data is then translated into computer usable form. In our study, we had access to six Vicon Motion Capture cameras for capturing the facial expressions and speech movements of the subjects while they were reading aloud a given text that contained the major phonemes in English language.

4.4.1 Vicon Motion Capture Camera.

As shown in Figure 4.6 the Vicon Motion Capture Camera uses surface-mount near infra-red Light Emitting Diodes (LEDs). Motion capture cameras are capable of capturing full body (*i.e* the whole person) as well as any part of the body. In this study, we are using the motion capture cameras to capture the facial expressions and speech (lips) movements only. Accurate capture of subtle facial expressions requires high resolution in order to capture subtle movements of facial features such as lips, eyelids, and eyes. The capture volume (*i.e.* the area to be captured by the cameras) and the size of the facial markers to be used on the subject should be small compared to the full body capture so that the facial expressions can be captured accurately.

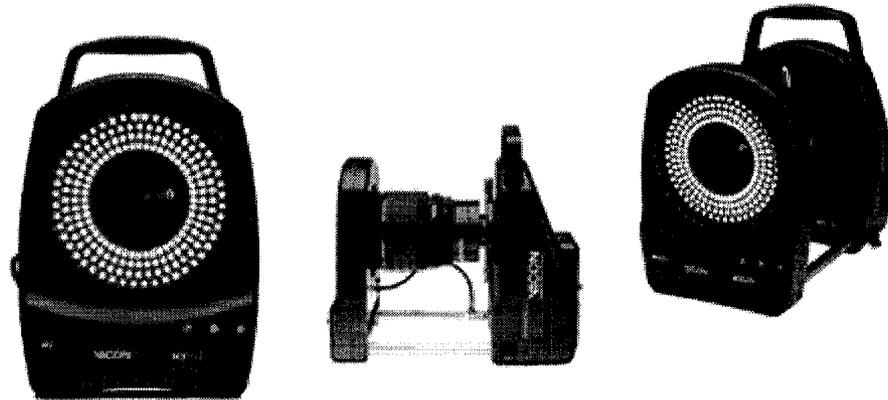


Figure 4.6: Vicon motion capture cameras [60]

4.4.2 Facial markers models

Initially, we used the 30 facial marker set along with 4mm retro-reflective markers for seven volunteers as shown in Figure 4.7. The 30 facial marker set has only seven markers around the mouth for detecting the orbicularis oris muscle movements, only two markers for the chin and some few markers around the cheek. As shown in Figure 4.7, the 30 marker set is not a truthful representation of the facial muscles and thus did not faithfully capture the subtle facial expressions, so a 128 facial marker set of Figure 4.8 was modeled for improved capturing of the facial expressions and lip movements.

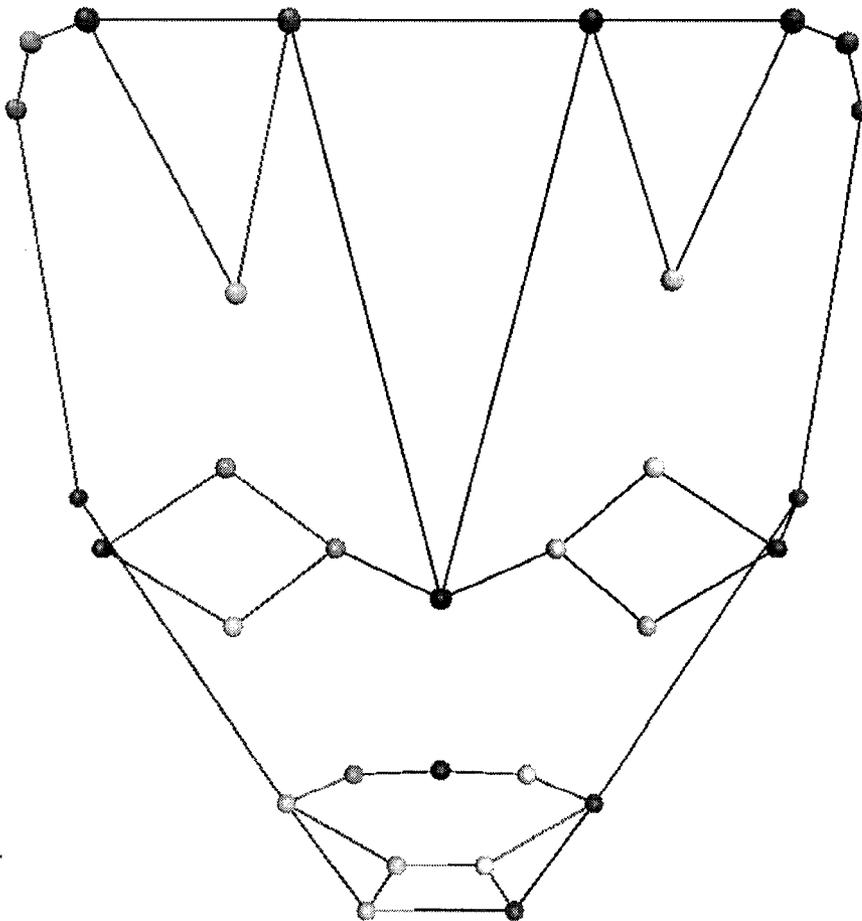


Figure 4.7 Thirty facial marker set model.

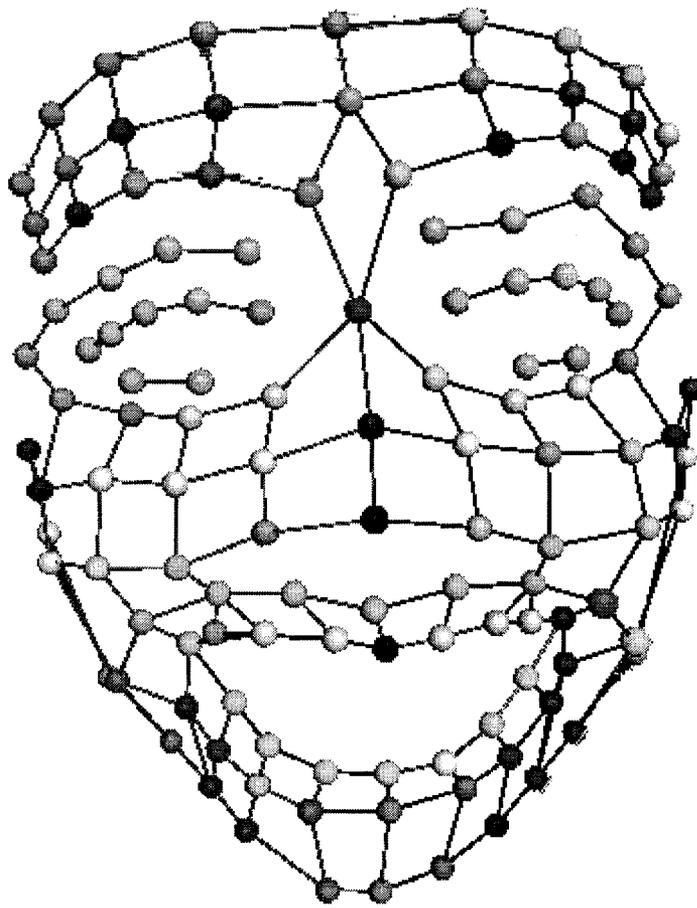


Figure 4.8 One-hundred twenty-eight facial marker set model.

The modeling of the facial marker set involves the creation of a musculoskeletal model; an XML file which contains the musculoskeletal model of the human head. The model comprises of chains of segments (that is bones), markers that reflect the location of facial muscles and “sticks” for joining the markers together. This is illustrated in Figures 4.9 and Figure. 4.11.

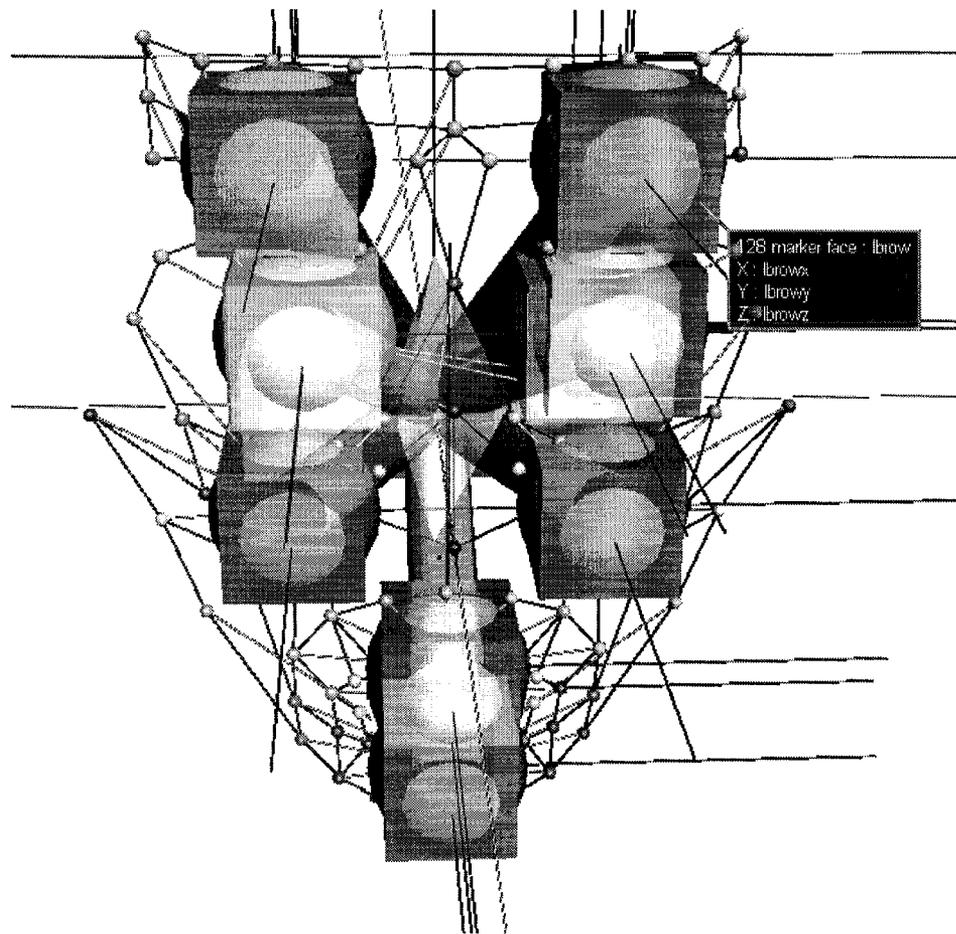


Figure 4.9 One-hundred twenty-eight facial marker set with segments

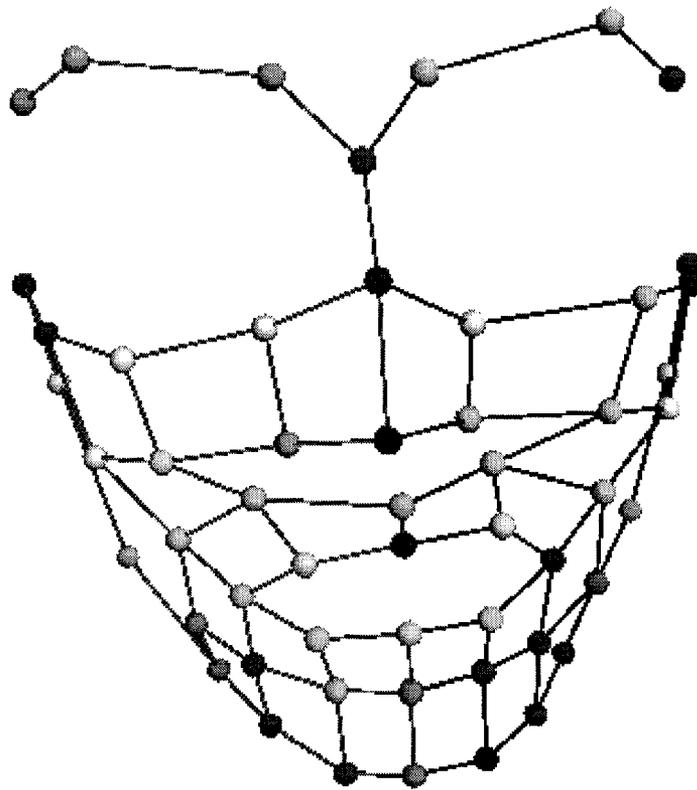


Figure 4.10 Fifty-four facial marker set model.

The segments were first defined to reflect the head (root segment), upper mouth, mid-mouth, jaw, cheek, forehead, chin, moustache, eyebrow and eyelid. Attached to these segments were the markers and finally the markers were joined together by “sticks”. We tried the animation using 128 facial-marker set model, but the post-processing of the motion capture data for this set was tedious as the number of the motion capture cameras used was not sufficient to accurately capture the subtle movements of the dense markers on the face. Due to the density of the markers, the system could not accurately trace the trajectories of each marker for the number of the available frames. There was the problem of trajectory crossovers around the eyelids and the bottom of the eyes, a situation that occurred when the system swapped markers for difference frames. An increased number of motion capture camera, say ten, would be able to capture the locations of these markers accurately and thus could prevent the problem of trajectory crossovers. Therefore, we decided to reduce the number of the facial markers to fifty-four to accommodate the six motion capture cameras available in the research laboratory. A fifty-four facial-marker set as shown in Figure 4.10 was modeled to reflect the major facial muscles (*orbicularis oris, zygomatic major, zygomaticus minor etc*) and the motion capture data obtained for three subjects. These data were post-processed and used for facial animation.

Table 3.1 compares the three facial marker set model. For the thirty facial marker set model, *epicranial aponeutosis, teporoparietalis, risorius, levator anguli oris, depressor labii inferioris, mentalis, oculi, p, palpebalis, palpebral ligament, depressor supercillii, procerus, labii.sup. alaequsi nasi* and *masseter* muscles are not represented. For fifty-four facial marker set model, *frontalis, epicranial aponeutosis,*

temporoparietalis, oculi, p, palpebalis, palpebral ligament, depressor supercillii are not represented. These are fewer than the thirty facial marker set model. All of the facial muscles are represented in the one hundred and twenty- eight facial marker set model.

