AN ANALYSIS OF THE CURRENT AND FUTURE STATE OF 3D FACIAL ANIMATION TECHNIQUES AND SYSTEMS

by

Chen Liu
B.A., Communication University of China 2006

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
School of Interactive Arts and Technology

SIMON FRASER UNIVERSITY
Fall 2009

APPROVAL

Name: Chen Liu

Degree: Master of Science

Title of Essays: An analysis of the current and future state of 3D facial animation techniques and systems

Examining Committee:

Chair: ______________________________

Dr. **Bernhard Riecke**
Assistant Professor, Interactive Arts and Technology

____

Prof. **Steve DiPaola**
Senior Supervisor
Associate Professor, Interactive Arts and Technology

____

Dr. **Chris Shaw**
Supervisor
Associate Professor, Interactive Arts and Technology

____

Dr. **Tom Calvert**
Internal Examiner
Professor, Interactive Arts and Technology

Date Defended/Approved: December-02-2009

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Simon Fraser University
Burnaby, BC, Canada

Last update: Spring 2010

ABSTRACT

Computer facial animation is now being used in a multitude of important fields. It brings a more human, social and dramatic reality to computer games, films and interactive multimedia, and is growing in use and importance. Authoring computer facial animation with complex and subtle expressions is still difficult and fraught with problems. It is currently mostly authored using generalized computer animation techniques, which often limit the quality and quantity of facial animation production. Given additional computer power, facial understanding and software sophistication, new face-centric methods are emerging but typically are immature in nature. This research attempts to define and organizationally categorize current and emerging methods, including surveying facial animation experts to define the current state of the field, perceived bottlenecks and emerging techniques. The thesis culminates in documenting this shared knowledge and making recommendations based on the data gathered, on possible new techniques for next generation, face-centric, computer animation systems.

Keywords: facial animation; animation systems; animation techniques; user survey; motion capture

DEDICATION

To my loving husband Lei Chen, who has supported me throughout this entire venture.

ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Steve DiPaola. His guidance has helped me complete this study as well as complete my entire graduate career. Thank you for your great help with the whole process from constructing a viable survey to finding professional subjects in animation.

Thanks also to the members of my committee and Tom Loughin who has assisted me through statistical analysis.

I also want to thank my parents who have raised me and have taught me from a baby. If it were not because of your teaching and moral example, I would not have pursued a Master degree. All my success in the future will also be your success.

In the end, I want to thank my husband Lei Chen for his support and love during my graduate studies. He gave me both economic and emotional support throughout these years.

TABLE OF CONTENTS

List of Figures

List of Tables

1: SURVEY OF 3D FACIAL ANIMATION

1.1 State of Facial Animation

1.1.1 Introduction

The living face is the most important and mysterious surface that human's deal with. "It is the showcase of the self, instantly displaying our age, sex, race, health and mood. It can send messages too elusive for science, and it bewitches us with its beauty" [McNeill 1998]. Every face is unique, 6.7 billion or so adorn the earth. It is a magnificent surface, and in the last 20 years, we have begun to understand the workings of the face more than the previous 20 millennia.

The rapid growth of visual communication systems has brought forth the emergence of virtual face-centric multimedia environments. Examples can be: video messaging on cell phones; dozens of online chatting expressions; faces in educational books, advertisements, video/online games; and online customer support agents. People are using virtual faces to send information, bypassing real human beings, for a variety of different purposes. Sometimes this is done online to protect a user's privacy; sometimes this is done in media like film and on television to obtain people's attention. Needless to say, computer facial animation is widely used in many industries, such as entertainment, customer service applications, human-computer-interaction and virtual reality systems. Research involving facial animation has been active in many disciplines, such as computer graphics, linguistics, and psychology.

Facial animation is one of the most complex, yet extremely effective communication tools in the creative animation field of realistic virtual avatars or social agents. It is very challenging work because it requires high quality animation involving every movement and element of the face such as facial muscles, facial bones and lip-synchronization of speech. Building such a sophisticated human face requires significant effort and time even for skilled animators. However, it is very important because facial expressions contain most of the information from which we recognize people's emotions and mental states. As Jeff Wilson, animation supervisor at Blur Studio said, "Creating believable facial animation is very important because the face is fundamental in understanding emotion" [Wilson 2005].

Due to our sensitivity towards facial expressions and the complexity of facial anatomy, it is generally accepted that there is not currently any real-time facial animation authoring system that can easily and expertly create realistic facial movement and consequently, subtle changes on the facial appearance of a character. Current facial animation systems either target a specific aspect or are unavailable to animators. This includes proprietary systems such as those only for internal use at game or film companies. Computer facial animation systems are hard to develop and are typically not extensible or easy to update with the latest advanced techniques. This current state of computer facial animations tools, where many different ad hoc or plug-in techniques must be used and mastered but typically only work in specific situations and applications, greatly effects individual researchers, animators and small studios who cannot afford to purchase and learn all the latest versions of the current systems. Furthermore, many of these systems or plug-ins rely on general or immature techniques that are not sophisticated or extensible enough to create high quality facial animation.

Based on our survey responses, and discussions with academic and industry facial animation professionals (animators, tool makers, managers, and researchers), we found that most believe that current facial animation systems lack a level of sophistication and specificity for facial animation needs. Furthermore, there seems to be a consensus in the field that given better scientific understanding of human facial processes, hardware speed increases and technical software improvements, along with developing 3D facial modeling and animation systems, we will, at some point, create real-time facial animation that can consistently generate believable facial emotions and expression sequences and thus, the subsequent models for them.

1.1.2 Computer Facial Animation History

The earliest work with computer based facial animation was done by Frederic I. Parke in the 1970s [Parke 1972, 1974]. Parke is considered the father of computer facial animation. The first computer-generated face images were done by Parke at the University of Utah in the early 1970s. There he used Henri Gouraud's smooth polygon shading algorithm to produce several segments of realistic facial animation. By 1974, Parke completed the first parameterized facial model, which is still generated today using photogrammetric techniques to collect facial expression polygon data from real faces and thus creates shape interpolation animations between facial expressions.

After this period, the three-dimensional facial animation field developed very fast. In 1980, the first physics-based muscle-controlled facial expression model was generated by Platt [Platt 1980]. Many major achievements have occurred since then. One facial animation milestone was the landmark film for facial animation called "*Tony de Peltrie*" in which facial expression and speech animation first appeared as a significant part of a short animated film at the SIGGRAPH 1985 computer graphics conference film show.

Keith Waters, who wrote the facial animation bible, Computer *Facial Animation,* together with Fred Parke [Parke 1996, 2008], also pioneered the muscle model approach in the mid 1980s [Waters 1987]. In it, a variety of facial expressions can be created by controlling the muscles underlying the skin. Next, we saw some development on the approach taken by Lewis [1987] and Hill [1986] to automatically synchronize speech and facial animation. In the milestone, Pixar's Academy Award winning short film, "*Tin Toy*", a muscle model was used to generate facial expressions. It is a powerful demonstration of the computer's capabilities regarding facial animation in a commercial medium. After this film, more experimentation and development on both facial animation techniques and facial animation movies were made, including movies involving facial animation such as "*Toy Story*", "*Shrek*" and "*Monsters Inc*". Meanwhile, facial animation was also used in the game industry and played an important role in computer games such as "*The Sims*" which used Artificial Intelligence techniques when generating families of corresponding related face models [DiPaola 2002].

Since the year 2000, facial animation has been extensively developed on many levels. One level is the appearance of many devices and systems that are used specifically for generating facial animation footage. For the movie "*The Matrix*", animators used optical flow techniques from several cameras to capture realistic facial expressions, which is part of the general category of performance-driven animation and is specifically a motion capture approach. Another movie, "*The Polar Express*", used a more standard type of optical motion capture system to record facial movements along with human body movements; although motion capture is still a quite immature technology for capturing faces. Another example of a milestone in the use of computer facial animation was in the film trilogy, "*The Lord of the Rings*". In these movies, new hybrid techniques like FACS from psychological research (discussed in a later section) mixed with motion capture and key frame animation helped bring facial animation to a new realistic and expressive leverage point. This hybrid technique was also applied in the movies "*King Kong*", "*Monster House*" and other films.

1.2 Application Areas

The largest application field for computer facial animation is certainly the animation industry. Facial animation appears in movies, and it develops, motivates and produces everything in the world of computer facial animation. Meanwhile, as a rapidly growing research area, facial animation has been adopted and developed in other industries or areas such as advertisements, games, web agents, as well as in medical and educational areas.

1.2.1 The Film and Advertisement industry

As the computer graphic technology developed, 3D animation films and visual effects in films became the facial animation's major application areas. The first time 3D characters were largely used in a movie was in "Toy Story" in 1995. It was also used in famous films like the "Lord of the Rings", "King Kong", "Finding Nemo" and most currently "Avatar". Numerous characters became famous, life-like and familiar to movie fans all across the world.

Another use is in the advertisement industry. Media professionals use 3D characters to produce advertisements for their clients. It is different and, at times, more interesting to use avatars than regular human beings. It can give visual effects that cannot be done by human actors and therefore, it can attract more attention as well as customers into new markets. Facial animation is not only about the 'human' faces, many likable animated characters and animals have been created. For instance, Telus is very well known for its different 3D animals in its Canadian commercials and as such, have become a trademark of their company.

1.2.2 Game Industry

Computer animation is also widely used in the game industry including computer/online games, video games and cell phone/PDA games. The close relationship between game design and computer animation forced these fields to develop interdependently and in a speedy fashion. facial animation in games is all about realistic human faces since a lot of games create characters like beasts or monsters. Needless to say, these animal facial expressions actually depend on the same facial animation techniques as human facial expressions.

One of the most famous facial animations applied to the character of a beast was created by Blur Studio using SoftImage's Face Robot facial authoring system in 2005 [Figure 1-1]. The animation was considered very realistic and lively, and its effects stayed consistent between movements. Although Face Robot is not available to all companies and users because of its cost and complexity, there are also other simple tools that can do a somewhat similar job. We can see more advanced facial animation techniques used on fantasy-based characters especially in the upcoming film, television and game fields.

For games that contain human looking characters instead of beast or fantasy characters, such as CS, FIFA or NBA live, there is also space for facial animation. As the recognition that communication between players is more and more important, game companies are providing facial expressions and dialogue functions for users to use when communicating between players. For Instance, when users win a game or progress to the next level, celebratory actions can be completed by the character which involves facial animation and communication. Overall, there are numerous possibilities for facial animation in games featuring non-human faces as well as games featuring human faces.

Figure 1-1 Rock Falcon, the poster boy represents Softimage's Face Robot technology.

1.2.3 Medicine

Computer animation plays a diverse role in medicine. For example, facial animation can support realistic models for medical research, teaching or clinical medicine. It is a new way to help doctors diagnose and treat patients without being limited by time or space, macroscopically or microcosmically. It can provide images and films, or real-time videos of the inside human body, which often can neither be seen by our naked eye or by photos.

Surgical simulation is another aspect of using facial animation in medicine. Facial animation can simulate the facial tissues of the human faces which are great helps in surgeries under the skin.

1.2.4 Social Agents and Avatars

Agents and Avatars are the new era of computational social sciences. The principle of social agents relies on the interaction between human beings. Agents can exhibit emotions and personalities. Agent systems can be as simple as following your instructions to open a folder or as complex as a personal director giving audible or visual responses or guidance, with proper characteristics and behaviours following your voice orders.

A lot of effort has been put into creating models that can heighten the believability of animated agents in social science. The development of needing interactive social models that contain social behaviours makes for another area of facial animation application.

1.3 Thesis Summary

Because of the widely used and rapid growth of computer facial animation, this thesis attempts to document this field and based on this shared information of current states of facial animation techniques and systems to perform a survey that help researchers and animators for future development and design. In chapter 2, the diverse and often overlapping categories of facial animation techniques are described and organized. In chapter 3, a brief description of facial animation systems and plug-ins are described and organized, showing how these mixed and emerging techniques are used in current commercial and academic systems. With the goal of contributing valuable suggestions regarding both the current and future trends of facial animation, given the current diverse and unclear situation, chapter 4 describes our original computer facial animation survey. The surveys goal is to 1) •to understand the subjects' most used and preferred facial animation techniques and system types, 2) to understand their most common issues on the techniques and systems they use and 3) to gather their thoughts on future trends in facial animation techniques, functions and interfaces. With these goals in mind, chapters 5 and 6 give recommendations based on the survey data and results as well as end with concluding remarks.