Facial muscles	Thirty facial marker set (number of markers)	Fifty-four facial marker set (number of markers)	One hundred and twenty eight facial marker set (number of markers)
frontalis	4	0	8
Epicranial aponeurosis	0	0	2
teporoparietalis	0	0	4
Orbicularis oris	7	11	22
Zygomaticus major	4	4	8
Zygomaticus minor	2	4	4
risorius	0	4	4
Levator anguli oris	0	2	4
triangularis	2	4	8
Depressor labii inferioris	0	4	4
mentalis	0	4	6
o.oculi. p.orbitalis	4	2	14
Oculi,p palpebalis	0	0	8
Palpebral ligament	0	0	2
Depressor supercilii	0	0	2
Corrugator supercilii	2	2	6
procerus	0	1	1
nasalis	1	2	2
Levator labii superioris	2	2	4
Labii.sup.alaeque nasi	0	2	4
masseter	0	4	5
temporalis	2	2	6

Table 41 Comparison of the facial marker sets

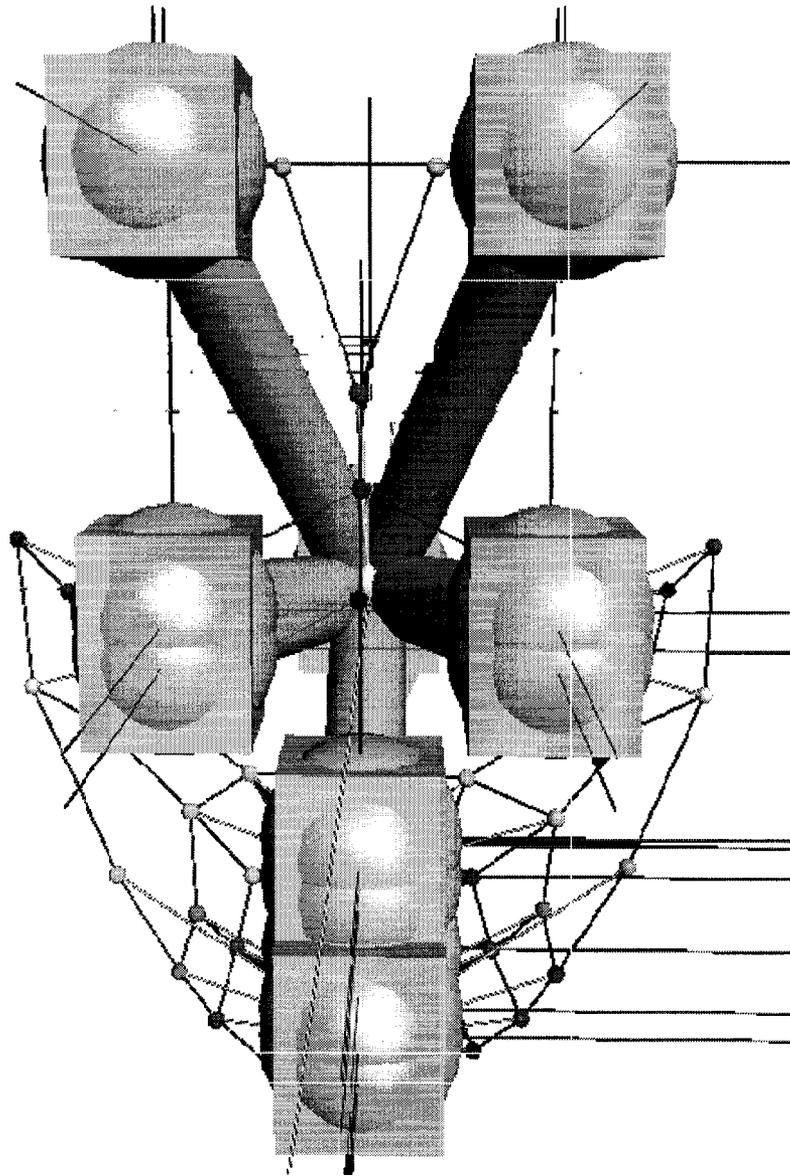


Figure 4.11 Fifty-four facial marker set model with segments

4.4.3 Compilation of text

There are forty-four major phonemes in English language which comprises of twenty vowels and twenty- four consonants. The text that we used for this study was compiled from the International Phonetic Alphabet which is shown in Table 4.1. The text incorporated all the forty-four major English phonemes.

vowels			consonants		
IPA	ASCII	examples	IPA	ASCII	examples
ʌ	^	cup, luck	b	b	bad, lab
a:	a:	arm, father	d	d	did, lady
æ	@	cat, black	f	f	find, if
ə	..	away, cinema	g	g	give, flag
e	e	met, bed	h	h	how, hello
ɜ:	e:	turn, learn	j	j	yes, yellow
ɪ	i	hit, sitting	k	k	cat, back
i:	i:	see, heat	l	l	leg, little
ɒ	o	hot, rock	m	m	man, lemon
ɔ:	o:	call, four	n	n	no, ten
ʊ	u	put, could	ŋ	N	sing, finger
u:	u:	blue, food	p	p	pet, map
aɪ	ai	five, eye	r	r	red, try
aʊ	au	now, out	s	s	sun, miss
oʊ əʊ	Ou	go, home	ʃ	S	she, crash
eə	e..	where, air	t	t	tea, getting
eɪ	ei	say, eight	tʃ	tS	check, church
ɪə	i..	near, here	θ	th	think, both
ɔɪ	oi	boy, join	ð	TH	this, mother
ʊə	u..	pure, tourist	v	v	voice, five
			w	w	wet, window
			z	z	zoo, lazy
			ʒ	Z	pleasure, vision
			dʒ	dZ	just, large

Table 4.2 International Phonetic Alphabet [57]

The text used for the motion capture session was:

“Little black tourist met Lady Yellow getting out of bed; hello Miss Rock, he said to the girl. Could you please check this large vision cinema? Call the man at five past eight, neither give him ten pieces nor let him touch the map with his

fingers. Turn left if you crash near the blue window at the zoo. Father, try not to hit the cat with both hands. Could you please heat up the red cup and then join the air balloon class in the lab. Do go home when you are done.”

4.4.4 Motion capturing session

The motion capture session included the following steps:

- The cameras were arranged in a semicircle of radius 1.5m around the subject as shown in Figure 4.13. This ensured that the slightest movements of the markers are captured. The cameras were arranged such that the heights of three of the cameras were set to view the subject’s face from above while the other three were set to view slightly below the horizontal mid line of the subject’s face as shown in Figures. 4.12 and 4.13.

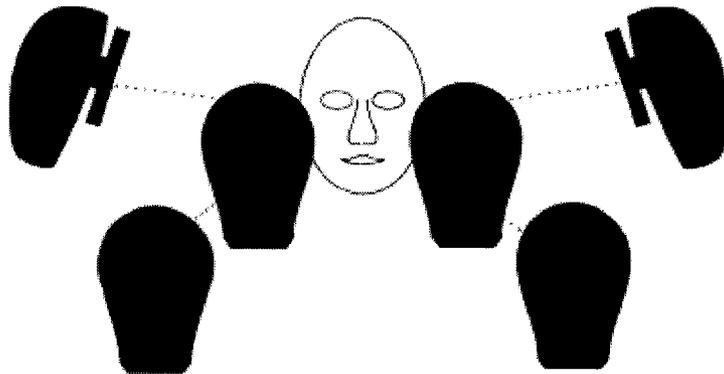


Figure 4.12 Set-up of motion capture cameras for facial capture [58]

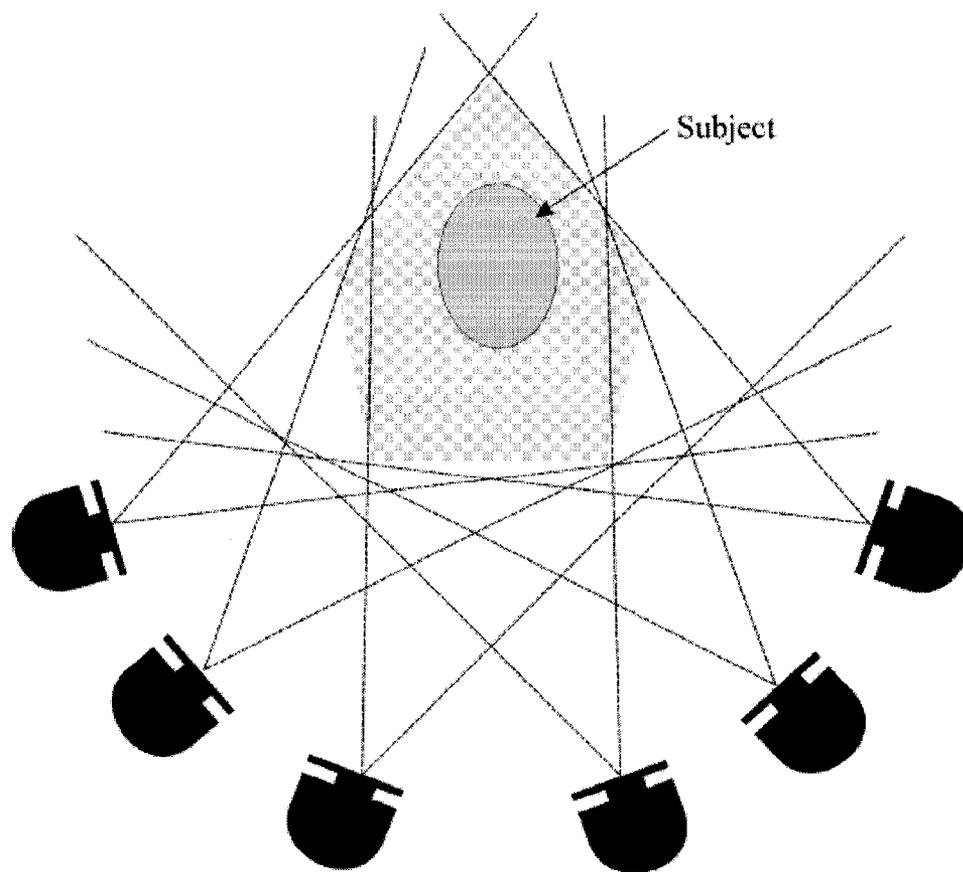


Figure 4.13 The positions of the six cameras.[58]

- The cameras were first focussed. Then the capture volume was cleared of any reflective object for accurate capturing.

- The system was then calibrated dynamically and statically. The dynamic calibration enables the system to define the capture volume and the relative positions and orientations of the cameras. The gathered information during this process was later used in the reconstruction of the markers, i.e. in the creation of the 3D motion of the markers. The static calibration was required for setting the origin and the direction of the axes.
- Facial markers were then attached to the subject's face, and the following routines were carried out
 - Range of Motion – This involves the exaggerated performance of the subject for all possible facial expression (anger, smile, happy etc).
 - Auto-labelling of the markers and the calibration of the subject.
 - Then each subject's facial expression and speech movements were captured as they read the text aloud.

The motion capture session was also recorded by a video camera. The audio part of the video clip was extracted and combined with the processed facial data. Figures 4.14 and 4.15 show a typical set up for a facial capture session whereas Figure 4.16 shows a volunteer with fifty-four facial markers placed on his face.

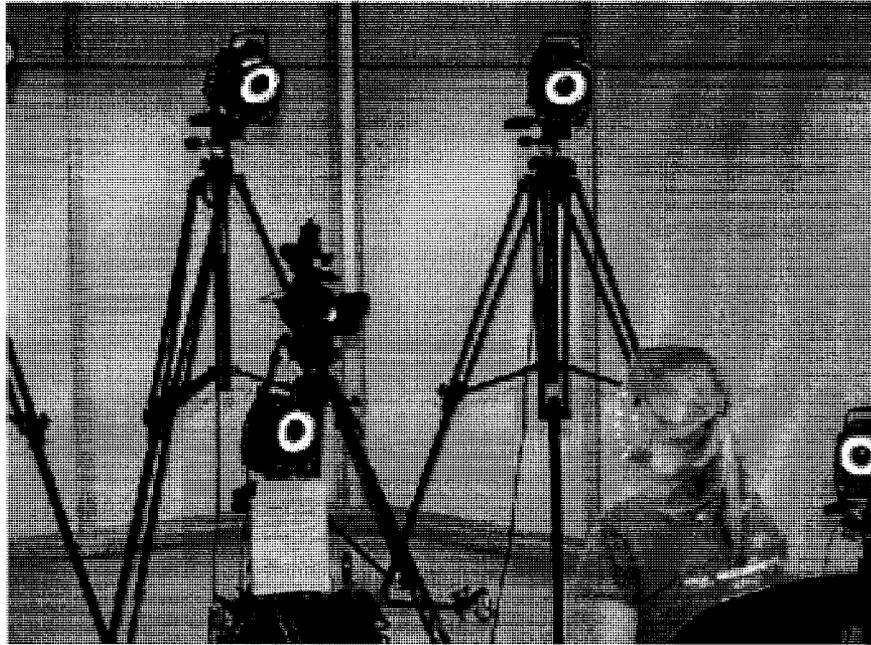


Figure 4.14 Facial capture session I

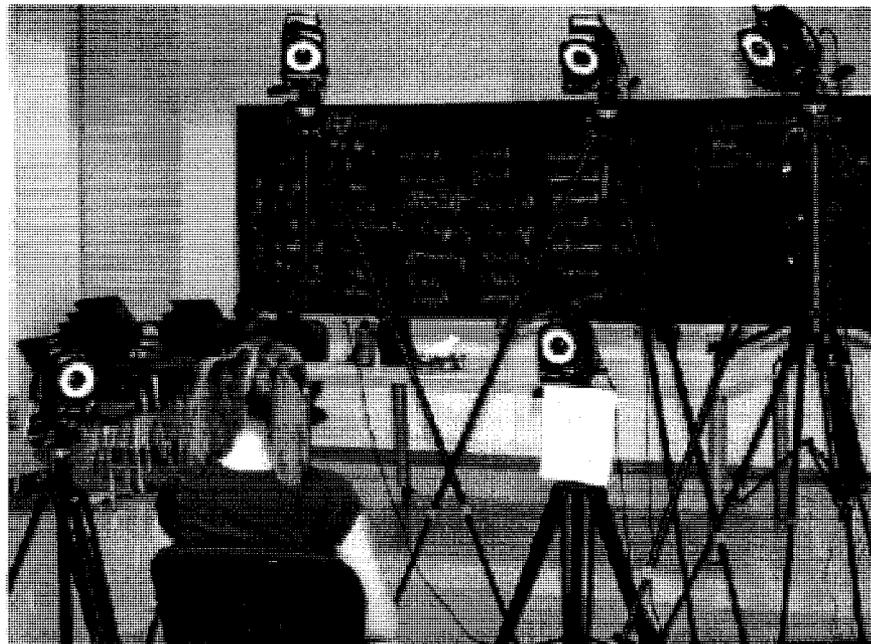


Figure 4.15 Facial capture session II



Figure 4.16 Fifty- four (54) facial marker set

The motion capture data were processed and exported to MotionBuilder Software for further processing. The skinned and rigged polygon mesh was also exported to MotionBuilder software, a character animation software. The motion captured data and the model were combined together for facial animation. In MotionBuilder the motion capture data was first mapped into an Actor face assets, then generic expressions were defined in the character face settings. As shown in Figure. 4.17, this was then used to animate the face model.

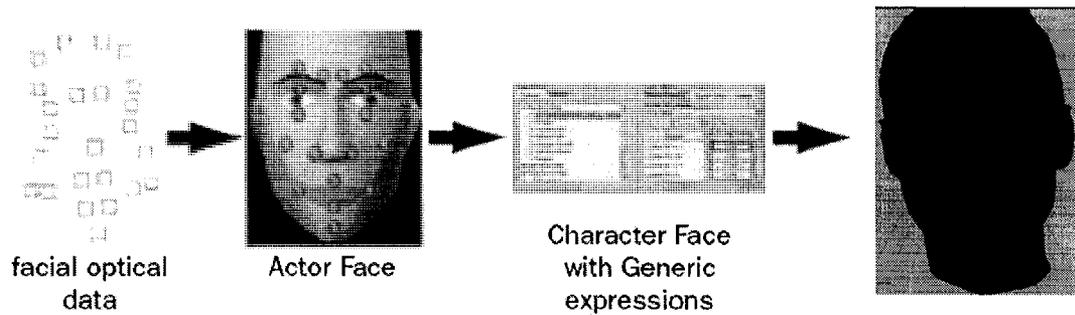


Figure 4.17 Facial animation with motion capture data workflow [59]

The process of producing the facial animation in the MotionBuilder involves the following steps:

- The motion capture data (.c3d file) was imported into the MotionBuilder.
- The Actor face was dragged on to the scene.
- A face reference was created with the markers on the right temple, left temple and the central head.
- A set of markers (known as marker set) were defined for specific locations of the head as follows:

1. right outer eyebrow
2. right inner eyebrow
3. left outer eyebrow
4. left inner eyebrow
5. right cheek
6. left cheek

7. right nostril
8. left nostril
9. right mouth
10. left mouth
11. upper mouth
12. lower mouth
13. chin

This is illustrated in Figure 4.18. This marker set was used to connect the optical data to the Actor face. The motion capture data was then plotted on to the Actor face.

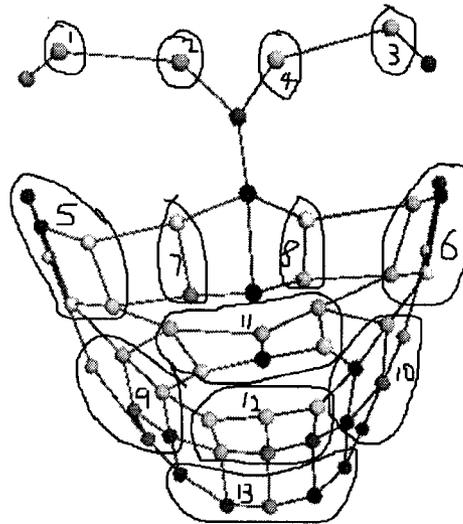


Figure 4.18 The marker set.

- The Head model was then imported and the Character face was set as follows:

- The Character Face was added on to the scene.
- Cluster groups of bones were created for the mouth, left eyebrow, right eyebrow, nostrils, right cheek and the left cheek.
- With these cluster groups, shapes were created for the
 - Mouth (open, whistle, left corner up, right corner up, left corner up, right corner down and square shapes).
 - Right eyebrow (up and down shapes).
 - Left eyebrow (up and down shapes)
 - Right cheek (up and out shapes).
 - Left cheek (up and out shapes).
 - Nostrils (left up and right up shapes).
- These shapes were then linked to the generic expressions.

Figure 4.19 illustrates the linking of the bones to the marker set. Bones were linked to the right outer eyebrow, right inner eyebrow, left outer eyebrow, left inner eyebrow, right cheek, left cheek, left cheek, right nostril, right mouth, left mouth, upper mouth, lower mouth and the chin. Same pattern as for the marker set of the Actor face. The Character face will then receive the motion capture data from the Actor face and combines it with the defined expression to animate the head model.

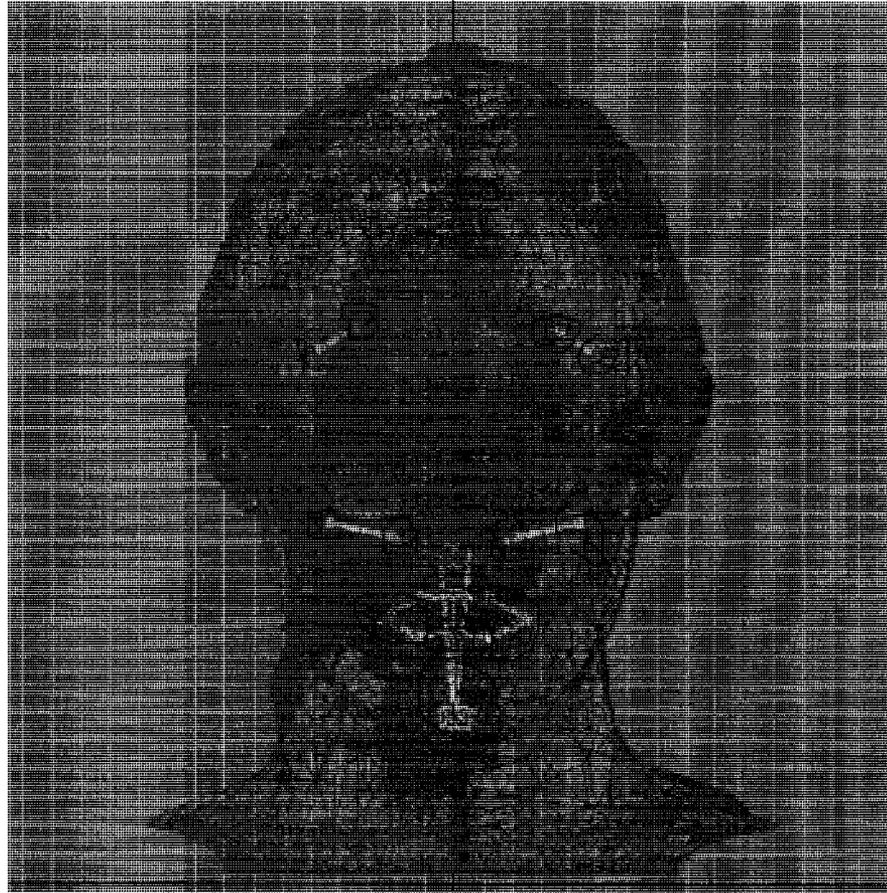


Figure 4.19 Bones/marker set linking

4.5 Rendering

The final step is rendering, during this stage the model, textures, lighting and animations are combined together into final sequence of images. The pixels for the diverse colors are created to form a complete image. For better results, the files were first processed as sequence of image files. Eight hundred and seventy five (875), eight hundred and nineteen (819) and seven hundred and seven (707) frames of images (.tiff files) were produced for the three volunteers. Furthermore these were combined

together along with the extracted audio for the production of a movie clip with the aid of Adobe Premiere, a video and film editing and creating application software as explained below:

- Imported the frames of still images, audio files into the Adobe Premiere application software.
- The recorded video was also imported into the Adobe Premiere.
- These files were assembled in the program sequence interface of the software application.
- The combined output program sequence was rendered and save as (.avi files)
- These movies were also compared with the original video by assembling the still images and the recorded video in the program sequence

4.6 Remarks

In this chapter, we described our approach based on the head anatomy for facial animation. This approach requires 3D head model to reflect the skull, rigging process to fix in the muscles, and skinning process to add skin to the head model. During animation process the issue of relaxation and contraction of the muscles is addressed. The final rendering process combines the images, lighting and animation together for the production of the movies.

We used Head Scanner to acquire person-specific 3D head model. The faces of three volunteers were scanned. The resulting data were further processed to remove noise. The resulting files of polygon mesh and texture were exported to 3DS

MAX, a computer application software. First, the polygon mesh was reduced to 10% of its original size. The texture was then applied using the Material Editor.

The next step is the creation of bone objects to mimic the muscles, then these bones were attached to the skin, so that the skin could move as the muscles “contracts and relaxes”. The motion capture data was then used to animate the bones. The resulting head model with muscles, and texture was imported to MotionBuilder so that the animation data from the motion capture cameras could be combined for animation of the face.

Animation data were acquired by using motion capture cameras to retrieve facial expression volunteers, while they read a paragraph of sentences which comprises of the forty-four major phonemes in English. Motion capturing requires facial marker set to reflect the location of the facial muscles. I modeled a hundred and twenty-eight and fifty-four facial marker set for this purpose and evaluated the accuracy of the fifty-four facial marker set. The fifty-four facial marker set model was used for capturing the facial expressions and lip movements of the subjects while they were reading the paragraph of sentences. A video recording of the motion capture session was done simultaneously. The data was analysed and exported to MotionBuider.

The motion capture data file and the polygon mesh were imported to MotionBuilder. The motion capture data was transferred to Actor Face. This is connected to the Character Face which defines the generic expressions on the 3D model so as to animate the model.

Finally the model, texture, animation and lighting were combined together to produce frames of still images. These were exported to Adobe Premiere along with the data from the video recorder for editing and final movie. The video recording made during motion capturing process was also combined together so that a comparison could be made between the original video and the processed movie (*i.e.*, the final facial animation). In the next chapter, we will discuss our facial animation results.

CHAPTER FIVE

5. RESULTS AND EVALUATION

In chapter 4, we described our approach for facial animation by using the modeling, rigging, skinning, animation and rendering processes. In this chapter we will discuss our results and address the problems encountered during this study and solutions to some of these problems.

5.1 RESULTS

The results for modeling, rigging, skinning, animations and rendering processes are presented below:

- **Modeling-** Figures 5.1, 5.2 and 5.3 show the raw scanned data while Figures 5.4, 5.5 and 5.5 show the processed scanned data. Comparing these set of figures, one can see the noise in the raw scanned data. After processing the resulting scanned data are a lot clearer and smoother than the raw scanned data.