2: FACIAL ANIMATION TECHNIQUES

When attempting to classify the categories of facial animation techniques, it is very difficult to precisely classify them because there is not a specific boundary between two methods, and recently most techniques actually synthesize several methods together to achieve better results. It is also rapidly changing, and fraught with archaic and informal name conventions. It is in fact one of the goals of this research to bring to light the many categories, naming conventions, and actual technical descriptions of these varying techniques in one place.

Traditional facial animation methods involve blend shape or shape interpolation known as morph target or moving vertices, which are curves or meshes of 3D face geometry. One common problem of these methods is that they can only affect one facial region at a time yet facial regions are intensely correlated with each other. To achieve realistic effects, animators need to be very careful around the edges of different facial regions -- the details and creases are extremely hard to represent naturally. Hence, newer technologies are being developed to achieve better results and provide more intuitive ways to create facial animations.

Currently there are many types of mixed and grouped techniques as well as a large number of animation systems using different forms of these techniques. First, large 3D package tools such as Autodesk Maya, 3DS Max or

Softimage use traditional shape interpolation, parameterization deformation or bone based techniques. These methods are being used widely to create both face and body animations. On the other hand, a number of small stand alone systems or academic systems that are specific for facial animations have abandoned these methods and have turned toward newer methods such as performance-driven, physics muscle simulation, expression coding and lip synchronization technologies. Currently large 3D tool makers have yet to fully apply these newer techniques to their systems. However, motion capture is becoming one of the most popular techniques in the field. Of course, as animators continue upgrading their skills and methods, they are also adventuring far more in this area. Lately, animators are concerned with achieving realism, eye and head synthesis, and transferring existing animations from old models to new models.

2.1 Shape interpolation or Blend Shapes

Blend shape is the most widely used animation method. Blend shape animation can also be considered as shape interpolation animation. It is applied by many commercial animation software packages such as Maya and 3D Studio Max". It basically involves distorting one shape, and at the same time, fading into another through marking corresponding points and vectors on the "before" and "after" shapes used in the morph. The principle of this method is that animators create several key poses of a subject and the animation system automatically interpolates the frames in- between. Technically, blend shape animation is point set interpolation where an interpolation function (typically linear) specifies smooth motion between two key sets of points, generally from models typically of the same topology [Pighin 1998]. This animation process involves modeling polygonal meshes of a human face to approximate expressions and visemes (lip positions that match speech) [Osipa 2007] and then blending the different sub meshes automatically, known as morphs targets or blend shapes. Because the blend shape method generally uses a linear algorithm of topology of the face meshes, animators can control the weights applied on the vertices of polygonal models and thus control the degree to which it is blended. Therefore, most of these automatically generated frames are simple transformations of movement, scale or rotation of rigid objects such as hierarchical parts of the human body like the arms or legs. However, there is not any specific, rigid part on the human face so the challenge of shape interpolation is that it really does not work very well on facial animation due to the complex construction and curves of the face, this method requires animators to adjust settings specifically to each face model and is labour intensive. It is especially hard to work on a model of a realistic human face rather than the face of a fantasy 3D character. For instance, we can create expressions on avatar's faces and we can exaggerate their expressions, and people may still think it is believable, such as the character Gollum, from "*The Lord of the Rings*". But when we are trying to achieve realistic human face animations, rigid interpolations generated by the blend shape approach will impair the authenticity of the animation. To solve this problem, a many techniques (5-10 approaches) based on shape interpolation are being generated to get a smooth animation of the human face.

One widely-used basic technique is to create a model of a face in a rest position. Then, using 3D modeling techniques, edit the points of a copy of that face to make other faces, typically with the same topology, and hence copy the rested face into different phoneme and expression states. Then, animate a facial animation sequence by morphing (point interpolation) between these sets of similar faces. The major disadvantage with this technique is that the animator is only picking from a set of pre-made face morphs and thereby limited to the expressions possible from that set in the final animated face sequence. There are several variants of this blend shape technique – most notably compound or hierarchical morph targets which allow the animator to blend several models together using differing weightings on selected areas of the face models. Again, all versions of this technique limit the creative process by only forcing the animator to pick from a pre-made set of expressions or cause the animator to stop the animation process and go back to a modelling process to create additional morph targets.

Other interpolation techniques, besides the linear methods, were designed for higher quality animations. Not only do we have key expression interpolation but we also have bilinear interpolation, pairwise interpolation or n-dimensional interpolation. For example, cosine interpolation function [Waters 1993] or spline interpolation can create faster animation effects at the beginning and slower effects at the end of an animation. Even bilinear interpolation can be used to generate more possibilities of facial expression changes [Arai 1996]. Although these interpolations are easy to compute, the set up before blending them together, giving the ability to generate a wide range of realistic facial expressions, is still limited. Besides, during the process, animators are always going back and forth to adjust the weights of the blend shapes and create more and more key poses to improve the parameter control effects, albeit only within a small range.

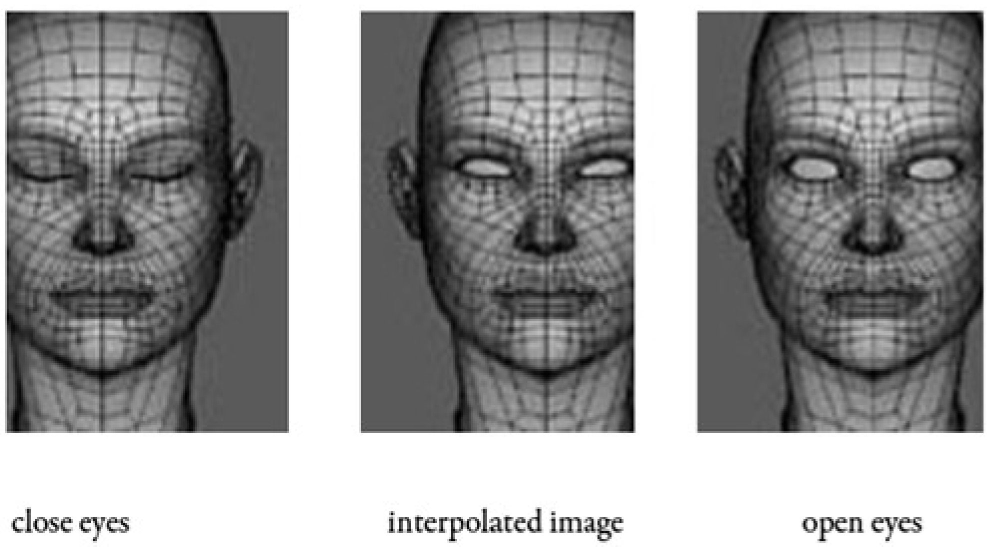


Figure 2-1 Linear interpolation is performed on eyes from close to open where interpolated_eyes(t) =(1-t) * closed_eyes + t * open_eyes. T is between 0 and 1.

2.2 Parameterization

Parameterization overcomes some of the limitations and restrictions of simple interpolations [Parke 1996, 2008]. Animators can control parameter values to control expressions. Fred Parke's early work in facial animation at the University of Utah and the NY Institute of Technology led to the first facial animation system. It used a control parameterized method, whereby the animation becomes a process of specifying and controlling parameter set values as a function of time. Ideal parameterizations specify any possible face and expression by a combination of independent parameter values [Parke 1996, 2008]. One limitation of shape interpolation is that animators cannot control the interpolation process of the animation. It is manipulated by the computer automatically. Although the computer will calculate and provide a proper interpolation, it limits the animators' manual adjustments. Parameterizations allow animators to have more freedom and specific control of facial configurations. With the parameters combined together, we can get a large range of facial expressions with relatively low computational costs.

However there is one big restriction on parameterization which limits its uses. Because it usually works on small facial regions, conflicts may happen. For example, as these parameters can only affect specific facial regions, there will be noticeable motion boundaries between these facial regions. Therefore, animators may not get enough freedom during parameter configurations.

FaceLift program used for "*The Sims*" [DiPaola 2002] uses a version of a production parameterization system which uses genetic programming and hierarchical parameterization to improve on past systems. Most parameter systems use Paul Ekman's FACS (Facial Action Coding System) [Ekman 1978] which describes facial muscle movement as a basis or starting point for specifying and defining the entire range of parameters. Ken Perlin's web Java-based system also uses a simple but very effective parameterized technique.

Parameter systems create very compact specifications for facial animation professionals and are therefore ideally suited for the web and games. And because all the parameter is related to the face model, it's highly dependent on the facial mesh topology. Therefore, parameterization is not a general method that can control the whole of animating the face model overall. The choice of the parameter setting requires the tedious manual tuning of each value for each specific area, but not the general control. Besides, due to the fact that it only affects specific facial regions, conflict may happen during set up. Hence it is also another time consuming method which does limit its general use; on the other hand, it leads to the development of other technologies such as deformation based technology. Recently, hierarchical and multi-dimension parameter space techniques have been introduced to overcome some of these issues [DiPaola 2002, Ayra 2007].

2.3 Bone-based Rigging Methods

Instead of creating from a million blend shapes, animators rig an entire facial setup to perform better animation results. This technique was developed later than blend shapes and is another widely adopted method [Schleifer 2004]. You can find it in many systems such as Maya, 3Ds Max, etc. It solves a lot of problems unsolved by shape interpolation. They can adjust all expressions on one fully rigged face model.

But they all require large amount of effort. The difference between face rigging and blend shapes is that when you are using blend shapes to create facial animation you have to create hundreds of key poses to assure the computer will blend them smoothly on every single part of the face and every single frame of the film. Even though, with facial rigging, it is certainly very hard and time-consuming to set up the bone structure for an entire face, once you have the set up you can control these bone structures and create animations easily afterwards. As well, it gives you more freedom when generating different expressions, and if you have done something wrong you can always go back and recreate it without costing too much time regenerating the whole expression. The reason being is that it is all done on the same face model and the bone structure has already been set up. You can do the adjusting and animation task based on this foundation. It is not as difficult as creating a brand new blend shape from zero.

One early example is a system developed by Hanrahan and Sturman, [Hanrahan 1985] which allows the animator to establish functional relationships between interactive input devices and control parameters. Modern interactive animation systems provide a wide variety of ways for linking user interactions to animation actions [Parke 1996, 2008]. The interaction can be through the use of articulated joints, the use of clusters or the use of interactive curve editors.

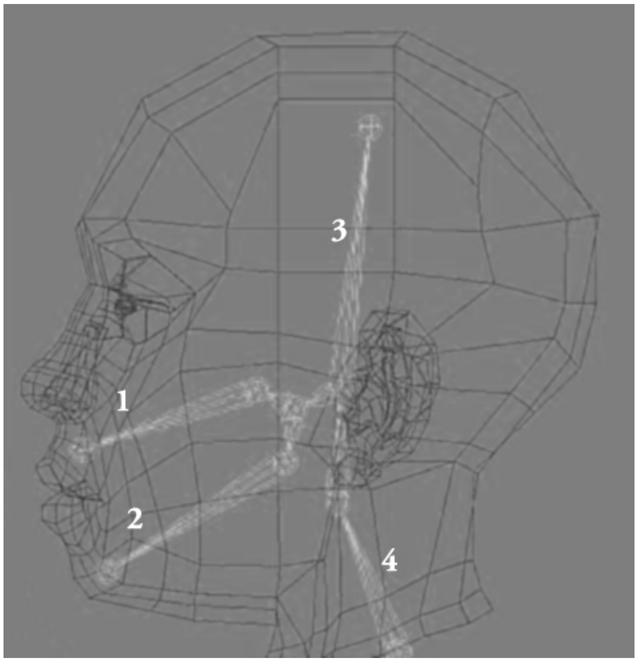


Figure 2-2 A simple illustration of how to use bones to create expressions. By dragging joints 1 and 2, one can control the open and closed mouth of the model. Also, dragging joint 3 will make nodding gestures. Joint 4 controls the motions of the neck. More joints can be applied together to gain better control over the model. Each joint controls a corresponding part of the facial region. One can drag these joints and create various expressions.

2.4 Deformation-based Approaches

This technique involves deformation on solid geometry meshes. The process only affects the thin shell mesh and ignores the muscle structure underneath. The facial expressions are all achieved by manipulating these surface meshes manually. Hence it can control all kinds of surfaces, planes, quadrics, surface patches or polygon meshes. This category includes morphing between different models and free-form deformations.