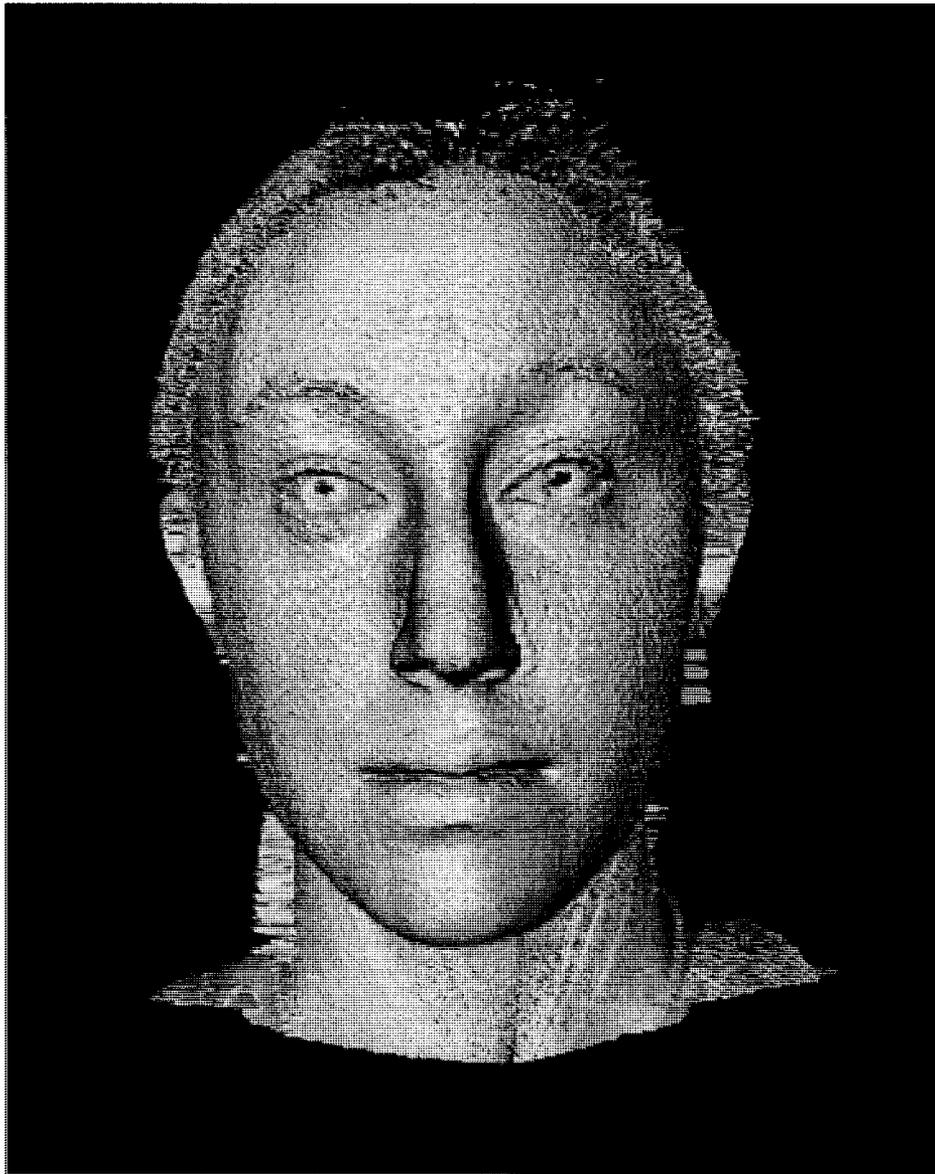


Figure 5.1 Raw scanned data I

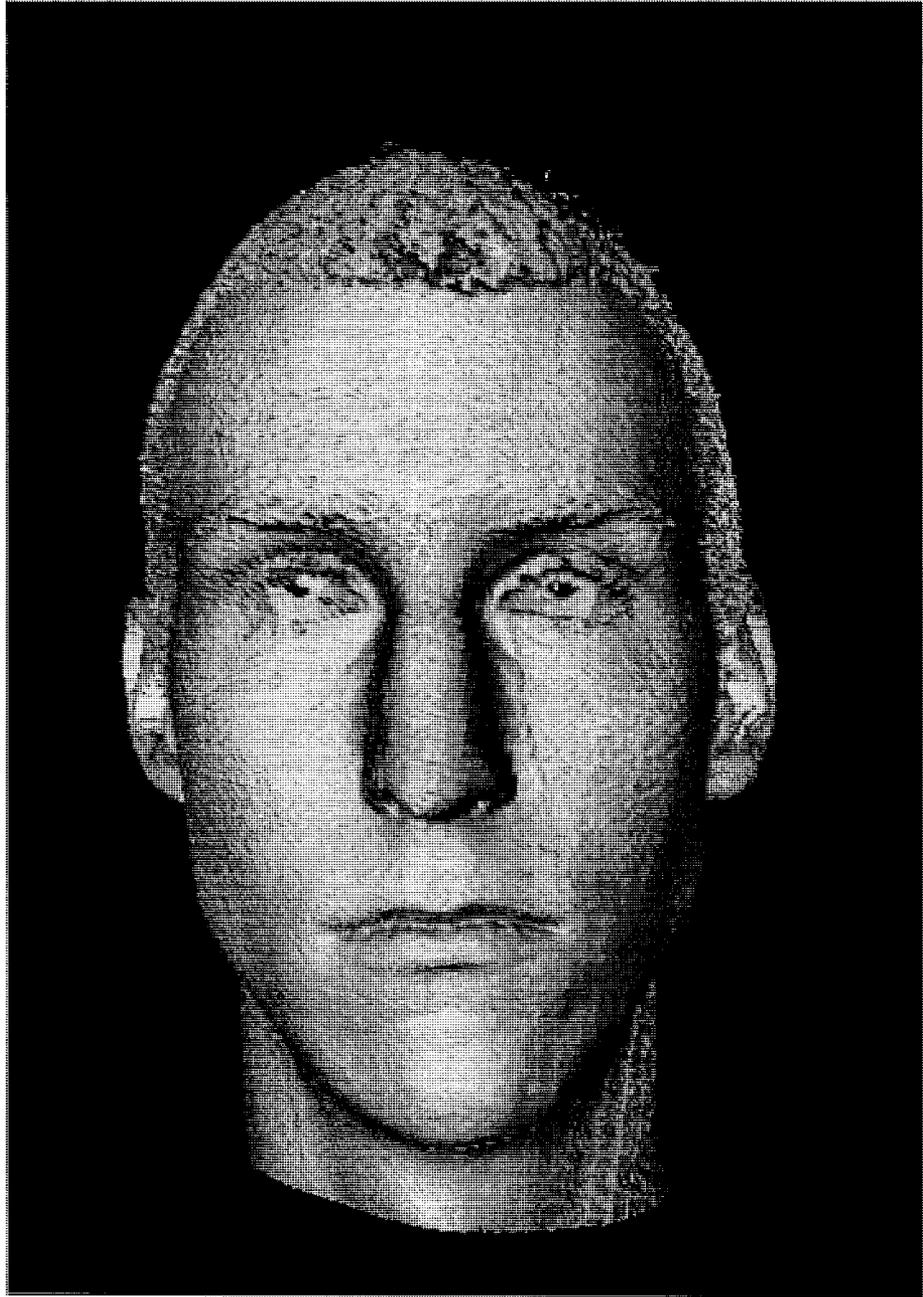


Figure 5.2 Raw scanned data II

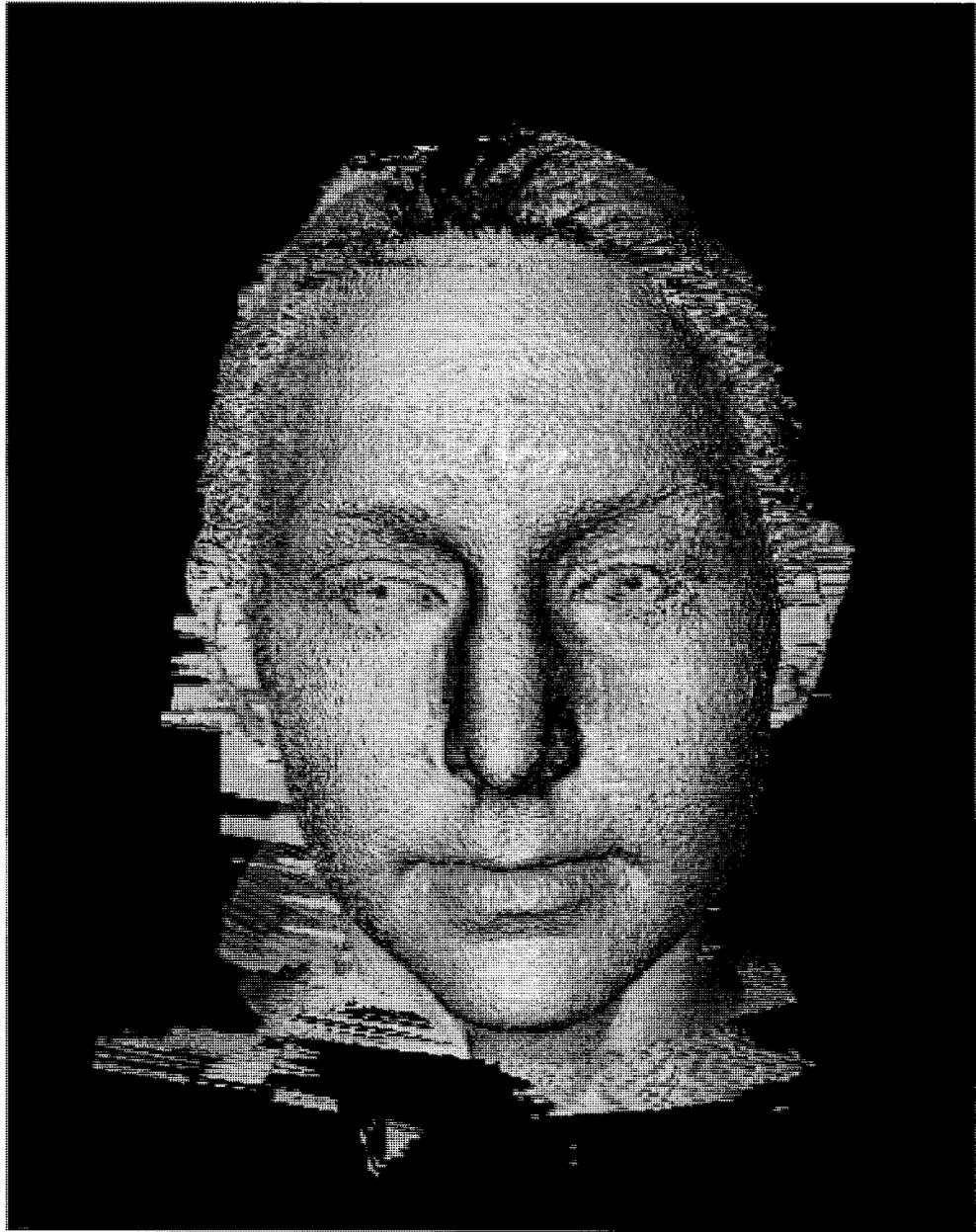


Figure 5.3 Raw scanned data III

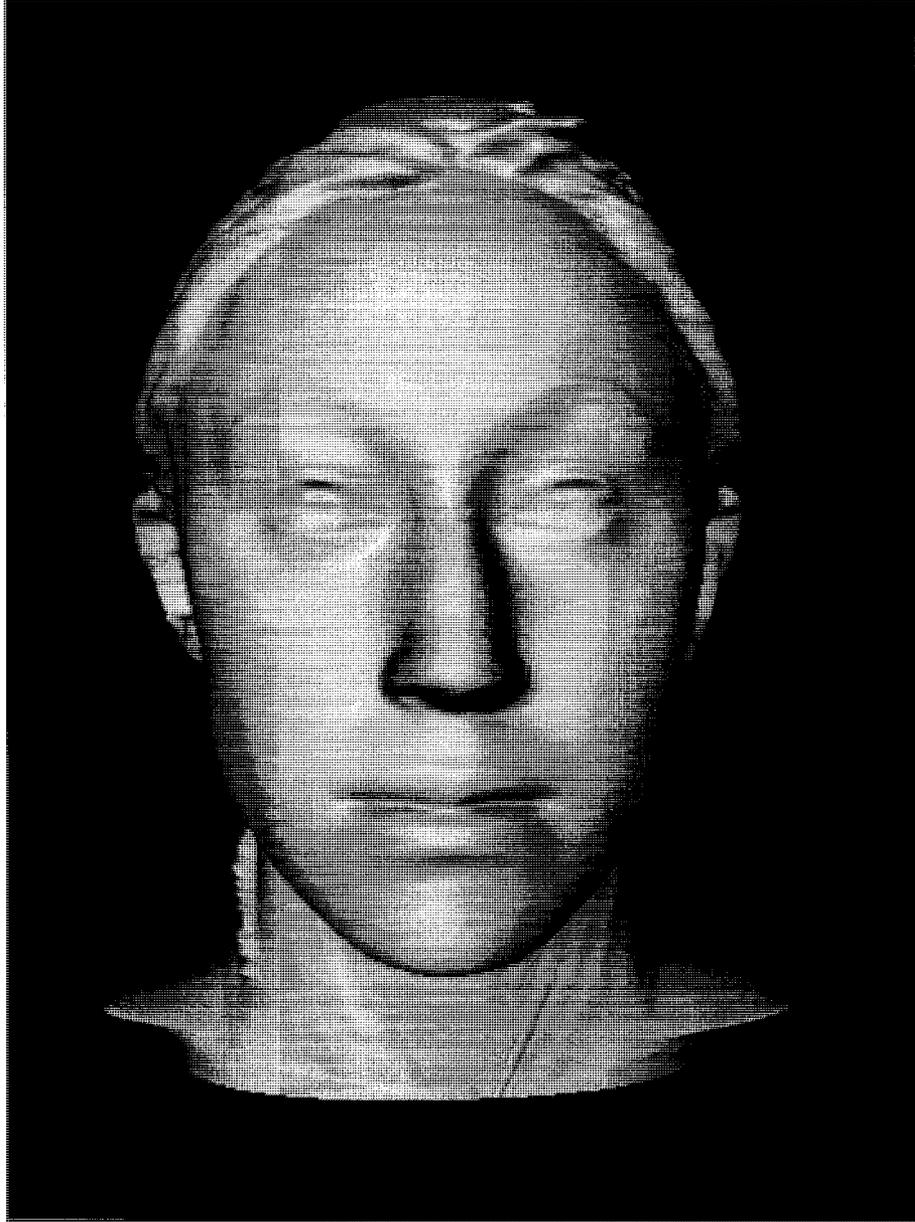


Figure 5.4 The processed scanned data I

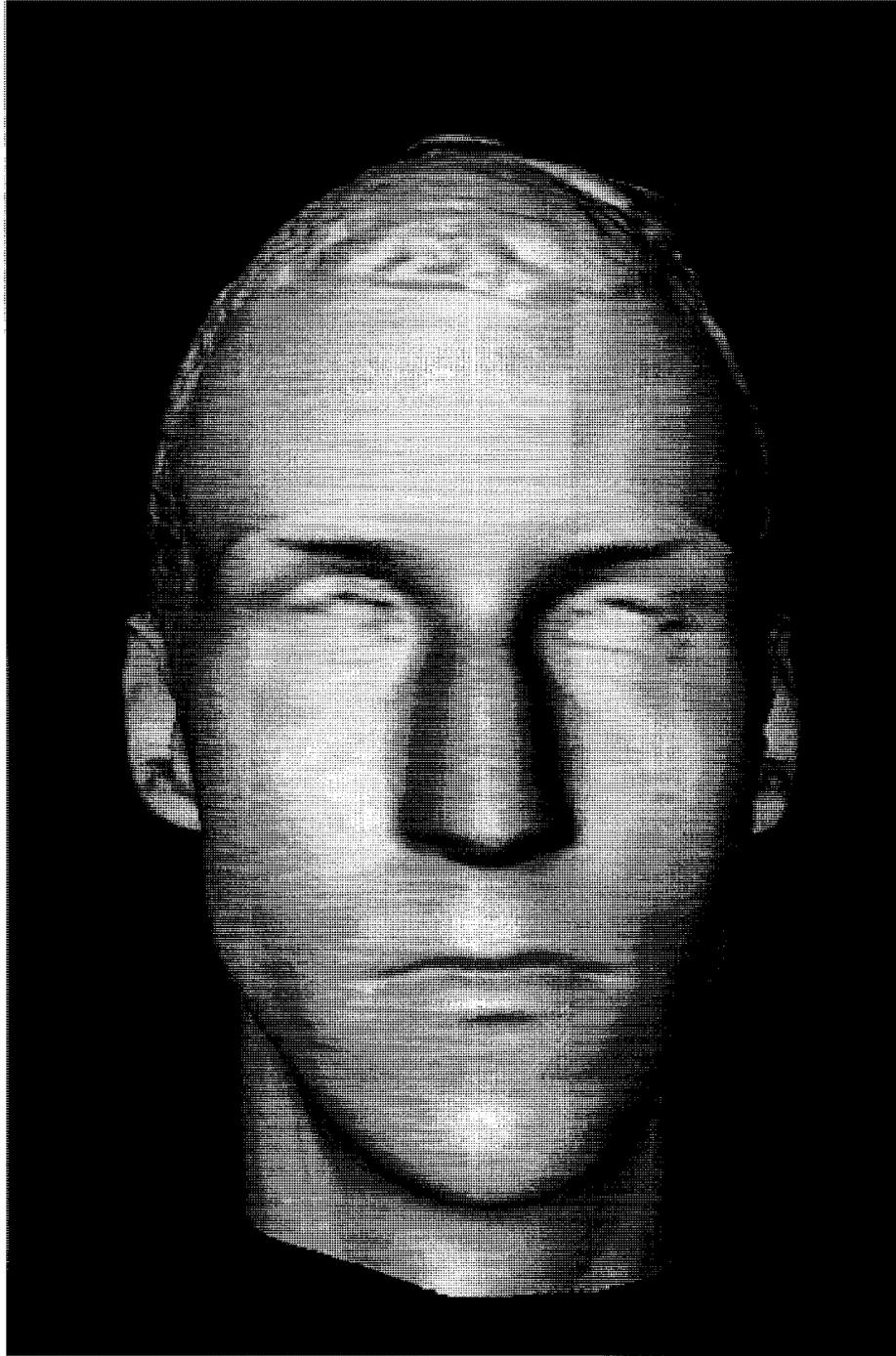


Figure 5.5 The processed scanned data II

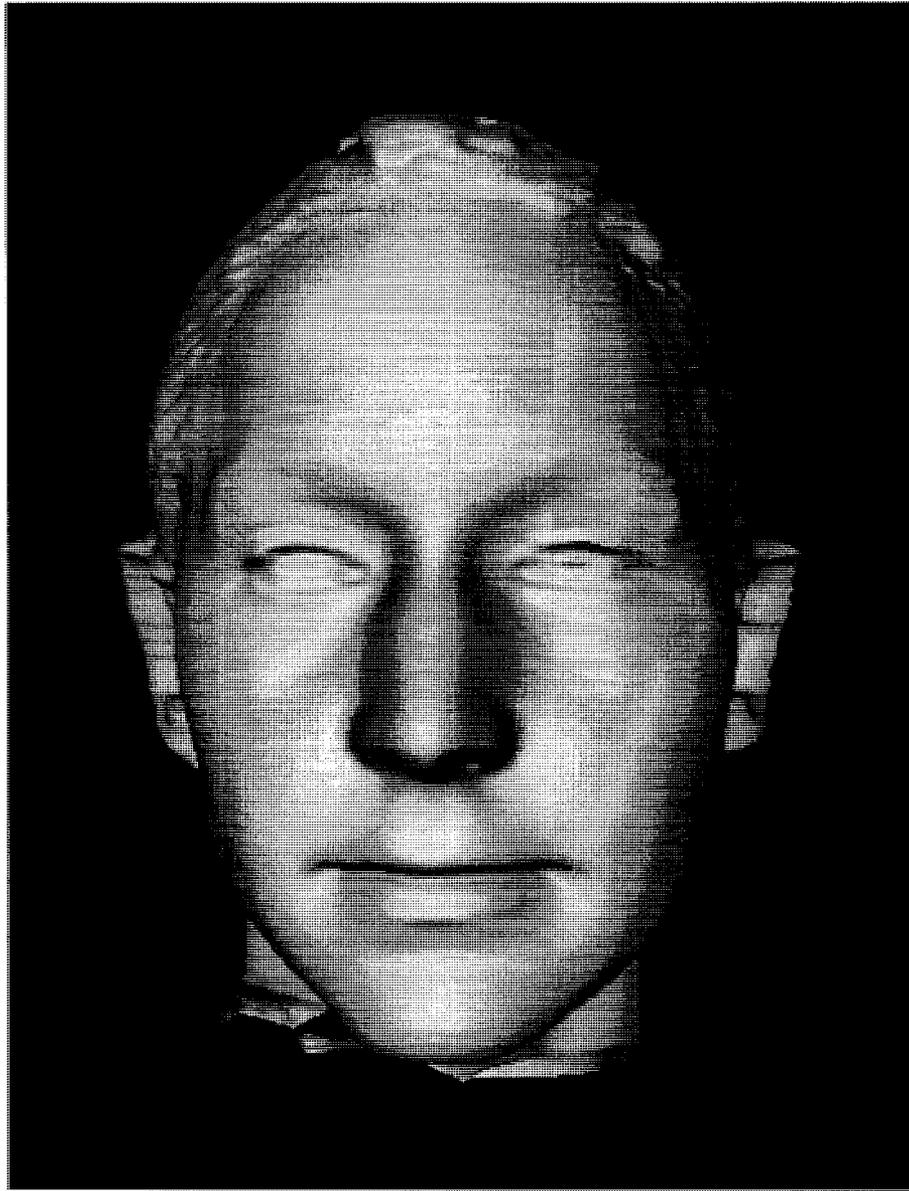


Figure 5.6 The processed scanned data III.

The major problem we encountered was the scanning of the hair especially with the female volunteers which introduces excessive noise. To overcome this, the hair was neatly packed and in some cases the process of scanning the face was repeated.

- **Rigging :** Figures 5.7 shows the reduced polygons with the attached bones to act as muscles. These are located around the mouth area to mimic the relaxation and contraction of the orbicularis oris muscle, around the cheek to mimic the relaxation and contraction of the zygomaticus major and minor muscles and around the eyebrow to mimic the relaxation and contractions of the orbicularis oculi muscles.

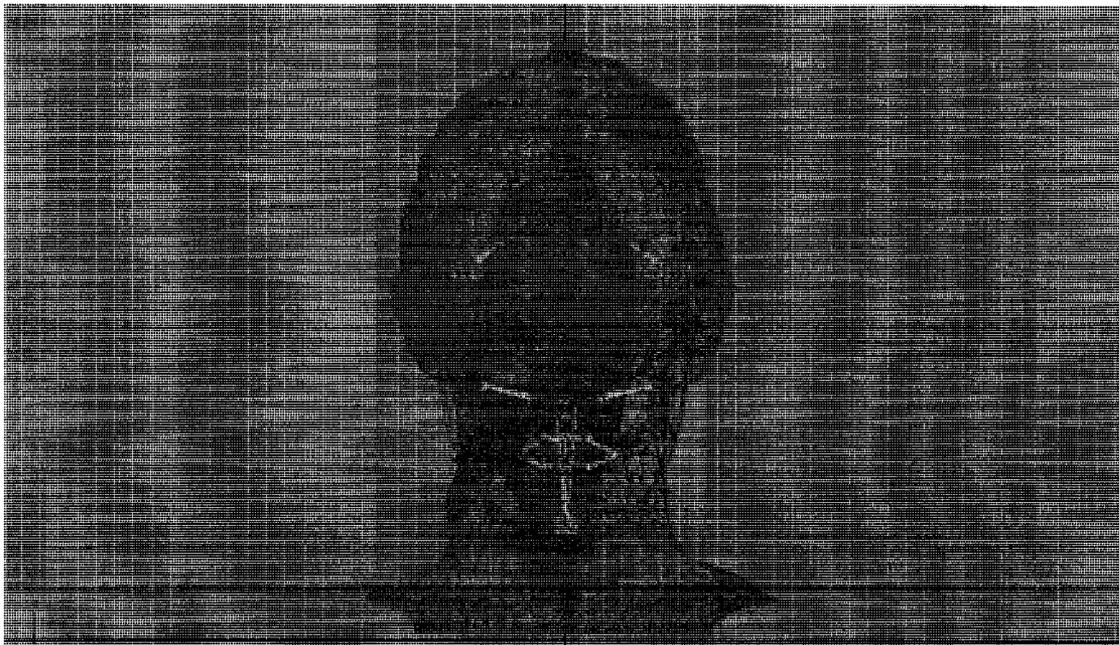


Figure 5.7 The rigged polygon mesh

- **Skinning:** Figure 5.8 shows the skinned polygon with envelopes that defines the area of influence of the eyebrow. The movements of the bone at the eyebrow locations also affect its area of influence.