2.4.1 Free-form deformation (FFD)

Free-form deformation (FFD) [Sederberg 1986] is a very popular tool in computer animation. It uses geometric algorithms to change or define the shape of a 3D model. It involves implementing a user interface that animators can use to modify shapes by moving the control points around the objects. The control points are arranged in a three-dimensional cubic lattice, while the lattice changes, the object changes accordingly. The lattice is three dimensional which means animators have to perform a three dimensional manipulation on a two dimensional interface. Ideally, animators can flexibly rotate this lattice within a three dimensional environment. But working in a two dimensional environment means you can only see one side at a time. It can be a challenge to make sure the front side and the back side is the same by hand. We could also depend on some parameter value adjustments, such as moving all the control points to a certain level without varying them. But that is not as intuitive as hand manipulation. FFDs can deform many types of surface primitives, including polygons, quadric, parametric, and implicit surfaces; and solid models. [Parke 1996, 2008]

The advantage of free-form deformation is that the manipulation of deformation is independent of surface geometry. Hence it is not related to the meshes and vertices. While creating a model, animators do not need to worry about future deformation. Because the manipulation is on the control lattice directly and the object indirectly, they can always go back or delete that lattice if they are not satisfied which will not cause any unrecoverable deformation of the objects. Also, the way to control the lattice is very intuitive. Simply pulling or dragging can modify the shape of the lattice and the object at the same time, which makes it very efficient and flexible.

On the other hand, due to the shape of the lattice, free-form deformation prohibits arbitrary shape deformation. It is not applicable on any kind of object. For instance, it would be difficult to create a torus on the surface. More precise deformations can be achieved if we make higher point counts and more calculations, which the computer could make along with it. More control points produce more precise deformations. As more of the control points are added, the object will also be harder to clearly see which would be another problem. Due to the fact that the lattice is covering the whole object and the control points on the lattice actually affect each other, sometimes when you deform one part of the object you will affect other parts. And this corresponding outcome may not always be the result you are expecting. In the case where you want to affect the arm of the character only, you may affect the face by using FFD. For instance, if you select the control points on the top facet to deform the cap of the teapot, as displayed in the picture, the spout of the teapot scales too.[Figure 2-3] It also tends to flatten the object, which is another limitation due to the fact that we are only able to work on a two dimensional plane. FFDs have also been used for simulations involving facial motion actions [Kalra 1992].

2.4.2 Extended free-form deformation (EFFD)

Extended free-form deformation (EFFD) extends the control point lattice to a cylindrical structure [Coquillart 90]. Cylindrical lattice provides for a better deformation on objects that are not cubic, such as a basket, a wheel, an arm, or anything that is not regularly cubic [Figure 2-4]. Also, it gives more flexibility to animators when they want to change shape deformations, not including cubic lattices. Similar to FFD, we can control the segments on both the vertical dimension and the surface dimension to control the precision of the lattice, which better deforms the object in it. As a result, the deformation is no longer dependent on the specific surface itself but depends on the parameters of the lattice, which is obviously easier to manipulate.

Again, there is a positive side and a negative side to extended free-form deformation due to the simple and intuitive manipulation on the lattice instead of on the surface itself. It is simple and easy to go back to the history, but at the same time, it does not provide a precise control of the actual surface. One point to remember is that this technique continues to cause the development of other facial animation techniques.

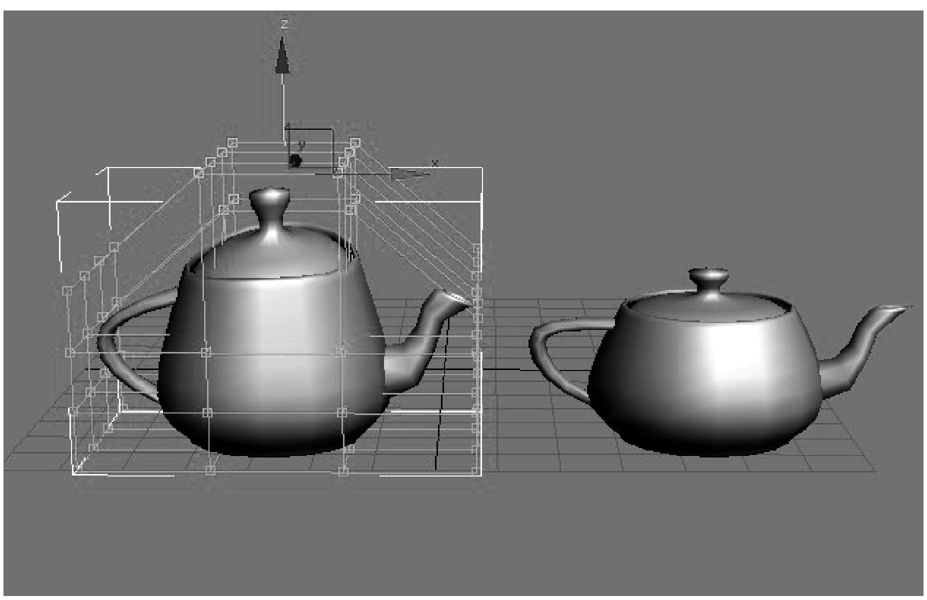


Figure 2-3 Free-form deformation. Select the controlling points of the lattice to deform the embedded object. When the controlling points are changed, so is the embedded object.

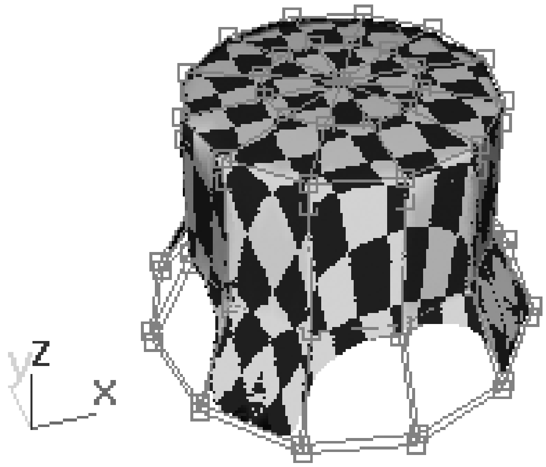


Figure 2-4 Cylindrical free-form deformation. Performs better on cylindrical objects than on cubic lattices.

2.5 Expression Coding Approaches

Expression coding techniques have provided a database of muscle actions of all possible facial expressions. Animators can compile desired expressions and let the computer animate them. The most popular expression coding systems are facial action coding systems and MPEG-4 facial animation models. Many other technologies involve expression coding in their applications.

2.5.1 Facial Action Coding System

The Facial Action Coding System (FACS) is a system measuring and describing facial behaviours through understanding each facial muscle, developed by Paul Ekman and Wallace Friesen in 1976. The first FACS Manual was published in 1978 [Ekman 1978]. It is a common standard about how each facial muscle changes the appearance of the face. It's derived from analysis of facial anatomy described how human face's muscles behaviour including jaw and tongue movements.

By studying facial anatomy, we can understand that facial expressions are caused by changes in facial actions. And to understand facial behaviours, FACS begins working from facial actions. Unlike other muscle-based techniques, FACS does not directly study muscle, but instead goes to another level which is the actions that are caused by the muscles. Since facial expressions are complex, and no expression is made only by one muscle, different expressions can use the same muscle. It is a very complex configuration. Hence, the study of each individual muscle would be difficult to facilitate a clear understanding of this area. However, facial action units (AUs) are created according to these actions and each AU can be involved with several facial muscles because if one expression is made, all of the muscles involved cannot be distinguished easily. Vice versa, one muscle can also be separated into two or more AUs to describe relatively independent actions of different parts of the muscle. FACS breaks down the human face into 46 action units. Each unit represents an individual muscle action or a set of muscles that represent a single facial posture. The principle is that each AU should be a minimal unit that cannot be divided into smaller action units. By elaborately categorizing different AUs on the face, FACS was capable of mimicking all the facial muscle movements.

Different combinations of action units generate different facial expressions. For example, combining the AU1 (Inner Brow Raiser), AU4 (Brow Raiser), AU15 (Lip Corner Depressor), and AU23 (Lip Tightener) creates a sad expression. A table of the sample action units and the basic expressions generated by the actions units are presented [Tables 1] [Table 2].

FACS was not originally designed to help make facial animations. It was intended only as a way to score or describe facial movements. But due to its wide range on as well as acknowledgement for facial actions, it has been used extensively in facial animation during the past few decades. FACS helps animators represent and create realistic facial expressions by describing all possible facial animation units performed by the human face including descriptions for measuring head and eye positions. The weakness of FACS is that it deals only with obvious facial changes and ignores all the subtle changes that might be difficult to distinguish. Also, it ignores any movements underneath the skin which are not visible but can affect facial expressions. FACS is very successful in distinguishing the actions of brows, eyelids and the forehead area, but not evenly concerned about the whole face. Especially, it does not include all of the visible, reliable, distinguishable actions of the lower part of the face which makes it not able to do the lip-sync requirement.

As lip-synchronization has become more and more important and has been obtaining most of the attention of researchers, FACS has showed its inferior position in this field due to its lack of muscle control over the lower parts of the face. Future facial animation techniques should be able to synchronize speech animation with voice easily. That limits FACS's ability to become the dominant techniques in the industry.

In summary, FACS covers most facial movements and expressions by coding facial actions. It can provide very reliable face codes on the upper parts of the face including foreheads, eyebrows and eyelids, but it does not provide sufficient face codes on the lower parts of the face. Although, FACS still overcomes the limitation of interpolation and has been a landmark in facial animation techniques. It surely encourages other similar techniques and is widely utilized with muscle or simulated muscle-based approaches.

AU	FACS Name	AU	FACS Name	AU	FACS Name
1	Inner Brow Raiser	12	Lid Corner Puller	2	Outer Brow Raiser
14	Dimpler	4	Brow lower	15	Lip Corner depressor
5	Upper Lid Raiser	16	Lower Lip Depressor	6	Check Raiser
17	Chin Raiser	9	Nose Wrinkler	20	Lip Stretcher
23	Lip Tightner	10	Upper Lid Raiser	26	Jaw Drop

Table 2-1 Sample single facial action unites

Basic Expressions	Involved Action Units
Surprise	AU1,2,5,15,16,20,26
Fear	AU1,2,4,5,15,20,26
Anger	AU2,4,7,9,10,20,26
Happiness	AU1,6,12,14
Sadness	AU1,4,15,23

Table 2-2 Example sets of action units for basic expressions

2.5.2 MPEG-4 Facial Animation

The MPEG-4 is an ISO standard for multimedia [MPEG41997]. It was released in 1999 and since then many research topics have focused on this standard because of the wide range of audio and video, as well as 3D graphics, it can be used in. It is also the only standard that involves facial animation hence new methods have been developed based on the MPEG-4 standard. In the area of facial animation, some research has been done to represent and describe certain facial actions with predefined sets of “codes”. The Facial Action Coding System is probably the first successful attempt in this regard (although not directly graphics and animation research). More recently, the MPEG-4 standard has been concerned more about the communication and integration of multimedia content and is gaining more and more attention in facial animation domains.

MPEG-4 Facial Animation (FA) defines various parameters of a talking face in a standardized way. I.e. it has defined Face Definition and Animation Parameters to encode facial actions. The head was divided into 84 feature points (FPs), each feature point describing the shape of a corresponding area of a standard face model. Therefore, they can be used to define the animation parameters on the face in order to conform to this standard when switching between different models. In this standard, there are 68 universal Facial Animation Parameters (FAP) on the face. Every FAP corresponds to a FP and defines low level facial actions like jaw-down, and higher level, more complicated actions like smiling. The FAP is independent of model geometry which means it needs to be calibrated before being applied on different models. This is done by using Facial Animation Parameter Units. This unit is actually the distance between facial features. Therefore FAPU is not a universal parameter; it is specific to the 3d face model that it is applied onto. So when we have a standard FAP, a corresponding FAPU, and corresponding FPs, we can adjust and decide the values on a new model freely exchanging information from the face models.

Together they allow a receiver system to create a face (using any graphics method) and animate that face based on low level commands in FAPs.

The Facial Animation Parameters are composed of two types. The first two are high level parameters that can represent facial expressions by themselves. It includes six basic emotions: anger, joy, sadness, surprise, disgust and fear. Animators are allowed to use only these two parameters to produce a relatively good animation by doing linear interpolation between each expression. The other facial animation parameters focus on specific regions of the face, for instance, the right eyebrow, left corner of mouth, bottom of the chin, etc, that can encode facial actions of a higher degree, with more sophistication. It should be noted that FAPs do not need to be used with a synthetic face and geometric models. They are independent of animation methods and simply define the desired movements. They can be used to apply pre-learned image transformations (based on detected location of facial features) to a real 2D picture in order to create visual effects like talking, facial expression, or any other facial movements [Ayra 2003] [Ezzat 1998].