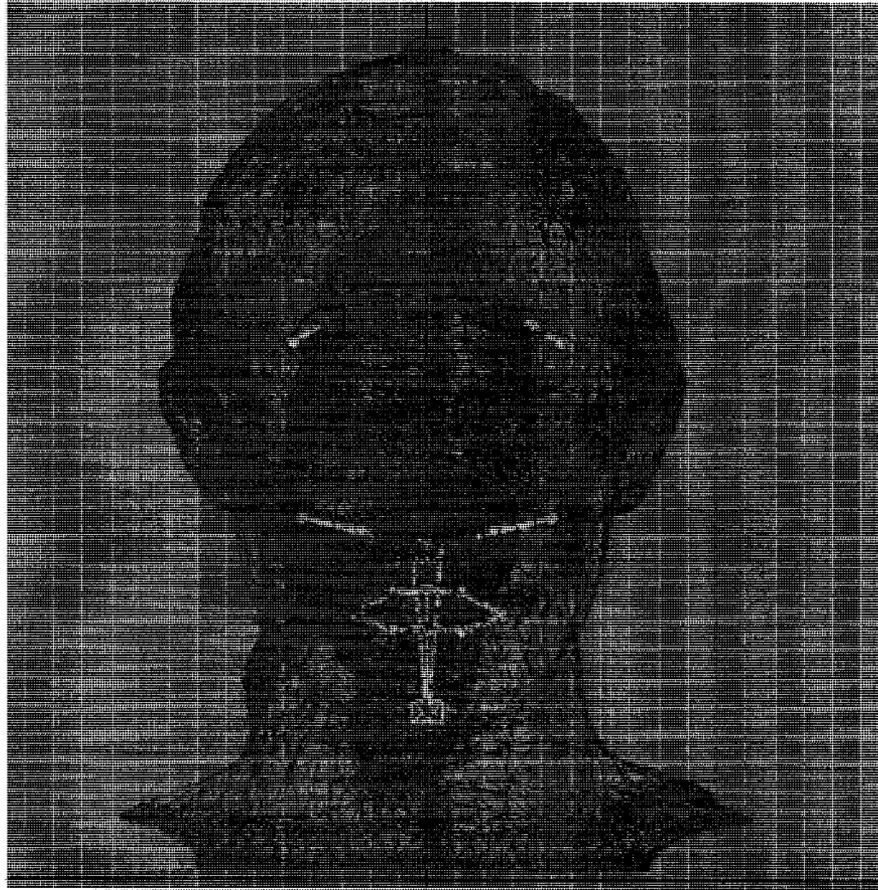
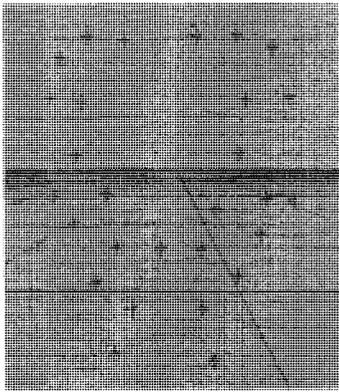
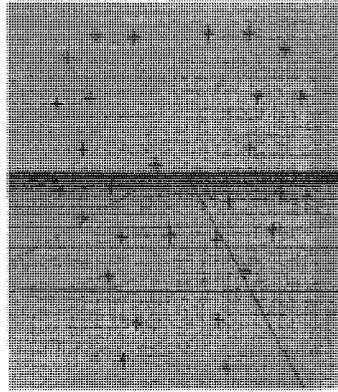


Figure 5.8 The skinned polygon with eyebrow envelope

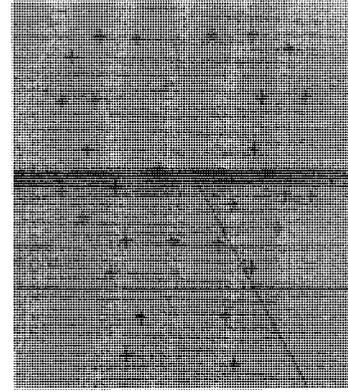
- **Animation:** Figures 5.9 and 5.12 show the exported motion capture data. The markers are shown as crosses on the screen for each animation frame.



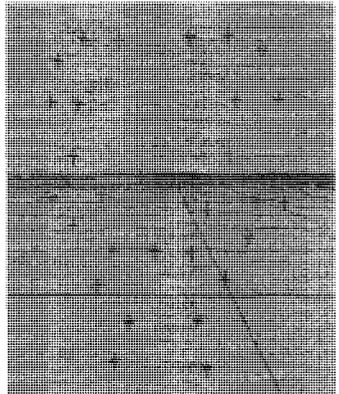
Frame 1



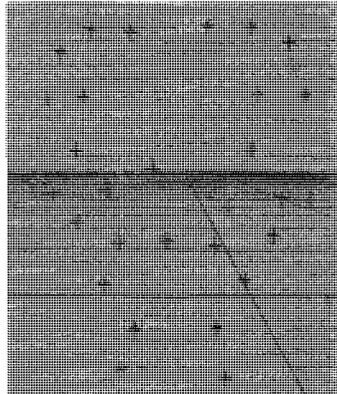
Frame 40



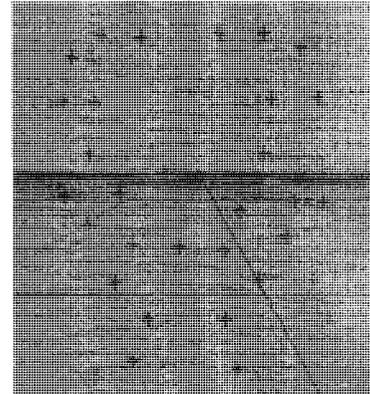
Frame 80



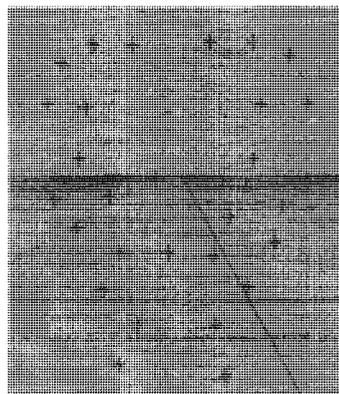
Frame 120



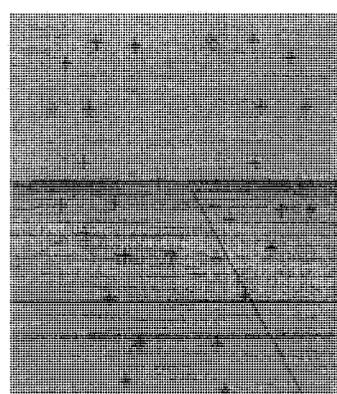
Frame 140



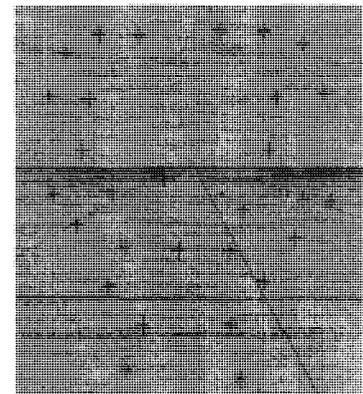
Frame 160



Frame 180

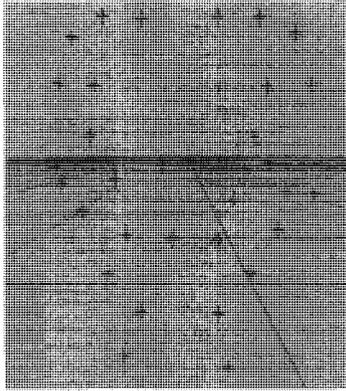


Frame 200

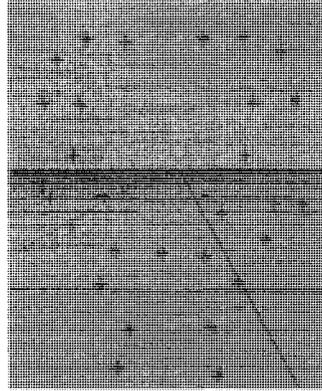


Frame 230

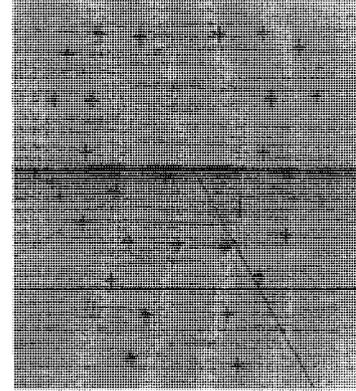
Figure 5.9 Selected frames for thirty facial marker set model I



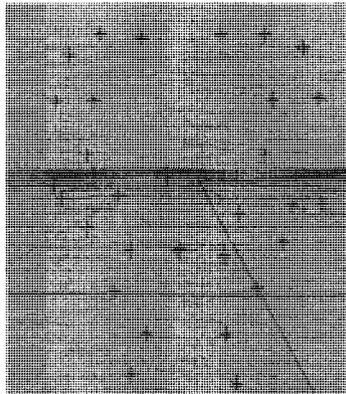
Frame 260



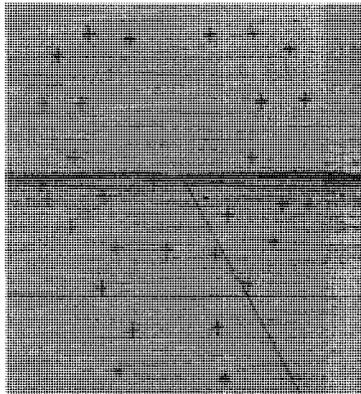
Frame 300



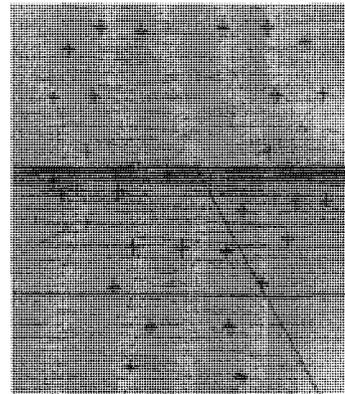
Frame 350



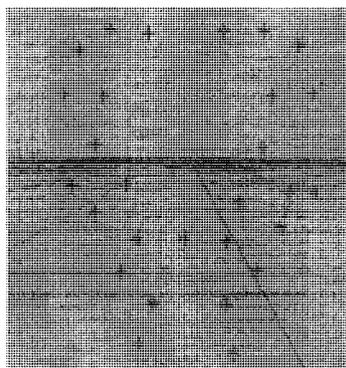
Frame 400



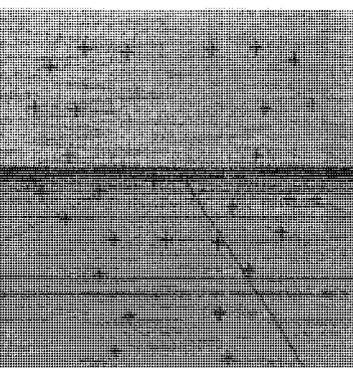
Frame 500



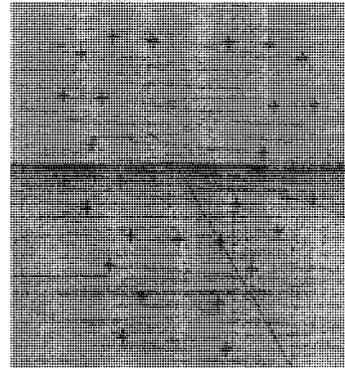
Frame 550



Frame 600

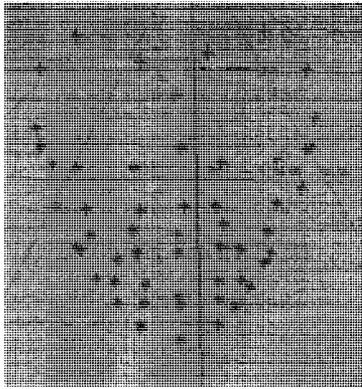


Frame 650

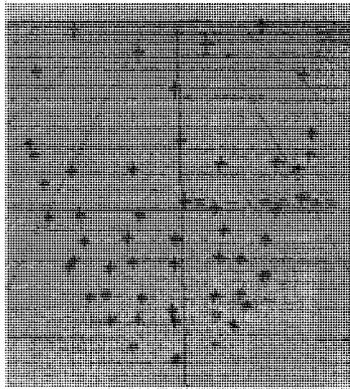


Frame 700

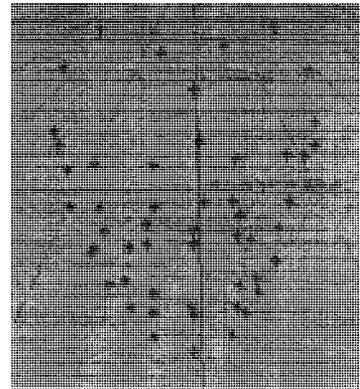
Figure 5.10 Selected frames for thirty facial marker set model II



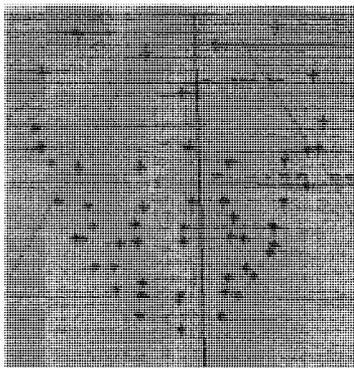
Frame 1



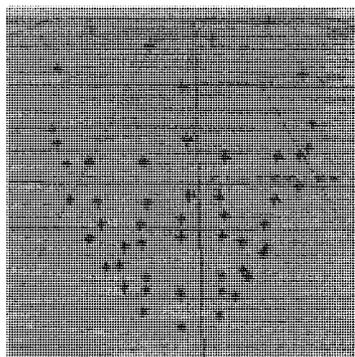
Frame 40



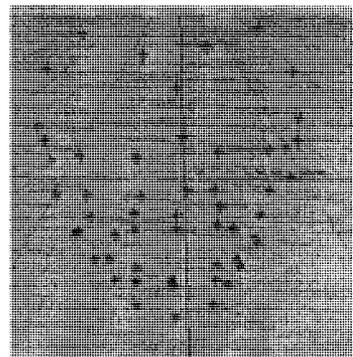
Frame 80



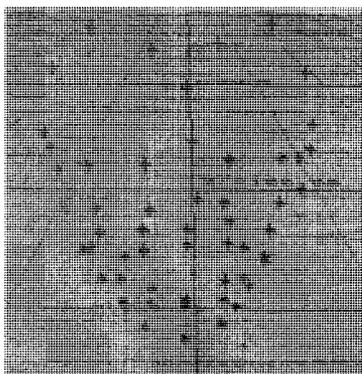
Frame 120



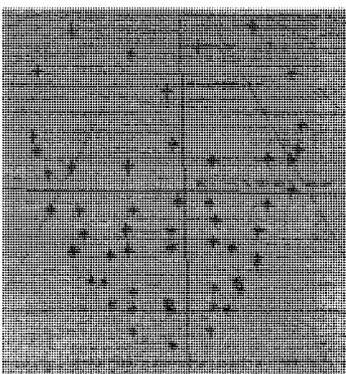
Frame 140



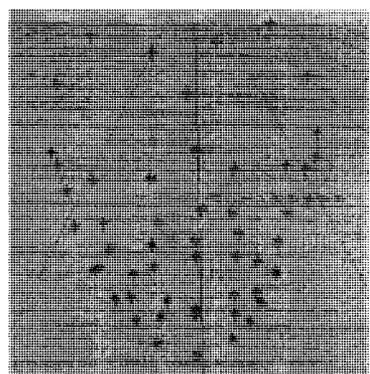
Frame 160



Frame 180

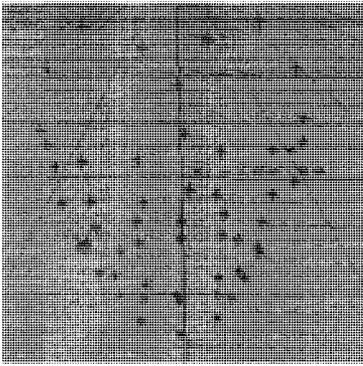


Frame 200

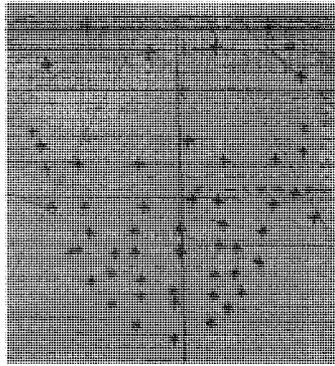


Frame 230

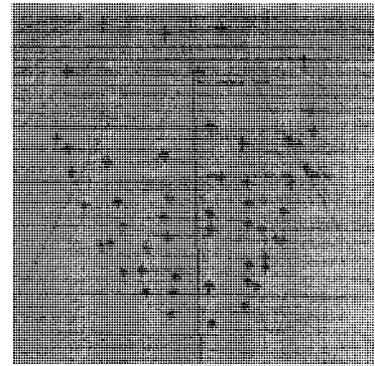
Figure 5.11 Selected frames for fifty-four marker set model I



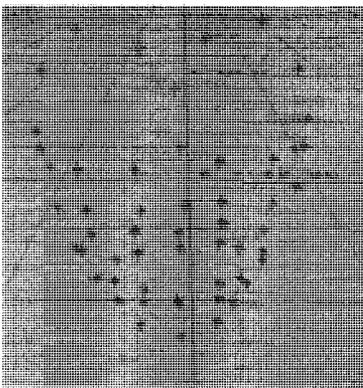
Frame 260



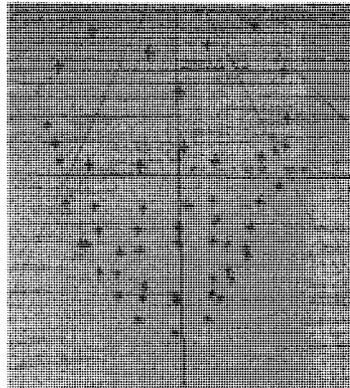
Frame 300



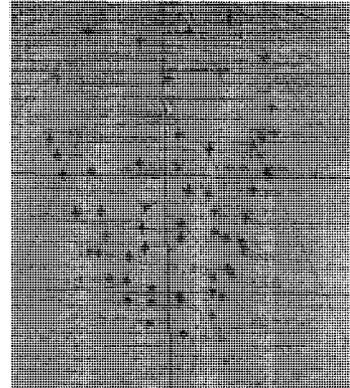
Frame 350



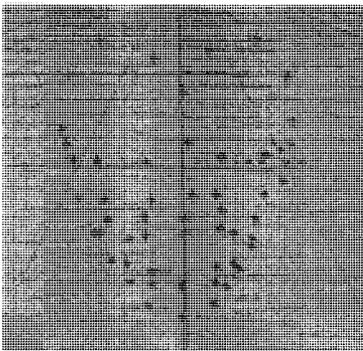
Frame 400



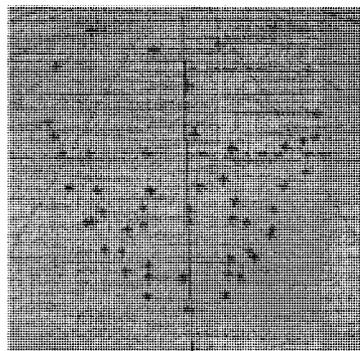
Frame 500



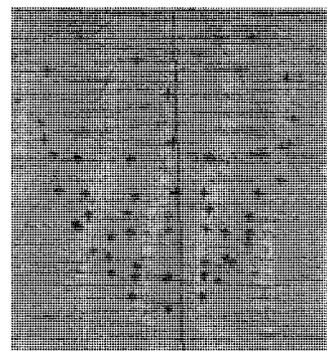
Frame 550



Frame 600



Frame 650



Frame 700

Figure 5.12 Selected frames from fifty-four facial marker set model II

Figures 5.9 and 5.10 show selected frames for the thirty facial marker set model while Figures 5.11 and 5.12 show selected frames for the fifty-four facial marker set model. A close look at these frames reveals that the changes between these marker sets especially around the mouth area is more pronounced for the fifty-four facial marker set. This proves that the fifty-four marker set model was able to capture the facial muscles around the mouth area more accurately than the thirty facial marker set model.