MPEG-4 also provides a scripting language that can generate parameters. The researchers also invented some similar languages, for example, VHML, APML, BEAT and XMT as a framework for incorporating textual descriptions in languages like SMIL and VRML. Together MPEG-4 face parameters define a very powerful low-level mechanism to control facial animation. However, MPEG-4 FPs and FAPs do not provide an animation language but only a set of low-level parameters. Although they exist as powerful means in facial animation, this mechanism lacks higher levels of abstraction, timing control, and event management. The content providers and animation engines still need higher levels of abstraction on top of MPEG-4 parameters to provide group actions, timing control, event handling and similar functionality usually provided by high-level language. In other words, MPEG-4 face parameters do not provide an animation language; whereas, XMT languages do not include any face-specific features, yet.

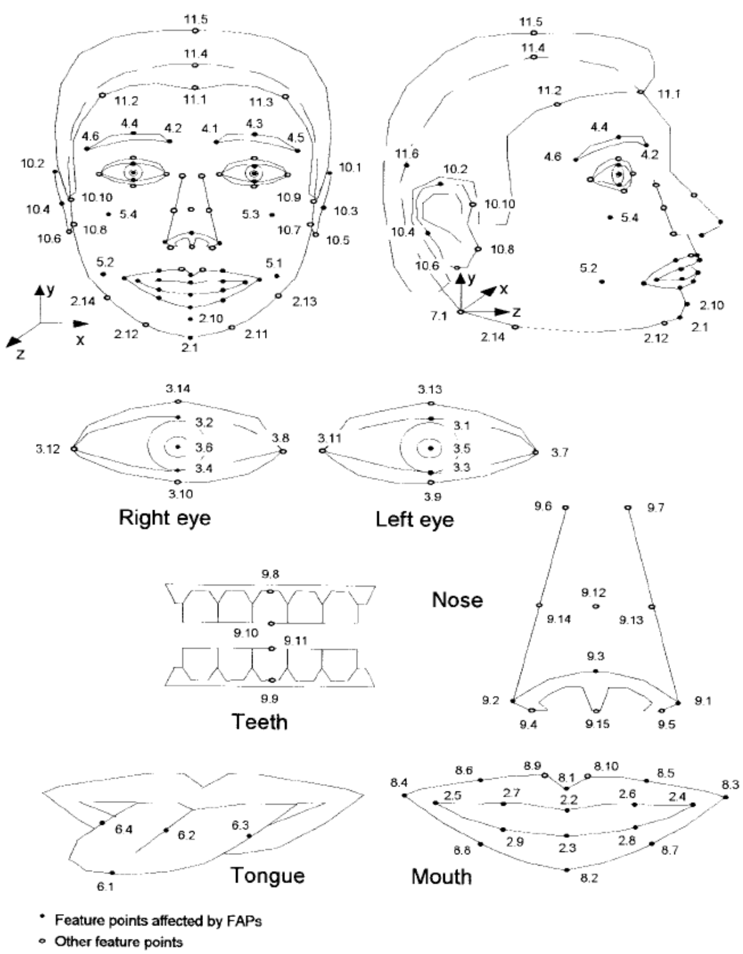


Figure 2-5 MPEG-4 Feature Points.

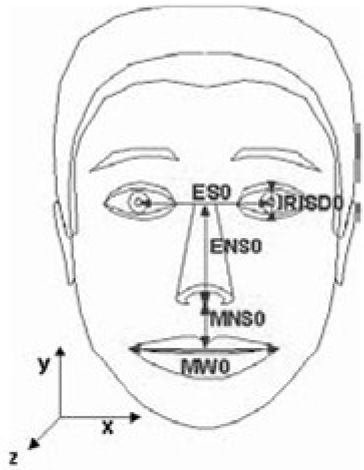


Figure 2-6 MPEG-4 FAPU description.

2.6 Muscle-based Modeling Methods

The human face is composed of rigid and non-rigid components, simple skeleton-based models often limit simulations of the face, while attempting such an anatomical model is challenging concerning these physical attributes. Until now, there is no model exactly representing the real structure and skin of the face. As a result, we have to sacrifice some true physical attributes and attempt to simulate as closely as possible to the real anatomy of a face. Muscle based simulation can help with this complexity. Generally three simulation methods are used for simulating muscles: mass-spring methods, vector simulations, layered spring methods.

2.6.1 Mass-Spring Models

Mass-spring methods represent solid objects by connecting point masses by springs. It simply simulates skin deformation by forcing these points into an elastic spring mesh. It was first proposed by Nedel and Thalmann [Ned 1998] modelling a real-time muscle deformation using a mass-spring system. They created action units to force the meshes into a deformable spring network. Waters also developed a mass-spring model for skin deformation. It allows for pulling linear muscles and squeezing sphincter muscles. The model's action unit is based on FACS. Platt and Badler [Platt 1981] are also focusing on muscle modeling since 1980s. Their model consists of 38 regional muscles interconnected by a spring network with action units constraining the deformation of the spring network.

The limitation of this method is that the deformation of an object depends on how the spring network is set up. And it is difficult to adjust the spring constants if different deformation is desired. Mass-spring methods cannot capture volumetric effects. But it is easy to implement and it is fast.

2.6.2 Vector Muscles

The muscle vector method deforms a facial mesh through muscle motions due to the vectors' influence with different directions and magnitude in both two and three dimensions. This method classifies muscles into linear and sphincter. Waters proposes a model, in which the muscles are independent from the underlying bone structure, with vector properties following the direction and insertion of muscle fibres [Waters 1987]. The vector controls directions through a point connection on the bone, and the magnitude of displacement through a muscle spring constant. Hence, each muscle is independent of the geometry and influence of a certain region. Muscle-based models only use limited parameters and are independent from facial topology. The Pixar movie "Tin Toy" used vector muscles when creating the baby's facial expressions.

2.6.3 Layered Spring Mesh Muscles

Since the mass-spring method cannot capture volumetric effects, layered spring mesh extends into three connected mesh layers to simulate the volumetric deformations, for the purpose of creating a better anatomical model. Terzopoulos and Waters [Terzopoulos 1990] proposed a hierarchical model of the human face that incorporates an approximation of both the anatomical structure and the dynamics of the human face. The three layers are skin, fatty tissue and muscles tied to bones. Lee also presents a face model that simulates the human face with physics-based synthetic skin and muscle layers, which do have constraints and emulate facial tissue for better movements and work according to the skull structure when making facial expressions [Lee.YC 1993, 1995]. This non-linear deformation property makes the tissue looks biologically good. The skull structure beneath the muscle layer enables the tissue to slide over it without penetrating it. The muscle layer knits them together. Since more layers are involved, this method costs more in regards to computation efforts.

2.7 3D Face Modeling

3D face modeling is the foundation of all possible facial animations. No matter what facial animation techniques people use to create styles or realistic animation, it all depends on a well measured and generated face model. If the face model is poorly created, such as in poor topology, and the meshes and controlling points are not suitable for animation techniques, it will come out to be a terrible model for animation. Hence 3D face modeling is the most significant part of building facial animations. Everything is constructed on this foundation. There are tons of existing systems and methods for modeling 3D faces and a few of them have become the most widely accepted systems and methods, for instance polygonal modeling, surface modeling or sub-division surface modeling.

There are several types of modeling data that are being used by major modeling software. For the film industry, people usually create polygon models first and convert it to nurbs data because nurbs data is faster when rendering and can achieve better results. For instance, if you are using the commercial rendering systems Renderman or Mental Ray, using nurbs data is nine times faster than using polygon data. But that does not mean nurbs data is better than polygon data. Actually all these different ways of modeling data are for different usages. When creating models for games, polygon is the first choice that animators choose because creating simple polygon models is faster than creating nurbs models. Furthermore, polygon models can represent accurate appearances that are not applicable for nurbs models such as machines or skyscrapers in the fields of architecture and mechanics. .

The difficulties of making facial animation depend on modeling regardless of what method people use to make animation. All the elements in modeling such as surfaces, polygonal, and sub-div surfaces do matter in the following steps of making facial animation. Even though we can develop plenty of approaches to make animation easier, they all rely on the model we create. Thus fundamental models need to be created along with the techniques of animation. Modeling has been spotty since the last century and there are lots of existing modeling systems in the research area such as Mudbox and Zbrush. Although modeling a specific person is a very significant problem in facial animation, especially modeling the 3D geometry of an individual human face, people have invented a few methods and applications for it such as a range scanner, digitizer probe, and stereo disparity that can measure three-dimensional coordinates. However, the results of these products are still not suitable enough for facial animation because it is important to not only include the facial surface data in facial modeling but also the data about facial structures. Besides, almost all the data from these products need to be reconstructed either due to part of a face is missing or the data distribution is poor or there are a lot of noises to clean and separate. There is no doubt that it is going to take a lot of time cleaning this data afterwards. These products generally do not provide color or texture information either.

2.7.1 Regular Modeling Methods

2.7.1.1 Polygonal Surfaces

The Polygonal surface is the most early and widely used modeling method for three dimension characters. Many famous characters and films are made based on polygonal surface descriptions. Polygonal modeling specifies exactly each 3d point, which are connected to each other as polygons and approximating modeling objects' surfaces using this kind of polygons [Gouraud 1971]. This is an exacting way to get topology (points) where you need it on a face rather than where you do not. These polygons are laid out in a way that allows good approximation of real facial muscle distributions [Figure 2-7]. With polygonal models, you can get specific control and placement of topology and mapping. Also, the topology of the polygons is required as to be manipulated naturally and flexibly by the animators. Because of good topology, animators can find it competitive when mapping textures on it compare to other modeling methods. Compared to free-form deformation, animators can manipulate every single vertex or edge directly to get precise shape deformation. Besides, polygonal modeling is well suited to scanline rendering and is the best method for real-time computer graphics. Due to this simplicity, it saves computational calculations while rendering polygons. So polygonal modeling is faster than other representations.

There are also many disadvantages using polygons to represent an object. The first problem is it is challenging to work with polygon data as you must work with it directly. Rather than nurbs modeling with a controlling point, polygonal modeling requires animators to manage vertices one by one. It certainly provides precise control of the vertices but it can also bring overwhelming adjustments while working. Due to the properties of polygons it is incapable of accurately representing curved surfaces. The resolution uses a large number of smaller linear polygons to approximate that curved surface. This representing method combined with adjusting vertices directly increases workloads immensely. Therefore, it is not applicable on all facial regions. Due to its rectangle shape, it usually performs unsatisfactorily when representing sphincter muscles. Triangular meshes or multi-angular meshes may be created on sudden change borders, eye areas, mouth areas or ears. These areas are very difficult to model using polygonal meshes because of the sophisticated surfaces. Not only does it require a large number but also will cause evident creases after rendering in the software. Using polygon meshes to model a face, foursquare meshes are the best choice when considering post rendering effects. Besides, it is against the physical reality of muscle distribution to use foursquare mesh to represent muscles that are not foursquare and that have acute corners. And eyes and mouths are so important that they are the major expressive areas of the face. Also, hair has been and continues to be a challenge for animators to model. Polygon is not a good representation for that either. Hence polygons are not the choice for character animations, organic objects, and facial animations. But polygons will still be spotted by researchers and users in the future, because of its simple and fast rendering attributes and it has lasted for a very long time in the history of face modeling. Almost all the large platforms for face modeling still have polygons as a big component of their systems. It is a good representation for models that does not require high resolution for smooth surfaces.

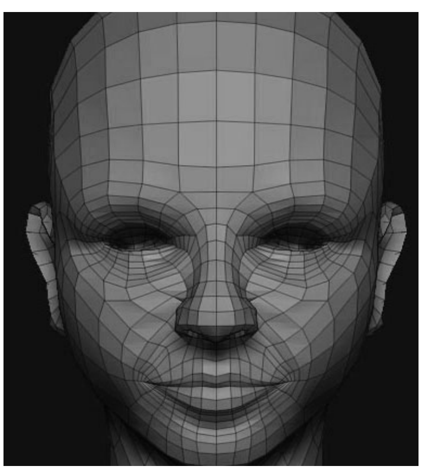


Figure 2-7 A facial topology simulating general muscle directions using polygons.

2.7.1.2 NURBs

Non-uniform rational B-spline (NURBS) is another method for modeling faces and characters. Unlike polygons, NURBS generates smooth curves and surfaces [Valle 1994] [Perez 1994]. It is designed similar to free-form deformation where animators work with controlling points to change the curve and surface shape indirectly. The interaction with NURBS is intuitive and predictable because the control points are connected directly to the curves or surfaces rather than a lattice or a box away from your objects. Hence they can be efficiently handled and easily determined by animators. And no matter how complex the curve or surface is, the controlling points are always a small number. These controlling points have different weights, so they have different effects on the curve or surface you are manipulating. The computer changes the shape of the curve based on the changes of the controlling points.

Usually several NURBS curves compose patches. These curvilinear patches are suitable for representing smooth faces. And it usually takes only a few patches to model a face [Figure 2-8]. In the baby's figure, four NURBS curves compose a large patch that simulates the cheek of the baby's face. If you use polygon mesh to simulate exactly the same size and place of the cheek, to achieve such a smooth effect, it may involve several dozen meshes. But with NURBS, due to the smooth attributes of itself, you get the same effect with only one patch, which can save a lot of time and calculation work while rendering. Sometimes the effect is even better. The cheek of the baby is only an example compared to the polygons. The Human face has areas of high detail and areas of low detail. Eyes, mouth and ears may still need many patches to achieve realistic results [Figure 2-9]. In figure 2-9, a very complex ear model is shown. The black curves drawn on the ear are not the real curves that can be used to create this model. However, the real curves may be in fact a large number due to the irregular grey surface. The basic idea of modeling this ear is to rotate a surface for 360 degrees horizontally and let the curves control the height of the surface. So the surface must follow the shape of the curves. And to get complex ear configuration, it will need many curves of different shapes to complete that. Again, this process may involve even more amounts of patches using polygons. Although NURBS is a good representing method, animators should still be careful about the distributions and geometric continuity of the patches and make sure the boundaries between patches are invisible.

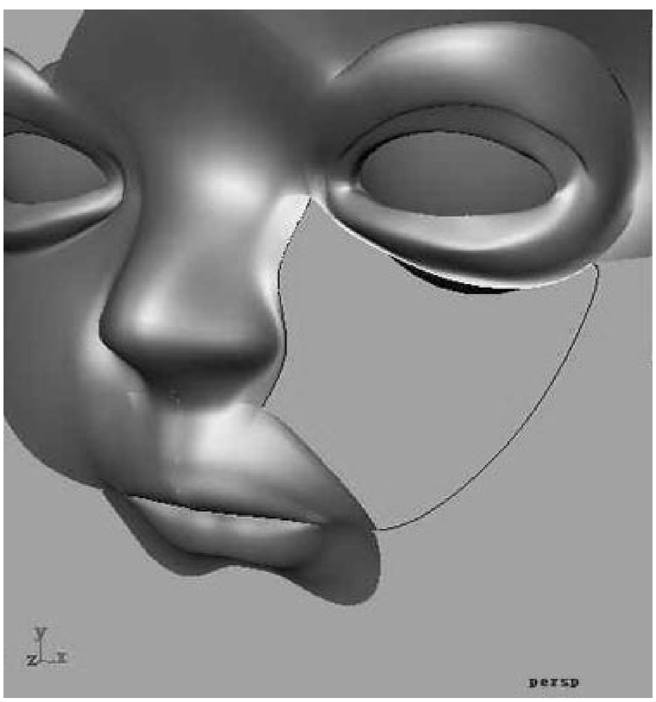


Figure 2-8 A baby's face modeled with NURBS patches. The blue curves compose one patch that covers the cheek of the face.

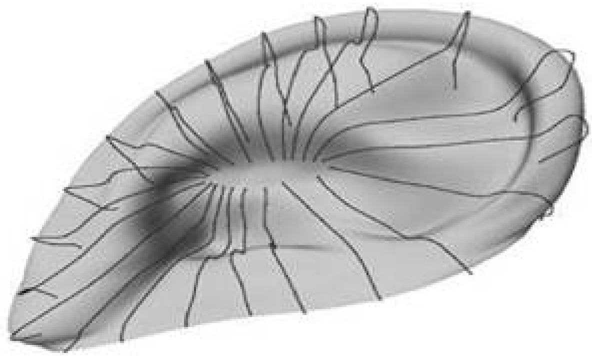


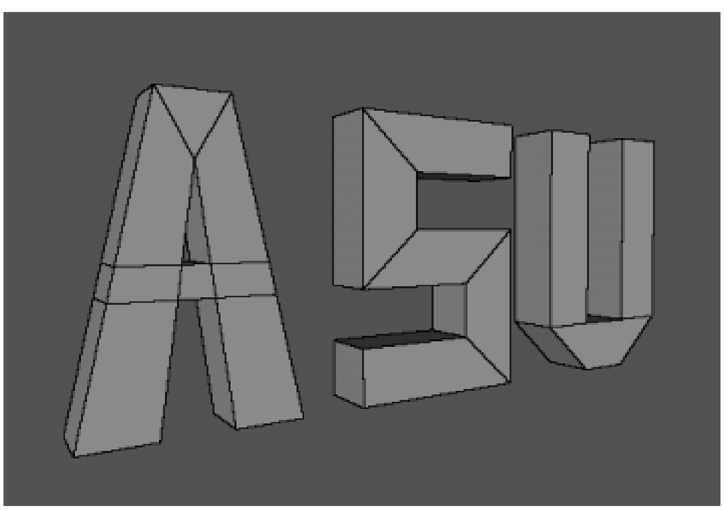
Figure 2-9 An ear modeled with NURBS patches. The black curves control the shape of this one surface that simulates the ear.

2.7.1.3 Sub-Division Surfaces

Subdivision surfaces are a surface refinement scheme describing a surface using a polygonal model. It is an iterative process of generating polygon meshes linearly to represent smooth surfaces. The polygons it used can be any shape or size and are not limited to rectangular patches. It approximates the original surface by subdividing each polygonal surface into smaller faces. To be put another way, one could say that by adding vertices on the original polygon, one can generate an even smoother approximate surface. The vertices are added to nearby original vertices. In some situations, the old vertices may also be altered to relatively approximate the same shape. The original polygon can be very coarse with only a few polygonal meshes. And the subdivision surfaces will finish the smooth work. Due to the fact that it is a linear process, the calculation speed is still not that slow although many more meshes are involved. And the complex smooth surfaces can also be generated in a predictable way from relatively simple meshes.

Subdivision surfaces was first published by Catmull and Clark and Doo and Sabin in 1978, and was about recursively generated b-spline surfaces on arbitrary topological meshes. There are different schemes for the actual subdivision process. The Doo-Sabin scheme and the Catmull-Clark scheme are the most early and widely used schemes [Doo 1978, Catmull 1978]. They were adopted on many commercial animation packages such as Maya, 3Ds Max, Lightwave and many game development engines. They are the tone that other schemes followed. Later on more schemes were developed such as Loop [Loop 1987], Butterfly [Dyn 1990] and Modified Butterfly [Zorin 1996], and Kobbelt [Kobbelt 2000]. The most substantial difference between these schemes is whether or not they retain the original vertices. For example, there are two basic ways, either the original vertices are not retained at the newer level of subdivision when approximating the same shape, or the subdivision happens around the original vertices and makes sure the original vertices are always unchanged when other vertices are interpolating around them. However, higher levels of subdivision surfaces add a lot more computational work. There have been many adaptive methods based on basic schemes that reduce meshes where it needs them the least. Ashish Amresh and Gerald Farin developed an adaptive method based on the Loop scheme, which reduces a lot of meshes compared to the normal scheme [Amresh 2003].

Figure 2-10 is a popular example for explaining the difference of the schemes. The first image is the base mesh of these three letters which has only sixty nine polygons. There are six images each illustrating a result from two different subdivision schemes. Each row represents one level of subdivision process and all the images on the left are processed using the Doo-Sabin scheme and all the images on the right are processed using the Catmull-Calrk scheme [Figure 2-10] [Homles 2009]. Three times the subdivision process has been applied on the letters and each time the letters have become smoother. However, the shapes of their letters have turned out to be slightly different in the two schemes. After doing the subdivision process three times, the left image now contains 4,452 polygons and the right image contains 4,448 polygons. The tips of the letter U on the left image are rounder than the letter U on the right image. Also the shape of the letter S is slightly different in these two pictures. This small difference is caused by different subdivision schemes. There are also schemes that work with triangular meshes only such as the Loop [Loop 1987] [Figure 2-11]. All the schemes and adaptive schemes develop subdivision surfaces into a very important method and are applied as a component in the modeling modes by most animation systems.



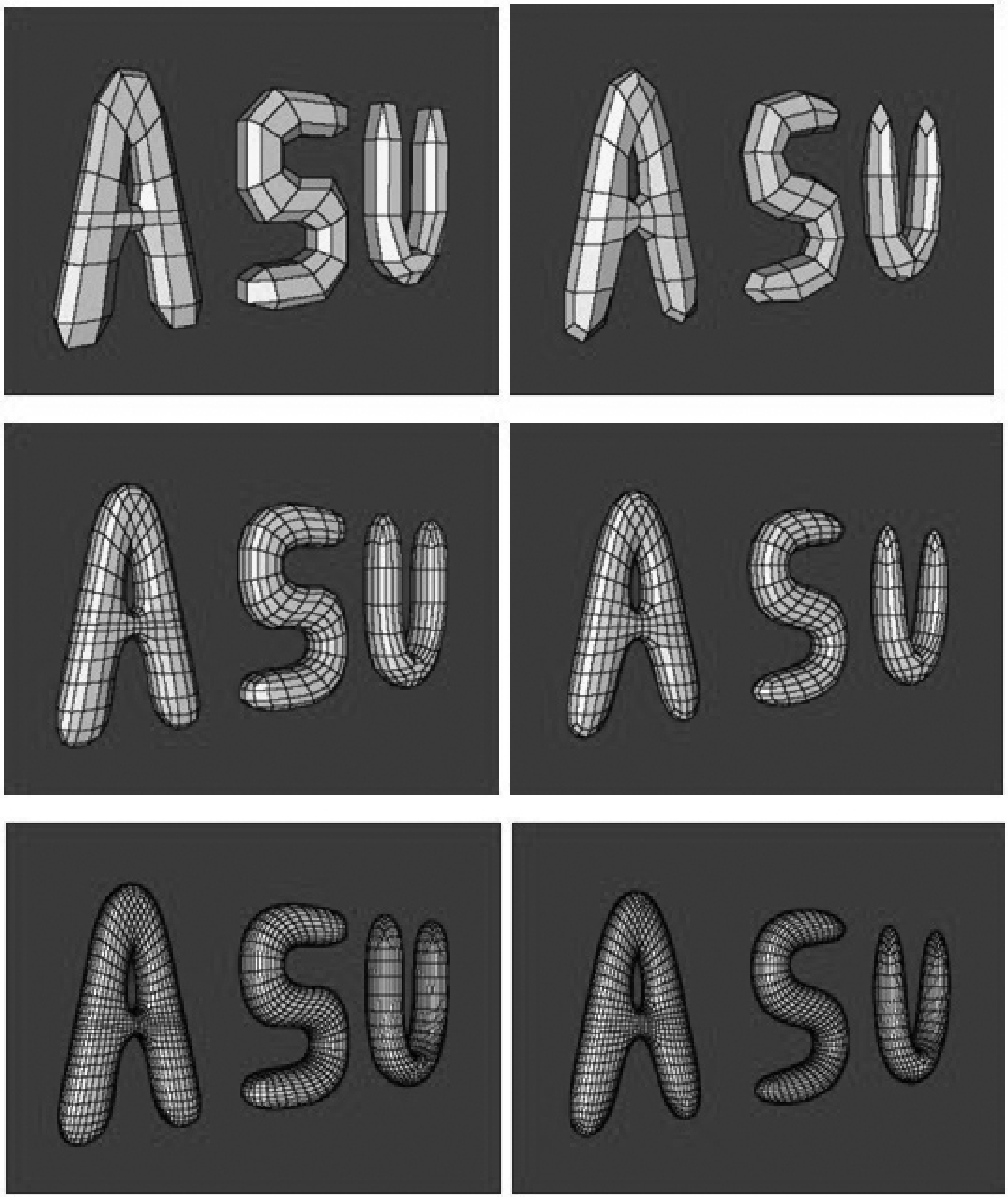


Figure 2-10 An example of the difference between Catmull-Clark scheme and Doo-Sabin scheme [Holmes 2009]