- **Rendering:** Figures 5.13, 5.14 and 5.15 show the final facial animation movies of the three volunteers. The length of the movie clips are twenty-nine seconds, twenty-seven seconds and twenty-three for the movie pictures I, II and III respectively. When the movie clips were played, the lips moved in synchronism with the speech that was recorded during the motion capturing sessions. The movies were also compared with the recorded videos and were found to be similar. The combined original videos and the result movies for the three volunteers are shown in Figures 5.16, 5.17 and 5.18.

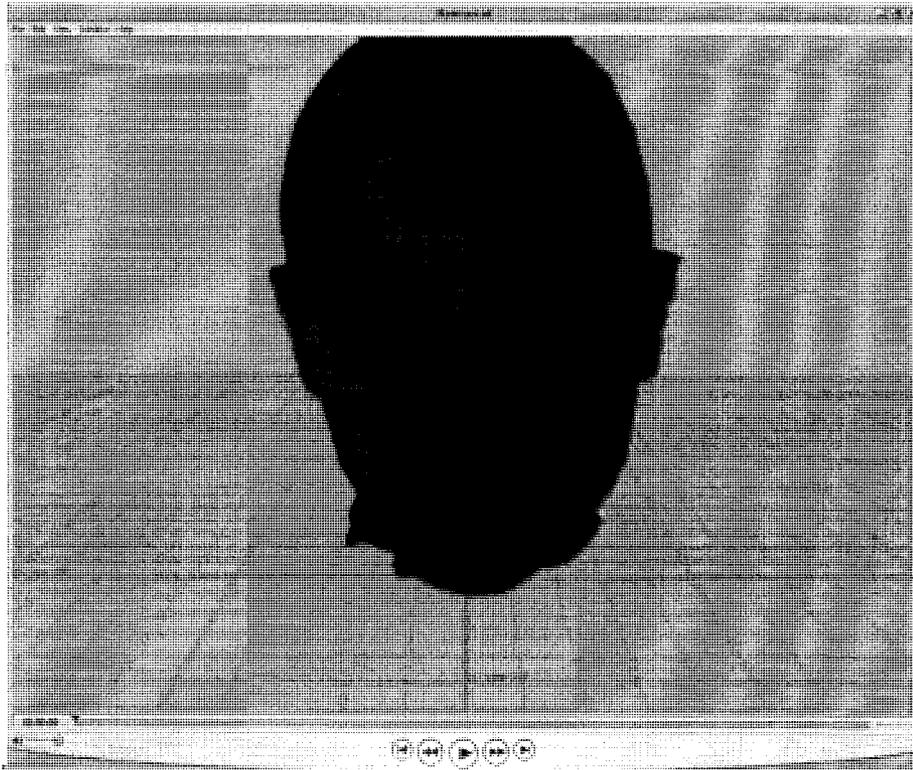


Figure 5.13 The Movie picture I.

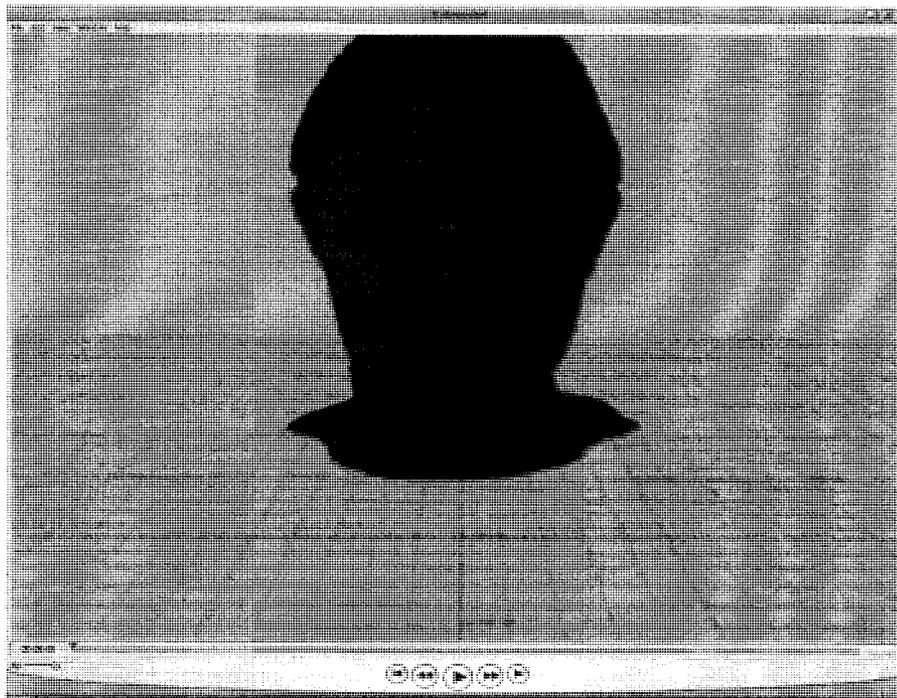


Figure 5.14 Movie picture II

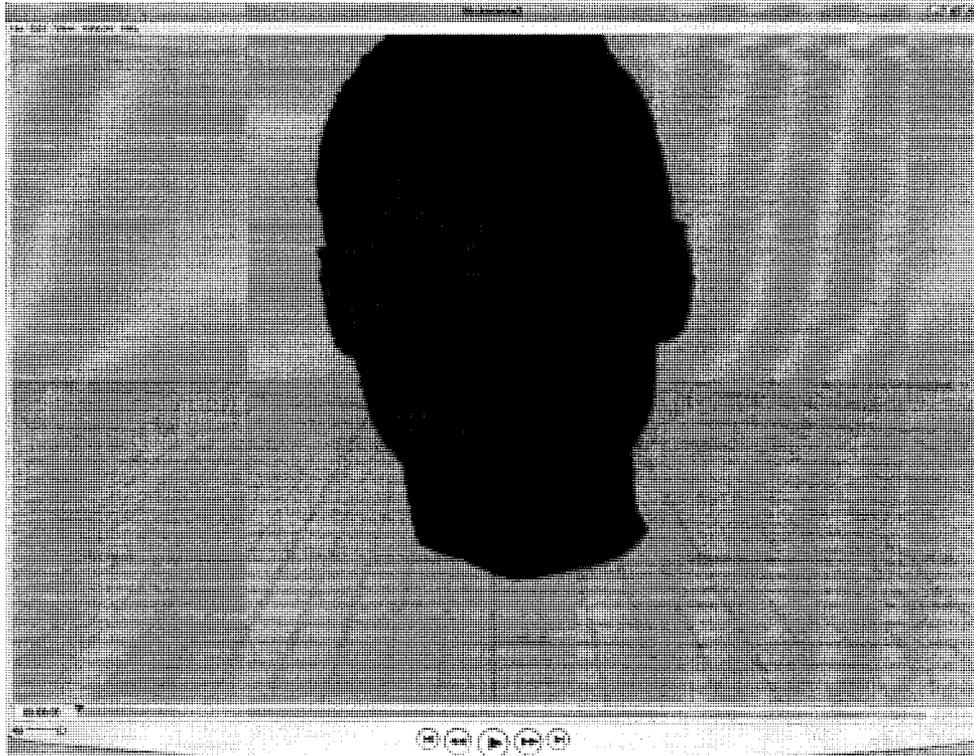


Figure 5.15 Movie picture III.



Figure 5.16 Combined video and result.I



Figure 5.17 Combined video and result.II

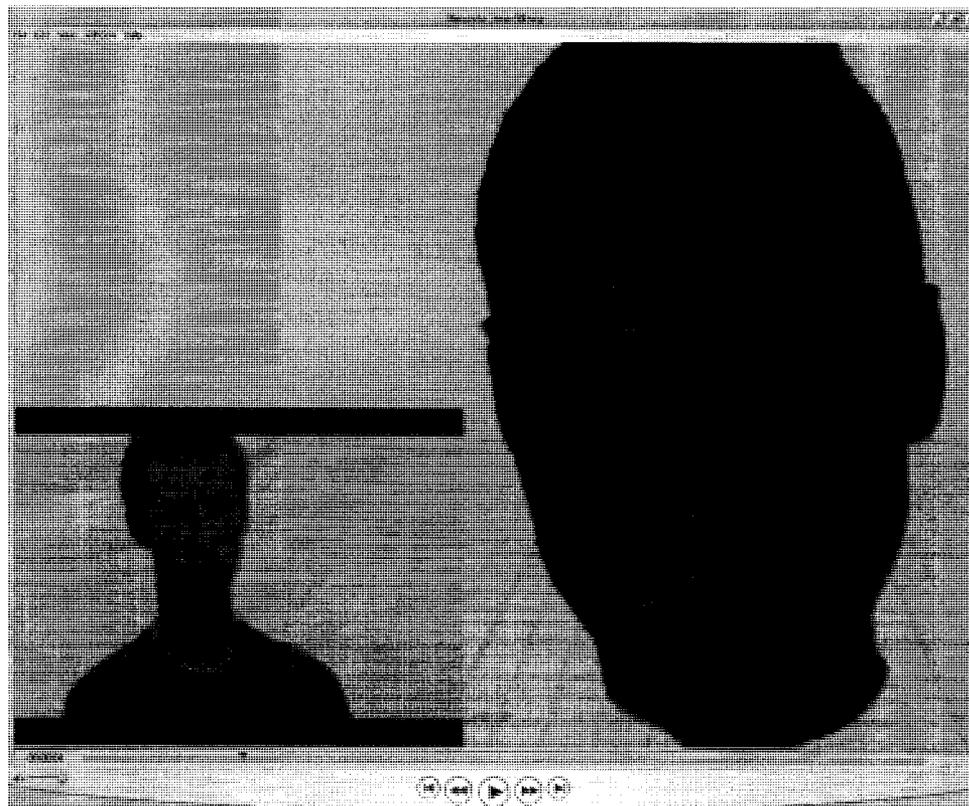


Figure 5.18 Combined video and result III

The final results show that the combination of the data from the head scanner and motion capture data along with audio data streams from video camera data was effective for facial animation. When the movie clips were played, the facial movements especially the lips movements were in synchronism with the spoken words which reflected the effectiveness of the motion capture data.

5.2 Evaluation

The result also proves that the fifty- four facial marker set which was modeled will definitely produce facial animations especially around the mouth. With the marker set the relaxation and contraction of the muscles around the mouth (that is *orbicularis oris*, *zygomaticus major*, *zygomaticus minor*, *risorius*, *levator anguli oris*, *triangularis*, *depressor labii inferioris* and *mentalis* (Figures 3.5 and 3.6)) was accurately captured by the Vicon motion capture camera.

The facial animation techniques (*i.e.* model-based and image based techniques discussed in the Chapter 2) attempt to mimic the dynamics of facial expressions and lip movements during articulations. However due to the complexity of the skin, that is its visco-elastic properties coupled with the complexity of the muscle's contraction and relaxation, the use of motion capture cameras produces better result. In this study, we show that the direct retrieval of both facial movements and facial physique with the aid of motion capture cameras and the head scanner can be used for animating the movements of facial muscles and the lips produced during speech.

CHAPTER SIX

CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

Achieving realistic facial animation is a great challenge especially as we all have a perspective of how ideal facial expression should be. The other issue that contributes to the problem is the complexity of facial shapes, muscles, articulation which varies among gender, race and age. The major work done in this present study is

- The modeling of one hundred and twenty-eight facial marker set which reflects the location of the facial muscles that lie beneath the skin especially around the mouth area. As a result of trajectory crossover, the effectiveness of the one hundred and twenty-eight facial marker set could not be verified; we then reduced the number of the markers to fifty-four.
- The modeling of fifty-four facial marker set. The effectiveness of the fifty-four marker set was verified by capturing the facial expressions and lip movements during speech for three volunteers, while they read aloud a text that includes all the forty-four major English phonemes. The motion capture data was combined with scanned data to produce facial animation tool.

- The development of a fast facial animation tool by combining the motion capture data with facial scanned data to produce facial animation tool thus making the mathematical model unnecessary.

The final movie clips proves that the fifty- four facial marker set is effective. The locations of the facial marker set reflected the location of the facial muscles especially around the mouth area. From the facial animation of the final movie we can conclude that the relaxation and contraction of these facial muscles was captured faithfully especially around the mouth area.

During the course of the project, it was discovered that the processing of the one hundred and twenty-eight facial marker set was difficult due to the inadequate number of the motion capture cameras. If the number of the cameras is increased to ten, it will then be possible to verify the effectiveness of the one hundred and twenty-eight facial marker set. In this case all of the facial muscles would be captured faithfully especially the upper part of the face that is not represented well in this project.

6.2 Future Work

The long time goal of the current study is the development an effective speech therapy tool. As mentioned in the introduction, this comprises of facial animation, speech recognition/synthesis/analysis and software tools for the development of speech therapy programs. The current study addressed the issue of rapid facial animation creation process by combining the motion capture data with that of scanned

data. In the process one hundred and twenty-eight facial marker set and fifty-four facial marker set were modeled. The fifty-four facial marker set was found to be effective. However, there are still more work to be done in order to accomplish the long term. Future work to be done

- With additional motion capture cameras, the effectiveness of the one hundred and twenty-eight facial marker set should also be tested in the same manner.
- A database of sub-clips of each diphone can be created and then used for generating other sentences and paragraphs. In this case the problem of co-articulations can be compensated by utilising a suitable co-articulation model. The sub-clips will probably be about three to eight frames of combined audio and video, reducing processing time. There is also the choice of a male voice or female voice.
- With the construction of the database of English diphone, a talking head application in speech therapy can be developed.
- Models of tongue, teeth and eyes to be included in the system.
- Software application tools for the development of speech therapy programs.