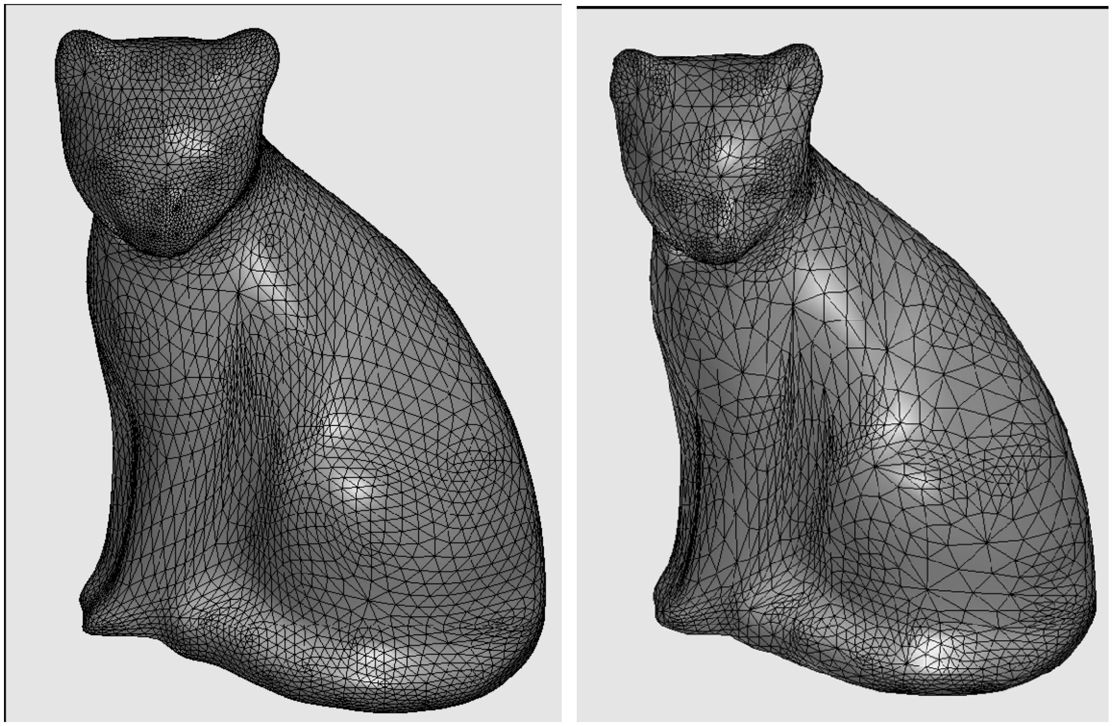


Figure 2-11 This is a cat represented by using the Loop scheme, the image on the left is using normal subdivision, the one on the right is using an adaptive method which reduces a lot of meshes.

2.7.2 Other Modeling Techniques

2.7.2.1 Three-dimensional digitizers

Three-dimensional digitizers are special hardware devices that rely on mechanical, electromagnetic, or acoustic measurements to locate positions in space [Park 1996, 2008]. These digitizers rely on different measurements of an object's three-dimensional positioning information of all the surface points when locating the object. Therefore, a large number of points are measured which requires a considerable time and computational load.

Mechanical digitizers measure the physical positioning information of an object at the same time the surface points are measured. The system can convert the mechanical position into electrical signals so that they can analyze these signals to determine the position of all the surface points of the object. Acoustic systems rely on sound sensors. Usually several sound sensors are used to measure one object to receive the sound reflections from different angles. Electromagnetic digitizers rely on generating electromagnetic fields and convert the signals that the fields provide into position information. The most widely used electromagnetically based digitizer is Polhemus 3Space [Polhemus 1987].

During the measuring process, these digitizers work better with sculptures or models that do not change shape or move positions. Rather than real faces, stable models are easier for measurement because if the position information is changing, it would be impossible to gather the precise positioning information of each point. As long as the object is stable, three dimensional digitizers can measure either polygonal vertices or control points over a NURBS surface which makes it rather applicable on current models. The character *Gollum* from "The Lord of the Rings" is an example of digitizing technology. First they make a sculpture of Gollum and then use the digitized data to define and control a subdivision surface model.

2.7.2.1 Photogrammetry

Photogrammetry, also known as image-based modeling, involves the capture of facial surface shapes and position data photographically. It is the first remote sensing technology in which the computer determines the parameters of a 3D model from photographic images to reconstruct objects. Several cameras can be used to capture one model from different views in order to have a desired three dimensional positioning data in space from using two dimensional photos.

First, we interpret the three dimensional coordinates and then locate the object's points in it. This method is applicable on both real faces and sculptures. Since it is still images, it does not matter if the object is still or not. But to measure the surface on the face using multiple pictures from different angles, an arbitrary marking method needs to be used. Usually the real face or the face sculpture has marks on it, and then by taking photos, the users can have an accurate measurement using many photos according to the corresponding marks.

Although it is applicable on both real faces and sculptures, it is still challenging work to get precise measurements. Besides, there is also other work such as lighting and camera settings to photograph which is another factor affecting the precision. After the photos are taken, there are also many calculations and analysis of the work so as to fit the three dimensional model in the computer. To determine precise coordinates and measurements, advanced technology of geometry, mathematics and physics may be used.

2.7.2.3 Laser scanning systems

A three dimensional scanner is a device that collects data from a real face by using laser scanning techniques. These 3D scanners can collect distance and positioning information and possibly color and texture information on the surfaces of the object. The laser rotates in a cylindrical coordinated system and transforms the scanning data into Cartesian coordinates to apply on the three dimensional models [Cyberware Laboratory Inc. 1990]. It is extensively used by the film and entertainment industry and video game companies. By using the scanner to collect the geometric information on the surfaces of the object, animators can reconstruct the shape of the three-dimensional face models and if color and texture information were also scanned, they can apply this information on the model as well.

The scanning process is very fast and simple. However, it usually takes multiple times, and multiple angles, to scan a complete model. Most of the time, a single scan cannot produce enough points' data with enough perspective views. What is important is that we have to obtain enough scanning data from different views to reconstruct the precise surface information for the models. For most situations, a single scan will not produce a complete model of the subject. Although the scanning step is simple and fast, the post-processing work is relatively large. The scanning data needs to be thinned out and spatial filtered. And these scanning devices usually produce a large number of meshes. To analyze efficient data from them is very time-consuming. Because the scanning device usually rotates in a cylindrical coordinated system, the problem with a laser-based device is missing data. It usually occurs on the top of the head, underneath the chin, in the pupils of the eyes, or in areas with hair. The scanned data becomes increasingly sparse especially near the top of the head, and the surface is poorly resolved in this region. Therefore, it also needs filling data in these areas during the post-processing step, known as the polar artefact. Otherwise, the data in this area is more prone to error. One common resolution is to ignore the missing data problem on the top of the head and substitute a newly constructed mesh that does not have the same problem. As in three dimension spaces, the control points and meshes should have the same normalities on one object or there will be display errors. However, the data from the laser scanners tend to be noisy and normalities might be different. Therefore in the post-processing step; the normalities of scanning data need to be filtered. Different scanning devices may have its limitations too, such as optical scanners encountering problems with shiny and transparent objects as the light reflecting on these textures is too difficult to handle. To solve these problems, a thin layer of white power may be covered on the surface to help with more reflections but still, it is not a desired solution.

2.8 Performance-driven Techniques

Performance-driven approaches for human motion synthesis have been the subject of increasing attention in the computer animation industry [Parke 1996, 2008]. Clearly, facial animation is one of the biggest challenges in computer graphics because the face has so many dynamic properties, and subtle emotions that cannot be adjusted by moving slide bars or dragging control points. So people come up with the idea that maybe the best technique to create facial animation is the face itself. We could capture movements directly from real faces and transfer that information digitally. It was first developed by Lance Williams, which involved using a few markers on the face and a video camera to scan the face and track the markers [Williams 1990]. Performance-based animation often uses interactive input devices such as Waldos [deGraf 1989], data gloves, instrumented body suits, and laser-or video-based motion-tracking systems [Parke 1996, 2008]. Generally the system contains of a video camera which points to the performer's face and moves with the head. The camera takes the position on the set of markers or dots drawn on the face and tracks the movements of real performers and drives the three-dimensional models accordingly. Currently, all face-tracking devices are still optical devices or video-based technologies. However, current systems such as Vicon can track more than 100 markers which lead to more precise and realistic facial animations. Because the capture process is unstable, it makes consistent and accurate tracking through the whole process very important when getting precise feature points and edges on the face, while maintaining a high-quality animation. The data generated by the camera can be used to create facial animation directly or to infer AUs of the FACS technique.

As for most popular facial animation techniques, there have been hundreds of published papers on methods to track face appearances. There are basically three types of performance-driven technologies: optical flow tracking, motion capture (with or without markers), and motion capture with keyframe animation. The main difficulty of performance-driven techniques is the data accuracy. The data captured by these systems need to be filtered prior to animation so that animators can use the data-driven animations and keying errors. Motion capture is a very promising technique and has been used by a long list of programs, both commercial and academic, such as Face Robot, Motion Builder, Zign track, Facefx, Image Metrics, Crazy talk etc. It is the most powerful facial animation technique so far and will continue to be one of the most popular techniques because of its ease of use; it's saving of tremendous effort and time, and its production of relatively high quality facial animations. The

Motion capture technique has been developed a lot in the past few decades. It dates from the 1980s or earlier, however, it has been applied in the movie industry only in the last several years. And the movies that have adopted motion capture techniques also made a lot of breakthrough success. For instance, in 2004, *The Polar Express* used markerless motion capture techniques to track the facial movements of all characters in the movie [Figure2-12]. Hundreds of motion points were captured to recreate the realism. *The Matrix* also used performance-driven avatars in a number of special effects shots. In the Matrix, they used a markerless dense motion capture system composed of five HD cameras and applied stereo to reconstruct the geometry of the models. Markerless motion capture is also being developed and used by researchers world-widely [Furukawa 2009] [Kaimakis 2004] [Hasler 2009] [Rosenhahn 2007, 2008] [Sundaresan 2005].

Figure 2-12 The conductor in *The Polar Express*, played by Tom Hanks. The avatar is driven by the performance of Tom Hanks. The Sony pictures movie *Polar Express* motion capture team leading by Ken Ralston and Jerome Cheng used a motion capture system called ImageMotion to simultaneously record facial and body motions of many actors at once. This new capture system used a lot of capturing cameras to cover 360 degree of actors.

2.9 Speech Synthesis

Speech synthesis is regarded as the visual motions on the face when people are talking, especially the mouth part. As the powerful hardware for 3D animation becomes more supportive, recent facial animation has developed from being concerning mostly about head and muscle movements to higher requirements such as lip synchronization, such as the synthesis of lip behaviour and the novel text. Because lip movements can be very fast and sophisticated and the connection between text and speech is relatively vague and hard to grasp, hence visual speech synthesis has been the most difficult task in the past few decades.

Making speech animation is different from making general facial expressions and muscle movements. The traditional way of making expressions is keyframe-based and that usually cannot provide satisfactory speech dynamics. The difference between facial expression and speech animation is that people tend to be lazy when they are talking. They always obviate some unnecessary syllables and sometimes even pronounce without movements. Therefore, we cannot make animation exactly the same as all the syllables. If you key every movement according to the text and syllable, the result will seems quite “busy”. So it’s important to be “lazy” while doing facial animations.

To simplify the lip movements, one usually classifies the lip shapes into must-see shapes and not important shapes. There are certain sounds that we make with our mouths that absolutely need to be represented visually, no matter what: visimes. [Osipa 2007]These are the sounds that must be made with specific mouth shapes so the visimes must be modeled and animated. The most important thing in the traditional way of making lip-sync animation is to study these visemes and key them manually, and to be aware of what they represent such as key poses or of the variation of these visemes when in different text circumstances. Key poses can be complex, it includes the position of the lips, jaw and tongue positions, and these properties are always changing in different text circumstances. You cannot memorize a certain shape of the same letter and apply that directly in the next sentence, because the connection of different letters actually changes the shapes and movements of the mouth. That is why visimes are the most concerned element that animators care about when there are no advanced lip synchronization techniques. And it is actually why speech animation is so challenging, and is often referred to as the coarticulation effect: the shape of the mouth that corresponds to a phoneme, but also on phonemes that occur before or after the current phoneme. It is the influence of the surrounding visemes on the current viseme. In linguistics literature, speech co-articulation is defined a little different: “phonemes are not pronounced as an independent sequence of sounds, but rather that the sound of a particular phoneme is affected by adjacent phonemes.” Taking coarticulation into account, current systems use longer units such as diphones, triphones, syllables or even words and sentence-length units to blend viseme keyframes. Normally systems provide phones and diphones to output broadly, but it may not be very explicit. For some situations, the production from large storage such as words or sentences actually has a high quality.

The most important qualities of a speech synthesis system are naturalness and Intelligibility. Naturalness describes how closely the output sounds like human speech, while intelligibility is the ease with which the output is understood. The quality of a speech synthesizer is judged by its similarity to the human voice, and by its ability to be understood. The ideal speech synthesizer is both natural and intelligible. Speech synthesis systems usually try to maximize both characteristics. There are a lot of computer-aided systems used for synthesizing the visual motions to speech animations [Lewis 1987]. The very first computer-aided speech synthesis system was invented in the early 1950s. Although the early quality of the synthesized sound was very robotic and not even close to realistic, as the development of the speech synthesis systems improved, the sound quality also improved and has reached a realistic level. Sometimes it is hard to distinguish it from actual human speech.

As speech animation becomes popular, a lot of research effort is being generated to improve this domain. There have been mostly two types of approaches: visieme-driven approaches and data-driven approaches. The first one is more of a traditional approach, involving the animators to key the shape of the mouth according to the visemes; every keyframe and the frames in between are automatically interpolated. The second type is more sophisticated, it involves a video-driven approach, a text-driven approach, or sometimes a muscle-driven approach. Video based techniques only require a pre-recorded facial motion database for the synthesis process afterwards [Figure 2-13]. Text-driven techniques are designed based on the typing input. Of course, each technology has strengths and weaknesses, animators usually make decisions upon the task requirements.

2.9.1 Text-driven Synchronization

The text driven approach can convert language text to speech. People have invented a language in the processing of synthesis called text-to-speech (TTS) [Klatt 1987]. Many systems are designed based on this principle. These systems are benefiting people with visual impairments and reading disabilities by using words on computers. The very first text-to speech system was developed as early as 1968 with the MITalk being the best known such system [Allen 1987]. It functioned as an engine which could translate the text from the front end into the speech information that the computer needed, to the back end. Within the translation process, there are three major steps. The first step is to analyze the text from the input which converts raw text with numbers or symbols into its equivalent and normalized words. Then we need to assign phonetic transcriptions to the words, named as the text-to-phoneme step. It involves dividing the text into some units like phrases, clauses or sentences and then assigning the phonetic information so that we can render out phonemes from text. This step is called linguistic analysis. As long as we have the phonemes, we can synthesize them to the text and convert the symbolic linguistic representation into sound which is speech. This step is called synthesizer [Figure 2-14]. There are also other ways to translate speech to visual movements such as rendering symbolic linguistic representations. One of the examples is to render phonetic transcriptions into speech.

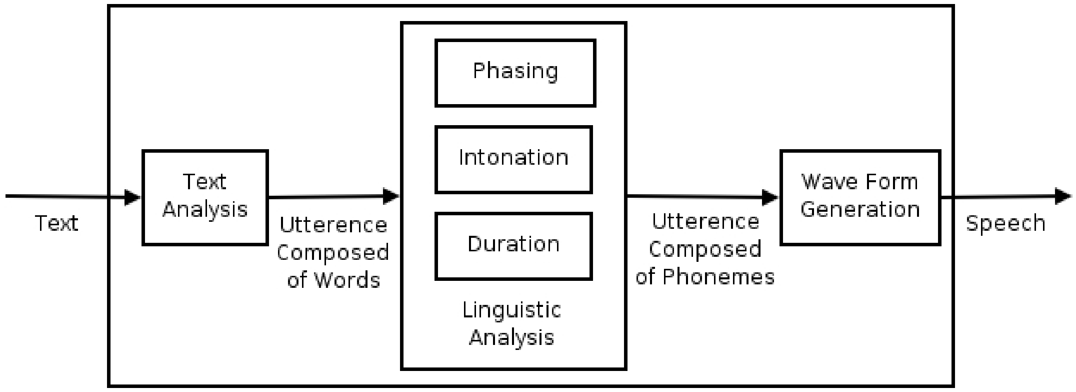


Figure 2-13 The image shows the process from input text to output speech. It contains three major steps, ext analysis, linguistic analysis and synthesizer.

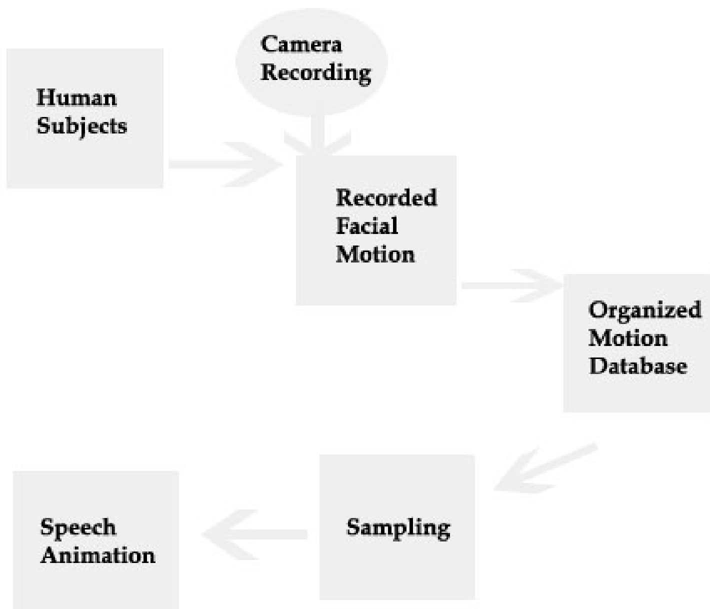


Figure 2-14 The image shows the process of data-driven speech animation which uses a pre-recorded motion database.

2.10 Transfer Existing Animations to Different Models

This approach mainly refers to transferring existing facial animations from a source model to a new target. Because the animation process is tedious and time-consuming, people invented the idea of using existing animation fractions on a different model, named expression cloning. The animations created by any tools or methods can be retargeted to new models. As a result, animators are free to pick the animations from a large, high quality, facial animation, motions database. The animations can be reused and composed to new models and it can make different expressions. Although after transferring there is a model-specific animation tuning step for the new model, it still can save a great deal of labour costs. Under this definition, performance-driven animation and data-driven animation can be conceptually regarded as a specific way of animation transferring. Although in this case, the source that is being used is the 2D video footage. Either 2D videos or existing 3D animated face meshes or 3D captured facial motion data can all be considered as the source. The target usually is a 3D face model.

Noh and Neumann proposed an expression cloning technique to make facial animations in 2002 [Noh 2001]. It transferred facial animations to a new model by using the motion vectors of the vertex from the source model. This approach constructs a vertex motion map from each vertex's motion vector to a target model's vertex even though the model may have different mesh geometry and vertex configuration. The mapping idea relied on the radial basis functions. Since then, a lot of approaches have been invented to help build the connections between the source model and the target model so that expression cloning can be utilized easily. For example, Sung Yong Shin proposed a blend-shaped model, containing 20 feature points and an algorithm to compute the displacement vector of the base model for the purpose of the blending and tuning process. It can automatically extract the feature points from the source model and blend facial expressions at the granule regions to the target model [Sung 2003]. However, this approach requires a proper blend-shape model to perform the action and it is not applicable on a pre-designed blend-shape model that does not have the right configuration and setting.

Igor Pandzic proposed another facial motion cloning method shortly after Noh in 2003 [Pandzic 2003]. This method not only transferred the animations to a new model, at the same time it preserved the compliance of the motion to the MPEG-4 Face and Body Animation (FBA) standard. The scale of the movement is normalized with respect to MPEG-4 normalization units (FAPUs), and together with the MPEG-4 facial animation tables, it is more flexible than Noh's method and more suitable for free facial expressions as well as for MPEG-4 compatible facial motions. Facial motions from video have also been transferred to other 2D or 3D models by researchers. Vlasic proposed a platform that uses multi-linear models to transfer scanned 3D face meshes to target models [Vlasic 2005]. The transferring process relies on the attribute parameters and the connection of these parameters such as identity and expression so that in the new model we can adjust the settings accordingly. The process is very intuitive and dynamic.

In summary, facial animation transferring provide an alternative method that minimizes the computational and manual costs while maintaining the accurate dynamics and visual quality of the facial motions produced by any available methods. However, this technique is very ad-hoc and rarely used on commercial systems. The lack of available platforms and facial motions limit the development of expression cloning.

2.11 Eye Emotion Synthesis

Facial animation is not always about lips and muscles. As a part of the face, the eyes are crucial and subtle. Eye emotion is the strongest cue when understanding the mental state of human beings. When people are talking, we look at their eyes to judge their emotions, interest and attentiveness and get all the information we need to respond. As reflected in the saying “windows to the soul”, eye gazes send strong and subtle signals. That makes eye animation the most important yet the most difficult factor in creating realistic talking agents. As long as the eyes are engaging, audiences actually do not pay much attention to the lip synchronization. What is ironic is that the entire animation and game industry is obsessed with lip synchronization rather than eye animation.

In character animation making, eye animation is either non-existent or very poor. However, eye animation is crucial when animating a talking head. One of the reasons why eye animation is not done properly is because large eye movements are not what makes the eyes. Eye gaze may be partly due to the eyelid motions and deformations, but the connection is not strictly deterministic. It is actually all the subtle little flicks inside that make the eyes. Hence, using direct motion capture data cannot get convincing and useful eye animation. In addition, currently a lot of methods are still stuck with motion capture because it seems like the only way to track pupil and iris movement. Even if we use motion capture to track the eye movement, a method to use the data in a pipeline of animation productions is another unsolved issue. Besides, keyframing eye animation is prohibitively expensive. If we use the captured data, we cannot use it directly but have to keyframe the movements afterwards. Also, "there are technical hurdles and performance-related issues that prevent you from using the data properly, even if you could get it" [Perrett 2008]. Another reason why eye animation is so complex is because it contains too many types of behaviours such as clear eye contact, Gaze direction, Eye convergence (focus of attention), Eyelid curvature and compression (squinting), and Eyebrow behaviour. All these behaviours cannot be done simply through video-based tracking or any other single computational method.

A number of research studies have been developed to model realistic eye gaze motions and head movements. For instance, Chopra-Khullar [Khullar 1999] presents a framework for computing gestures like eye gazes and head motions of three-dimensional avatars in dynamic environments based on high-level scripts provided by users. Vertegaal [Vertegaal 2000, 2001] designed an experiment of user studies to validate the role of eye gaze direction cues to understand whether it is the reason people can recognize who is talking to whom in a multi-party conversion environment. Lee [Lee S.P 2002] proposes statistical models named "Eyes alive" for analyzing eye gaze details. In their approach, they demonstrate that synthesized eye gazes are necessary for achieving noticeable realism and conveying appropriate mental states. However, only first-order statistics with a number of empirical parameters are used and not all eye behaviours are considered in their work, such as eye convergence and eyelid curvature and compression. Text has been one of the best methods applied further to synthesize eye animation. Deng [Deng 2003, 2005] proposed a text-based method that can simultaneously synthesize eye gaze and blink motion, accounting for any possible correlations between the two. And they also demonstrated that a text-based method is effective and applicable in synthesizing eye movements as well as the correlations between blinks and gazes. Saccades are also important in eye performance and the rapid movements are usually accompanied with a head rotation which can convey various states with various rapidity and timing. It also influences the mental states of human beings and the complexity of making eye animation. Freeman [Freeman 2000] claimed we can predict saccade kinematics if head movements are taken into account as the eye-head interaction can be helpful in understanding saccades. But there is still some vacant holes left as to how in particular the saccade and gaze both change the eye behaviour. So far we assume that eye gazing and blinking can cause the shape deformation of the eye lids. The correlation should be considered when animating 3D talking heads, especially the direction of eye gazes which cause subtle shape changes that are not easy to grasp. As the importance of eyelid shapes in eye animation synthesis becoming recognized, studies have developed some data-based methods for constructing the geometric deformations of the avatar's eyelids. Elisei and Bailly [Elisei 2007] presents a data-based construction method which constructs an articulatory model of the shape deformation for the eyelids of a 3D face that takes into account the eye gaze direction. And the model they present can be used to analyze cognitive activities. This model is driven by three low level parameters, two for eye gaze direction and one for the opening/closing of the eyelids.

An advantage to use texture in generating eye animation is that the eyes are small parts of the face. Considering the physical size of the eyes, it is easy to arrange the texture without looking roughly unrealistic. Another benefit is that it does not need huge computation. The motion-capture based method is the most straight and quick way. However, the captured data usually does not present a perfect psychological performance of human beings. There are also statistical models which must take order into account. Current eye animation synthesis methods clearly cannot produce full, realistic eye movements with various phenomena. Because the movements of the muscles around the eyes are not strictly linear, we cannot predict the precise deformation of the eyelid shape changes using simple algorithms. In future work, we hope that more work can be verified to create eye animation with different mental states, which can also generate according head motions. In summary, eye animation is still subjective and experimental but it will become more and more important in the future character animation area.