REFERENCES:

- [1] Moore, D.; McGrath, P.; Thorpe, J.; “Computer-Aided Learning for People with Autism – A Framework for Research and Development”, *Innovations in Education and Training International*, vol. 37 (3), 2000, pages 218-228.
- [2] Sutton, S.; Cole, R.; De Villiers, J.; Schalkwyk, J.; Vermeulen, P.; Macon, M.; Yan, Y.; Kaiser, E.; Rundle, B.; Shobaki, K.; Hosom, P.; Kain, A.; Wouters, J.; Massaro, D.; Cohen, M.; “Universal Tools: the CSLU Toolkit,” In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, November 1998, pages 3221-3224.
- [3] Turk, Oytun; Arslan, Levent M.; “Software Tools for Speech Therapy and Voice Quality Monitoring”, 13th *European Signal Processing, Eusipco 2005*, Antalga Turkey, 2005
- [4] Noh, Jun-yong, Neumann, Ulrich; “A survey of Facial Modeling and Animation Techniques”, *USC Technical Report 99-705* 1995.
- [5] Parke, F. I.; “Computer Generated Animation of Faces”, *Proceedings ACM annual Conference (ACM/CSC –ER archives)* 1972, vol. 1, pages. 451-457.
- [6] Cosatto, E.; Ostermann, J.; Graf, H.P.; Schroeter, J.; “Lifelike Talking Faces for Interactive Services”, *Proceedings of the IEEE, Special Issue on Human-Computer Multimodal Interface*, vol. 91, no.9, September 2003, pages 1406-1429.
- [7] Ostermann, J.; Weissenfeld, A.; “Talking Faces- Technologies and Applications”, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR’04)*, vol. 3, Aug. 2004, pages 826-833.

- [8] Radovan, M.; Pretorius, L.; “Facial Animation in a Nutshell: Past, Present and Future”, Proceedings of South African Institute of Computer Scientists and Information Technologists (SAICSIT) 2006, pages 71-79.
- [9] Lee, Won-Sook, . Magnenat-Thalmann, N.; “Fast Head Modeling for Animation”, Journal of Image and Vision Computing (SCI), ISSN 0262-8856, vol.18, no.4, Mar 2000, pages 355-364.
- [10] Breidt, Martin; Wallraven, Christian; Cunningham, Douglas W.; Buelthoff, Heinrich H.; “Facial Animation Based on 3D Scans and Motion Capture”. SIGGRAPH '03 Sketches and Applications, 0-0 (Eds), Campbell , N-ACM Press, New York, NY, USA, July 2003, page 1.
- [11] Borshuknov, George, Piponi, Dan; Larsen, Oystein; Lewis, J.P.; Tempelaar-Lietz, Christiana; “Universal Capture – Image-based Facial Animation for “The Matrix Reloaded” “,SIGGRAPH '03 Sketches and Applications, 0-0 (Eds), Campbell , N-ACM Press, New York, NY, USA, July 2003, page1.
- [12] Zhu, Lijia.; Lee, Won-Sook; “Modeling and Animating for the Dense Laser-Scanned Face in the Low Resolution Level”, Proceedings of the 17th IASTED International Conference Modeling and Simulation, May 2006, pages 101-106.
- [13] Goto, T.; Lee Won-Sook, . Magnenat-Thalmann, N.; “Facial feature extraction for quick 3D face modeling” Signal Processing: Image Communication, ISSN 0923-5965, vol.17, issue 3, March 2002, pages 243-259.
- [14] Parke, F.I. “Techniques of Facial Animation” In. Magnenat-Thalmann N. and Thalmann, D editors- New Trends in Animation and Visualization, 1991,Chapter 16, pages 229-241, John Wiley and Sons. Referenced in 4.

- [15] Parke, F.I.; "Parameterized Models for facial animation" IEEE Computer Graphics and Application vol.2, no.9, Nov. 1982, pages 61-68.
- [16] Ekman, P.; Friesen, W., "Facial Action Coding System: A Technique for the Measurement of Facial Movement", Palo Alto, CA: Consulting Psychologists Press, 1978: Referenced in 4
- [17] Kaehlr, Kolja; " A Head Model with Anatomical Structure for Facial Modeling and Animation", 2003 PhD Thesis
- [18] Lee, Yuencheng; Terzopoulos, Demetri; Waters, Keith; "Constructing Physics-Based Facial Models of Individuals, Proceedings of the Graphics Interface 93 Conference, May 1993, pages 1-8.
- [19] Magnenat-Thalmann, N, Primeau, E., Thalmann, D; " Abstract muscle action procedures for face animation", The Visual Computer, vol. 3, 1988, pages 290-297.
- [20] Patt, S.M.; Badler, N.I.; "Animating facial Expression", Computer Graphics, vol.15, no.2, Aug. 1981, pages 245-252.
- [21] Waters, K.; "A Muscle model for Animation Three-Dimensional Facial Expression", In Computer Graphics, SIGGRAPH 87 Conference Proceedings, July 1987, pages 17-24.
- [22] Terzopoulos, D.; Waters, K.; "Physically-Based Facial Modeling, Analysis and Animation", The Journal of Visualization and Computer Animation, vol. 1, no.2, 1990, pages 73-80.
- [23] Terzopoulos, D; Waters, K.; "Techniques for Realistic Facial Modeling and Animation", IEEE Computer Animation, 1991, pages 59-74.

- [24] Koch, R.M.; Gross, M.H.; Bosshard, A.A.; “Emotion Editing using Finite Elements”, Computer Graphics forum- Proceedings of Eurographics 1998, vol.17, no.3, pages C295-C302.
- [25] Kaehlr Kolja; Haber Jorg; Yamauchi, Hitoshi; Seidel, Hans-Peter; “Head Shop: Generation animated head models with anatomical structure”, Proceeding 2002 ACM SIGGRAPH symposium on Computer animation, pages 53-63.
- [26] Kaehlr Kolja; Haber Jorg; Seidel, Hans-Peter; “Geometry-based Muscle Modeling for Facial Animation”, In Proceedings of the Graphics Interface, 2001, pages 37-46.
- [27] Sederberg, T.W.; Parry, S.R.; ”Free Form Deformation of Solid geometry models”, Computer Graphics, Siggraph 1996, vol. 20(4), pages 151-160.
- [28] Coquillart, S. “Extended Free Form Deformation: A Sculpturing Tool for 3D Geometric Modeling”, Computer Graphics, 1990, vol. 24, pages 187-193.
- [29] Kalra, P; Mangili, A.; Thalmann, N.M.; Thalmann, D; “Simulation of facial muscle actions based on rational free Form Deformation”, Proceedings of Eurographics 1992, pages 65-69.
- [30] Ezzat, Tony; Poggio, Tomaso; “ Facial Analysis and Synthesis using Image-based Models” , Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Oct. 1996, pages 116-121
- [31] Ezzat, T.; Geiger, G.; Poggio, T.; “Trainable Videorealistic Speech Animation”, Proceedings of 29th annul Conference of Computer Graphics and Interactive Techniques, 2002, pages 388-398.

- [32] Terzopoulos, D.; Waters, K.; “Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no.6, June 1993, pages 569-579.
- [33] Casatto, E.; Graf, H.P.; “Sample-Based Synthesis of Photo-Realistic Talking Heads”, Proceedings of Computer Animation, Philadelphia, Pennsylvania 1998, pages 103- 110.
- [34] Cosatto, E. ; Graf, H.P.; “Photo-Realistic Talking-Heads from Image Samples”, IEEE Transactions on Multimedia, vol.2,no.3, September 2000, pages 152-163.
- [35] Chang, Yao-Jen; Ezzat, Tony; “Transferrable Videorealistic Speech Animation”, ACM SIGGRAPH Symposium on Computer Animation, 2005, pages 143-151.
- [36] Ezzat, Tony; Poggio, Tomaso; “MikeTalk: A Talking Facial Display Based on Morphing Visemes”, Proceedings of IEEE Computer Animation, 1998, pages 96-102.
- [37] Bregler, C.; Covelt, M.; Slaney, M.; “Video Rewrite: Driving visual speech with Audio: Proceedings of ACM SIGGRAPH 1997, pages 353-360.
- [38] Ali, Arya; Babak, H.; “FIX: Feature-based Image Transformations for Face Animation”, IEEE International Conference on Information Technology: Research and Education, ITRE-2003, Aug.2003, pages 500-504.
- [39] Williams, Lance; “Performance- driven Facial Animation”, International Conference on Computer Graphics and Interactive Techniques, vol.24, no.4, August 1990, pages 235-242.

- [40] Kass, M.; Witkin, A.; Terzopoulos, D.; “Snakes: Active Contour Models”, International Journal of Computer Vision, vol.1, no.4, 1987, pages 321-331.
- [41] Optical Flow, Wikipedia Free Encyclopedia, http://en.wikipedia.org/wiki/Optic_flow
- [42] Choe, Byoungwon; Lee, Hanook; Ko, Hyeong-Seok; “Performance-Driven Muscle-Based Facial Animation”, The Journal of Visualization and Computer Animation, vol.12, no.4, May 2001, pages 67-79.
- [43] Essa, I. A.; Darrell, T.; Pentland, A.; “Tracking Facial Motion”, Proceedings of IEEE Workshop on Non-rigid and Articulate Motion, Nov.1994, pages 36-42.
- [44] Haber,Jorg; “ Anatomy of the Human Head” In SIGGRAPH 2004 course notes Facial Modeling and Animation..
- [45] Lay D. Le; “Teach yourself human anatomy and physiology”, Hodder and Stoughton Educational, fourth edition, ISBN 0-340-78960-3, 2001
- [46] Patton, K. T.; Thibodeau, G.A.; “Mosby’s Handbook of Anatomy and Physiology”, Mosby Inc. ISBN 0-323-01096-2, 2000.
- [47] Lucero, J.C.; Munhall, K.G.; “A Model of facial biomechanics for speech production”, The Journal of the Acoustical Society of America, vol.106, no.5, Nov. 1999, pages 2834-2842.
- [48] Cross section of the Skin, http://www.infovisual.info/03/036_en.html
- [49] Waters, Keith; “A Physical Model of Facial Tissue and Muscle Articulation Derived form Computer Tomography Data”, Proceedings of SPIE - Visualization in Biomedical Computing, volume 1808, Sept. 1992, pages 574-583.

- [50] Larrabee, W.F.; “A Finite Element Model of Skin Deformation. I. Biomechanics of Skin and Soft Tissue”, The Laryngoscope, vol.96, no.4, 1986, pages 399-405.
- [51] Branon, H.; “Skin Anatomy”, <http://dermatology.about.com/cs/skinanatomy/a/anatomy.htm>
- [52] Muscle structure http://www.spring8.or.jp/en/current_result/press_release/2002/020806_fig/fig-1-e_gif
- [53] Sobotta, J.; “Atlas of Human Anatomy, Vol.1; Head, Neck, Upper Limb, Lippincot Williams and Wilkins, 13th English edition 2001, referenced in 17
- [54] A Human Face- Muscle Details in the upper Face, <http://face-and-emotion.com/dataface/anatomy/upperfacemusclesdetails.jsp>
- [55] A Human Face- Lower Facial Muscles-Details, <http://face-and-emotion.com/dataface/anatomy/lowerfacemusclesdetails.jsp>
- [56] Muscle Action; <http://www.face-and-emotion.com/dataface/expression/muscles.jsp>
- [57] The sounds of English and International Phonetic Alphabet. <http://www.antimoon.com/how/pronunc-soundsipa.htm>
- [58] VICON Motion Systems, “Advanced Capture Techniques.
- [59] MotionBuilder 7.5 Users’ Guide, version 7.5, March 2006, Autodesk Inc.
- [60] Basu, S.; Oliver, N.; Pentland, A.; “3D Modeling and Tracking of Human Lips” , Proceedings of the IEEE International Conference on Computer Vision, Mumbai India, Jan 1998, pages 337-343.

- [61] Zhang, Yu; Prakash, Edmond C.; Sung; Eric; “ A New Physical Model with Multilayer Architecture for Facial Expression Animation using Dynamic Adaptive Mesh, IEEE Transactions on Visualization and Computer Graphics, vol.10, no. 3. May/June 2004, pages 339-352
- [62] Waters, K; Terzopoulos, D; “Modeling and Animating faces using Scanned data”, Journal of Visualization and Computer Animation, vol.2, no.4, Oct. 1991, pages 123-128.

APPENDIX A

ANATOMICAL BASED MODEL

Physics-based models are based on the anatomical structure of the head which comprises of the different layers of muscles, cartilage, bones, nerves, blood vessels, glands, fatty tissue, connective tissue, hair and skin [25][26]. For computer analysis most studies concentrated on the skin, muscles and the skull [18]. The human face comprises of wide range of muscle types which are rectangular, triangular, sheet, linear and sphincter [17][21,], however most studies are based on linear, sheet, and sphincter [21][18]. As reported in the medical literature, muscles lie beneath the skin, often originate from the bone and inserts into the skin. When the facial muscle contracts it becomes shorter (linear/sheet muscles) or thicker (sphincter muscles) and at the same time interacts with other muscles within its influence thus resulting in the movement of skin/tissue layer to produce facial expression like fear, anger, sadness, surprise and pain. During muscle contraction, the movement at the point of origin is static while the insertion point experiences all the movements. Similar process is accomplished during speech production.[17]

Physics based models is typically accomplished by mass spring [49][23] or finite element method [60]24][50] to model the visco-elastic properties of the skin.

A.1 Head Structure

Most of the study [25][18][61] considered three conceptual layers based on the anatomy of the head, namely

- Skin/tissue layer to represent epidermis, subcutaneous fatty tissue.
- Attached to the skull and inserted into the skin is the muscles layer.
- Finally the bone structure which comprises of a rigid skull with a rotating jaw.

A.2 Skin/tissue:

As illustrated in Figure A-1 the skin/tissue is the physics based model's top layer which connects to muscles and bones. The elastic properties of the dermis, epidermis and the fatty subcutaneous layer are also taken into consideration for the skin model. The model also includes local volume preservation and skull penetration constraints [22][26][18]

A.3 Muscles:

Muscles are represented as piecewise linear fibres which contract isotonicly by the shortening of the linear segments with linear muscles contracting towards the

insertion point while sphincter muscles contract towards a point. These linear muscles ‘fibres’ can be combined into groups to form sheet muscles. Muscles are laid out in different configurations to reflect their anatomical structure; as shown in Figure A-2; long and thin strands to represent *zygomatic major* , broad sheets to represent *frontalis*, curved muscles to represent *levator labii superioris alaque nasi* and sphincter to represent *orbicularis oris*. The model also incorporated the merging and intertwining of muscles thus producing coupled actions. [25][26]

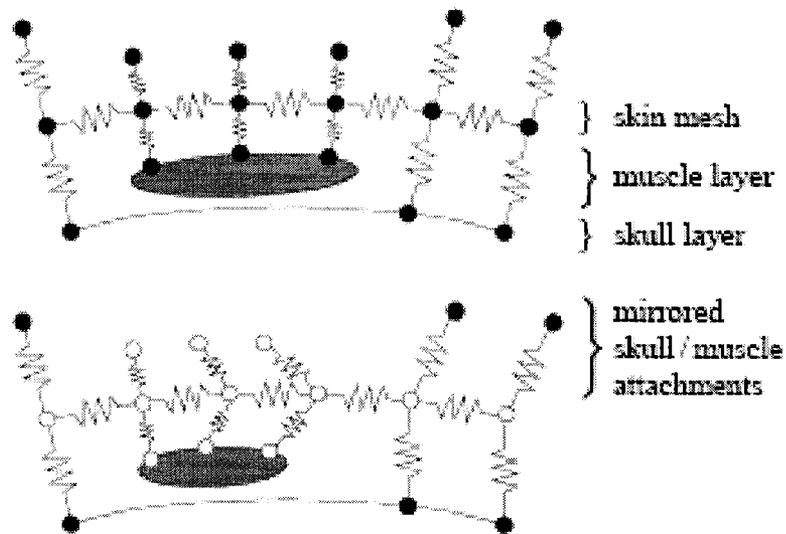


Figure A-1 Mass-spring system [26]

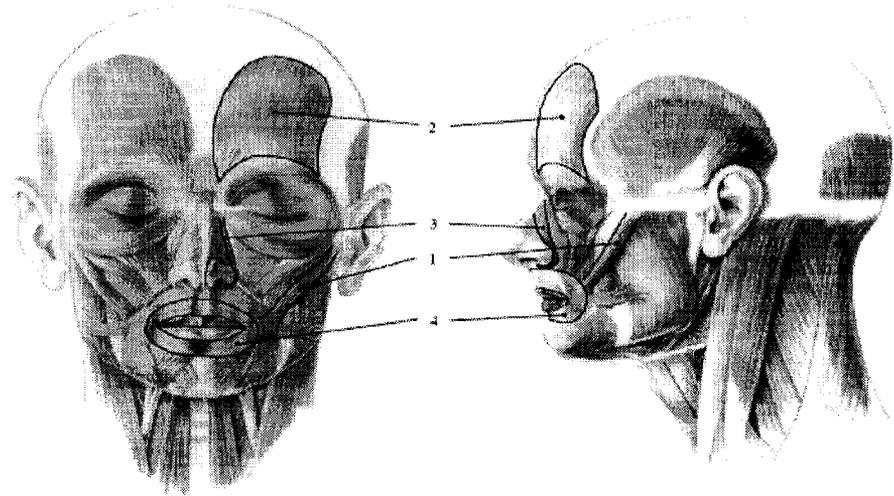


Figure A-2 Variety of supported muscles [26]

(1). linear muscles, (2). sheet muscles (3) curved muscles (4) sphincter muscles

A.4 Skull and Jaw:

The anatomical model made a distinction between the skull and movable jaw and their meshes are used to determine if part of skin or muscle layers lie over the skull or the jaw. The mass spring mesh handles the skull penetration constraints internally.

APPENDIX B

Mathematical Models

B.1 Skin /Tissue Model

As mentioned previously, skin is made of epidermis, dermis and the subcutaneous cellular tissue. The dermal layer (dermis) exhibits a non linear stress-strain relationship which is reflected in the Waters and Terzopoulos's model [62]. Figure B.1 shows the various layers; the skin (comprising of epidermal surface and dermal-fatty layer), muscle layer and the skull surface as represented by the nodes.

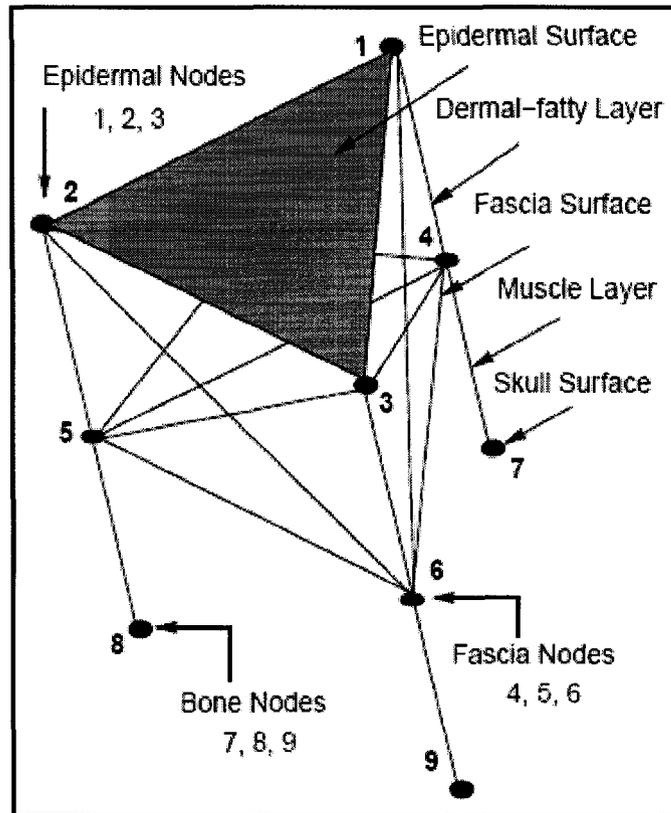


Figure B-1 Triangular skin tissue prism element [18]

Assume a node i of a mass m_i with a 3D position given as

$$x_i(t) = [x_i(t), y_i(t), z_i(t)]^T$$

The node velocity and its acceleration is given as $v_i = \frac{dx_i}{dt}$ and $a_i = \frac{d^2x_i}{dt^2}$

respectively. [5, 20] A spring k of length l_k^R and stiffness c_k which connects nodes i and j together. The vector separation of the nodes is represented by r_k and is given as

$r_k = x_j - x_i$ the actual spring length l_k is $\|r_k\|$ and spring deformation e_k is expressed

as $e_k = l_k - l_k^R$ then the force s_k that the spring exerts on node i is defined as

$s_k = \frac{c_k e_k}{l_k} r_k$. The overall force on node i resulting from the other springs which

connect to node $j \in N_i$ in the deformable lattice is defined as $g_i(t) = \sum_{j \in N_i} s_k$ [18][

22]

The modified version incorporated the biphasic dermal fatty springs by introducing two spring stiffness for the dermal layer.

$$c_k = \begin{cases} I_k K_2 / l_k^R & \text{for } l_k > l_k^R (1 + B_p^k) \\ I_k K_1 / l_k^R & \text{for } l_k^R (1 + B_p^k) > l_k > l_k^R \\ I_k K_1 \Phi(l_k, l_k^R) / l_k^R & \text{for } l_k^R > l_k \end{cases}$$

where K_1 and K_2 are the two spring stiffness for the biphasic dermal fatty springs

B_p^k is the spring's biphasic point and it is in terms of I_k / I_k^R

While the spring stiffness for other springs is

$$c_k = \begin{cases} I_k K / l_k^R & l_k > l_k^R \\ I_k K \Phi(l_k, l_k^R) / l_k^R & l_k^R > l_k \end{cases} \quad \text{K is the spring stiffness for other springs.}$$

K for epidermis is 60, dermal fatty 30 and 70, fascia is 80 and the muscle is 10.[21]

The equation of motion for the tissue model is given by the formula below

$$m_i \frac{d^2 x_i}{dt^2} + \gamma_i \frac{dx_i}{dt} + \tilde{g}_i + \tilde{q}_i + \tilde{s}_i + \tilde{h}_i = \tilde{f}_i \quad \text{where}$$

- m_i = nodal mass.
- γ_i = damping coefficient..
- \tilde{g}_i = total spring at node i
- \tilde{q}_i = total volume preservation force at node i
- \tilde{s}_i = total skull penetration force at node i
- \tilde{h}_i = total nodal restoration force at node i
- \tilde{f}_i = total muscle force applied at node i

[21]

B.2 Muscle Model.

Kahler et al's muscle model [26] incorporated the anatomical structure of muscles. Muscle was built as piecewise linear segment comprising of individual

fibers which represented muscle fibers and sacromere. Figure B.2 shows the relaxed muscle polygon, that is the rest state, while Figures B.3 and B.4 show the contracted muscles for both linear and sphincter muscle polygons

- Assume a control polygon that comprises of n control points $p_i \in \mathbb{R}^3 (i = 0, \dots, n-1)$ for contraction value $c = 0$ (no contraction) and $c=1$ maximum contraction, a new polygon Q can be constructed such that $Q = \{q_i\}_{i=0}^{n-1}$ with each control point p_i assigned a parameter t_i such that

$$t_i := \begin{cases} 0 & , \text{ if } i = 0, \\ \frac{\sum_{j=1}^i \|p_j - p_{j-1}\|}{\sum_{j=1}^{n-1} \|p_j - p_{j-1}\|} & , \text{ else.} \end{cases}$$

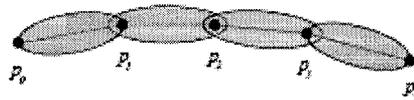


Figure B-2 Relaxed muscle fibre control polygon P [26]

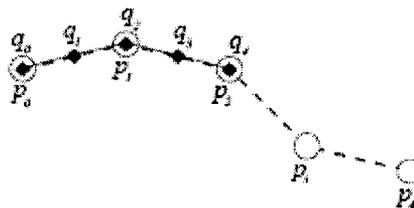


Figure B-3 Contracted linear muscle polygon q ($c=1/2$) [26]

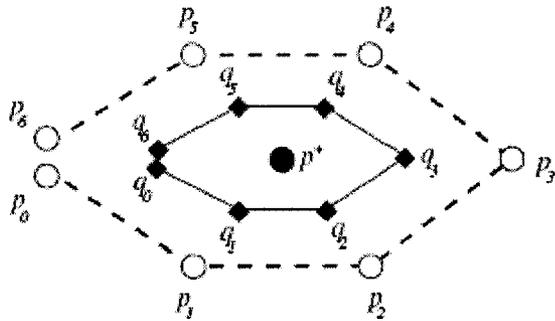


Figure B-4 Contracted sphincter muscle polygon q ($c=1/2$) [26]

- To avoid too much shrinking the parameters t_i are scaled by the contraction factor $1-c$ and then clamped to $[0.01, 1]$ i.e. $\hat{t}_i := \max\{(1-c)t_i, 0.01\}$
- Each parameter \hat{t}_i is mapped to the index $k_i \in \{0, \dots, n-2\}$ of the starting point

$$\text{of the segment that has } \hat{t}_i \text{ i.e. } k_i := \begin{cases} 0 & \text{if } i=0; \\ m : t_m < \hat{t}_i \leq t_{m+1} & \end{cases}$$

- Use linear interpolation to compute new control points q_i

$$q_i := p_{k_i} + (p_{k_i+1} - p_{k_i}) \frac{\hat{t}_i - t_{k_i}}{t_{k_i+1} - t_{k_i}}$$

- Sphincter muscles are contracted towards a centre point

$$q_i := p^* + (1-c) \| p_i - p^* \|$$

- Each muscle segment height $\overline{p_i p_{i+1}}$ is scaled by $(1+2si)$ to achieve bulging.

- Use affine transformation to transform the original segment $\overline{p_i p_{i+1}}$ to $\overline{q_i q_{i+1}}$ [2]

B.3. Zhang et al's muscle model

Zhang *et al.* [61] improved on Waters and Terzopoulos muscle model [62]. Their facial muscles modeling considered three muscle types namely linear, sphincter and sheet muscles.

B.3.1 Linear muscle model

The linear muscle model is shown in Figure B.5

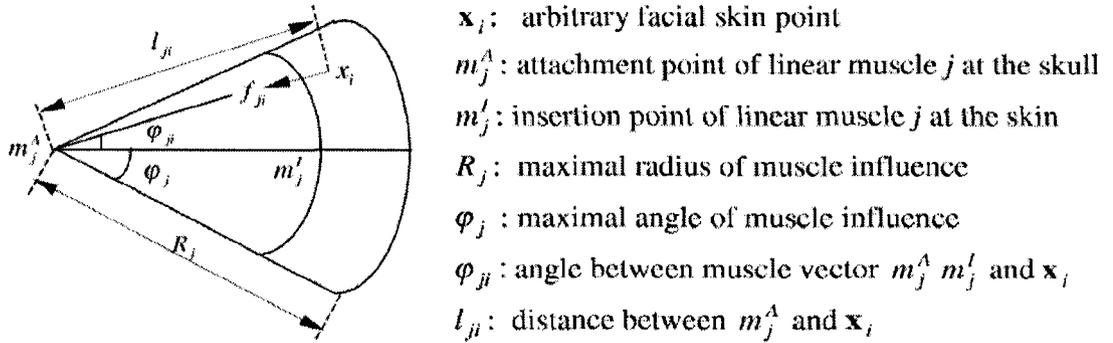


Figure B-5 Linear Muscle model [61]

From Figure B-5 above, The applied muscular force at vertex x_i is given as

$$\vec{f}_{ji} = \alpha_1 \Theta_1(\lambda_{ji}) \Theta_2(\gamma_{ji}) \frac{(m_j^A - x_i)}{|m_j^A - x_i|},$$

The length factor $\lambda_{ji} = \frac{l_{ji}}{|m_j^A - m_j^I|}$ defines the longitudinal distance between the muscle j and vertex x_i and the angular factor $\gamma_{ji} = \frac{\varphi_{ji}}{\varphi_j}$ defines the latitude distance between the muscle j and vertex x_i

α_1 is muscle contraction rate which controls the muscular force magnitude, the muscle forces due to length ratio and angular ratio are scaled functions Θ_1 and Θ_2

respectively and are defined as $\Theta_1(\lambda_{ji}) = \begin{cases} \cos(\frac{\pi}{2}(1 - \lambda_{ji}^{\eta_j})) \dots\dots\dots 0 \leq \lambda_{ji} \leq 1 \\ \cos(\frac{\pi}{2}(\frac{\lambda_{ji}^{\eta_j} - 1}{\delta_j^{\eta_j} - 1})) \dots\dots\dots 1 < \lambda_{ji} \leq \delta_j \end{cases}$,

$$\Theta_2(\gamma_{ji}) = \cos(\varphi_j \gamma_{ji}) \cos(\frac{\pi}{2} \gamma_{ji}) \dots\dots\dots 0 \leq \gamma_{ji} \leq 1$$

$$\delta_j = \frac{R_j}{|m_j^A - m_j^I|} \quad [61]$$

The strength of muscle j is defined by η_j which is inversely proportional to the muscle influence along the longitude.

B.3.2 Sphincter Muscle Model.

The sphincter muscle model is shown in Figure B.6

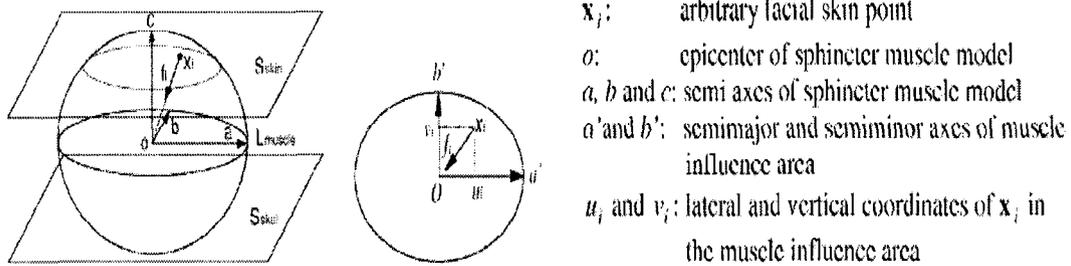


Figure B-6 Sphincter muscle model [61]

Applied muscular force at vertex x_i is $\vec{f}_i = \alpha_s \Theta(r_i) \frac{(o - x_i)}{|o - x_i|}$

where α_s is the contraction rate of sphincter muscle and

function $\Theta = \cos\left(\frac{\pi}{2}(1 - r_i)\right) \dots 0 \leq r_i \leq 1$ [7]

B.3.3 Sheet Muscle Model

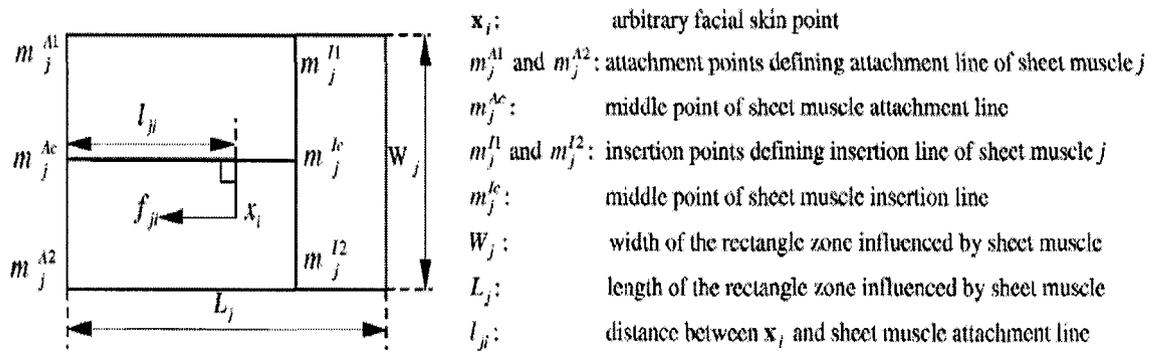


Figure B-7 Sheet muscle model [61]

Figure B-7 shows the sheet muscle model modeled as a series of almost parallel muscle fibres spread over a rectangular area. Applied muscular force at vertex x_i is given as

$$\overline{f_{ji}} = \alpha_{st} \Theta(\lambda_{ji}) \frac{(m_j^{Ac} - m_j^{lc})}{|m_j^{Ac} - m_j^{lc}|}$$

$$\text{function } \Theta(\lambda_{ji}) = \begin{cases} \cos\left(\frac{\pi}{2}(1 - \lambda_{ji}^{\eta_j})\right) \dots\dots\dots 0 \leq \lambda_{ji} \leq 1 \\ \cos\left(\frac{\pi}{2}\left(\frac{\lambda_{ji}^{\eta_j} - 1}{\delta_j^{\eta_j} - 1}\right)\right) \dots\dots\dots 1 < \lambda_{ji} \leq \delta_j \end{cases} ,$$

$$\lambda_{ji} = \frac{l_{ji}}{|m_j^{Ac} - m_j^{lc}|}$$

$$\text{and } \delta_j = \frac{L_j}{|m_j^{Ac} - m_j^{lc}|} \quad [7]$$