2.12 Head Motion Synthesis

Natural head motions play a crucial role in making lifelike conversational agents and an engaging human-computer interface. It is also facial gestures that convey important and distinctive nonverbal information in human communication. A simple example is seen in some basic virtual environment head motions and can be the easiest way to present meaningful performance. For example, we can use a basic nodding and rotation action of a 2D smiley face to represent right or wrong in an online question game. The users will understand even without any expression change on the smiley face. Hence heads should be appropriately modeled and be involved in making realistic facial animations besides just the lips, eyes and muscles to effectively mimic facial behaviours. Creating head movements seems rather straight forward, i.e. either nodding or rotating. However, when animators are making head movements they usually take the entire facial expression into account. Researchers who are developing these kinds of techniques also have to understand perfectly the correlation between head movements and facial expressions and how head movements can be helpful for the audience to grasp the avatars' mental states. A facial gesture is typically considered as a gesture containing facial muscles and facial movements, eye animation, and long head movements, etc. A lot of research has been done to perform this task.

Early head motion techniques could only generate simple actions such as nodding. But more and more researchers are creating head motions combined with other facial factors. Busso found that head motion patterns with neutral speech significantly differ from head motion patterns with emotional speech, in motion activation, range and velocity [Busso 2007]. They also presented a HMM framework that provides the best description of the dynamics of head motions and generates natural head motions directly from acoustic prosodic features. Figure shows the overview of this HMM-based head motion synthesis framework [Busso 2007]. Hiroshi developed a mechanism for controlling the head movement of photo-realistic lifelike agents when the agents are in various modes; idling, listening, speaking and singing [Lewis 2005]. It imitates the manner of a head motion of an existing person to give the agents virtual personalities. Graf analyze quantitatively head and facial movements that accompany speech and investigate how they relate to the text's prosodic structure [Graf 2002]. Based on their statistical study, head movements vary depending on various verbal conditions. Hence they estimate the distribution of head movements in different pitch accents and spoken text. Deng [Deng 2004] presented a data-driven audio-based head motion synthesis method that first captured a real human face speaking then extracted audio features from it. The data was analyzed through a K-Nearest Neighbours (KNN) algorithm to synthesize head motions from audio input data.

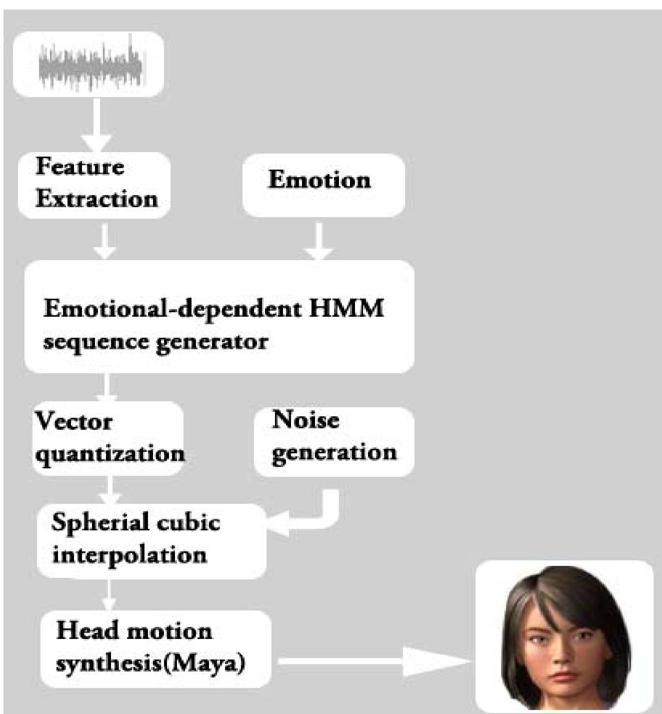


Figure 2-15 A HMM-based expressive head motion synthesis framework [Busso 2005, 2007].

3: CURRENT FACIAL ANIMATION SYSTEMS

Facial animation has been the focal issue of generating 3D avatars in computer graphics for a few decades. The way animators used to create facial expressions for agents varied from earlier blend shapes and shape interpolation to later motion capture technology. These techniques burst into many 3D systems that supported a number of facial animation approaches. Among these systems, there are large and comprehensive packages such as MAYA or 3Ds Max, and also stand alone systems that are specifically designed for creating facial animation such as Face Robot, Facefx, and Dio-matic Facial Studio; not to mention, hundreds of academic open source applications developed by researchers using various facial animation technologies. This chapter attempts to summarize generally how animators are doing facial animations nowadays and describe the theory of these systems in order to have a better understanding of current facial animation systems so that we can know what is working well and what is not. Also, we are hoping both industry and academia can gain more information regarding the gap between commercial systems and academic systems.

3.1 Large 3D Packages

The most popular large 3D animation systems usually have very complex functions combined together. It is a “whole” pipeline from modeling, texture, lighting and animation to rendering. They provide a very comprehensive three-dimensional working environment for computer graphic users. You can use them to create humans, animals, buildings or any other objects existing in the real world and you can animate them. Due to fact that these packages are designed for general three-dimensional creatures, sometimes it restricts the freedom and capabilities when making facial animations. Because facial animation is one of the most complex and distinct domains in the virtual environment, large 3D packages cannot always perform satisfactory results and high quality animations. For example, there is a specific character body animation platform called CAT in 3Ds Max but there is no platform specific for doing facial animation. Hence, to generate facial motions animators have to use several traditional methods such as shape interpolation or parameterizations. These methods are very time-consuming and also are expensive in labour costs.

In the most currently used professional system Autodesk Maya, the animators prepare several head models copied from a neutral expression face and adjust the expressions on these heads to the largest degree. Each one represents different facial expressions and is named accordingly, such as a frown, open jaw, etc. Because these expression models have the same geometry as the neutral model, it is easy to link the affection between them directly. As long as the relationship is created, animators can adjust each expression on one head to control the corresponding expression on the neutral head. You now have full control over the character's expressions because you have the ability to flex any of the facial muscles by simply dragging a slider bar. Users can experiment with different settings and combinations of poses by manipulating blend shapes. They can either use the slider bars to blend from 0 to 1, or type in a value, or type a value bigger than 1 or smaller than 0 to get an exaggerated effect. However, if animators want to create more realistic expressions, the expression models have to be a lot more than a couple of still models. Although animators can have relatively good facial expressions using blend shapes, for lip synchronization, having this many blend shapes to keyframe could be very time consuming.

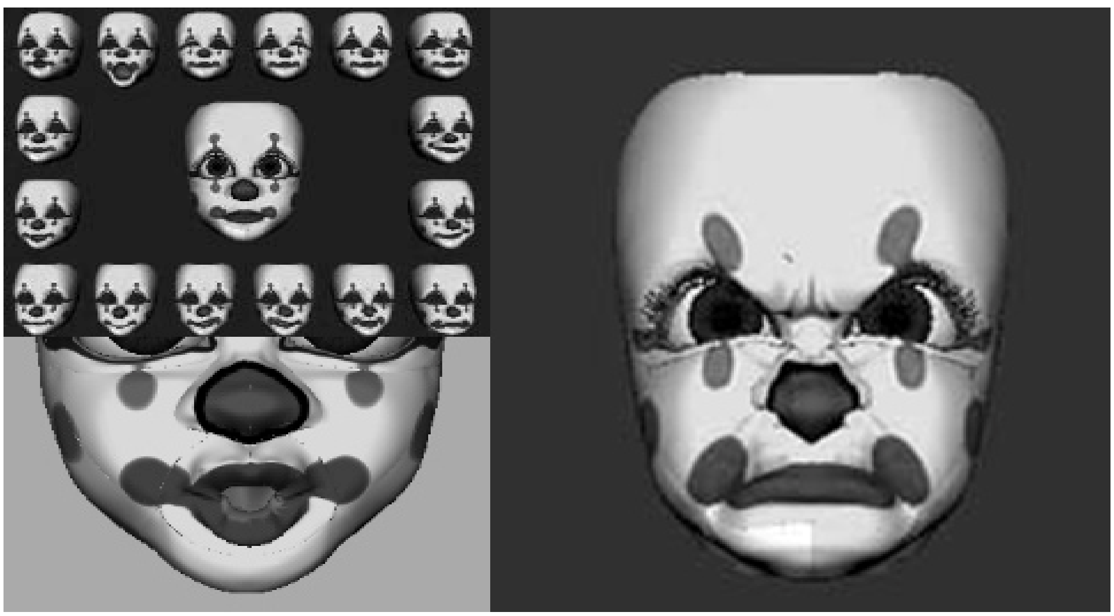


Figure 3-1 Top left shows the general steps of creating facial animation using blend shapes in MAYA. In the center of the screen you can see the base version of the head. Around the border are 16 different versions of the head. Each of the heads contains one facial muscle fully flexed. Right shows an angry expression generated by dragging the parameter bars of 16 different blend shapes. Bottom left shows a snapshot from a lip synchronization footage created using blend shapes.

The example in MAYA shown above illustrates that although we can produce facial animation in large 3D packages, the process takes a lot of effort. For instance, the number of blend shapes may rise to a hundred in some circumstances. Accompanied with the high requirement of realistic facial animation, general 3D software are not attractive to animators anymore. As soon as people found the importance of facial animation, they invented standalone systems to create facial animation specifically. Compared with the general 3D packages, standalone systems are easy to use with higher efficiency and also produce animations faster as well. Animators can import the animation data generated from these standalone systems into large 3D systems to finish adding light, texture, rendering steps, and export the data into the final movies. This pipeline makes it easier and faster to create facial animation, which also speeds up the entertainment industry overall so as to produce more and more digital character movies and games.

3.2 Commercial Standalone FA Systems (issues)

As mentioned in the above, more and more animators recently have chosen standalone facial animation systems to perform their tasks. Both the specialized techniques and features for facial animation are the major reasons why animators abandon large 3D packages. Current standalone facial animation systems are either commercial or research systems. There are standalone commercial systems for face modeling tasks such as Autodesk Mud Box and Facegen, or for facial animation tasks such as SoftImage Face Robot, ImageMetrics and Facefx. Currently, many commercial systems rely on performance-driven techniques and can provide different qualities of capturing facial animation data such as Proface, Zigntrack, Meta Motion, etc. These systems help animators to avoid the overwhelming setting and keyframing in traditional 3d packages and encourage researchers to develop more and more customized platforms and frameworks for facial animation. They also benefit both the animation producing studios as well as audiences to experience more realistic facial animations. However, current commercial systems are not quite mature yet, systems emerge and disappear either because of lack of fame, bad user interface design, or they do not perform better than existing comprehensive 3D systems, etc. The industry needs extensible systems that are customized for facial animation and that are also developed based on user's concerns. Among all the commercial standalone facial animation systems, Face Robot is the most alike system that may satisfy all customers.

Face Robot was released by Softimage and Blur studio during the Siggraph conference of 2005. It soon became an outstanding and powerful facial animation system. The emergence of Face Robot has been a milestone of stand-alone facial animation systems because it beats other systems on the performance of making realistic facial expressions as well as on the control of facial muscles. Several facial animation techniques have been integrated including the performance driven approach, while keyframe and parameterizations give users optional choices for making facial animation. Also, all the technology is built on a computer face model that mimics the real facial soft tissue of the human face which lets users gain a direct and intuitive control of facial expressions with a high resolution. With that help, motion capture animators can work with fewer markers thereby reducing setup and cleanup time. Facial expressions emerge primarily from deformations of the soft tissue on the face, which is nearly impossible to capture with other systems. Although Face Robot does not support face modeling, face models need to be imported from traditional systems or other face modeling systems into Face Robot. Conflict may happen during this process and the head models also need to be adjusted and tuned before being applied to create facial animation. Face Robot is also strong at tuning face models through some particular control boxes. These control boxes are distributed according to facial anatomy and can control a region of face muscles that moves as groups [Figure 3-1]. The small region and detailed performance of the controlling boxes are significant in the tuning process since realistic and believable muscle movements are basic factors in producing high quality facial expressions. The animation can be applied easily on the model after tuning and rigging either using motion capture data or keyframe animation or both. Face Robot infuses significant power into the animation field and much experience in how to make facial animation software. Although it has limitations and has not been a perfect system, it surely sets a good example which deserves attention and research.

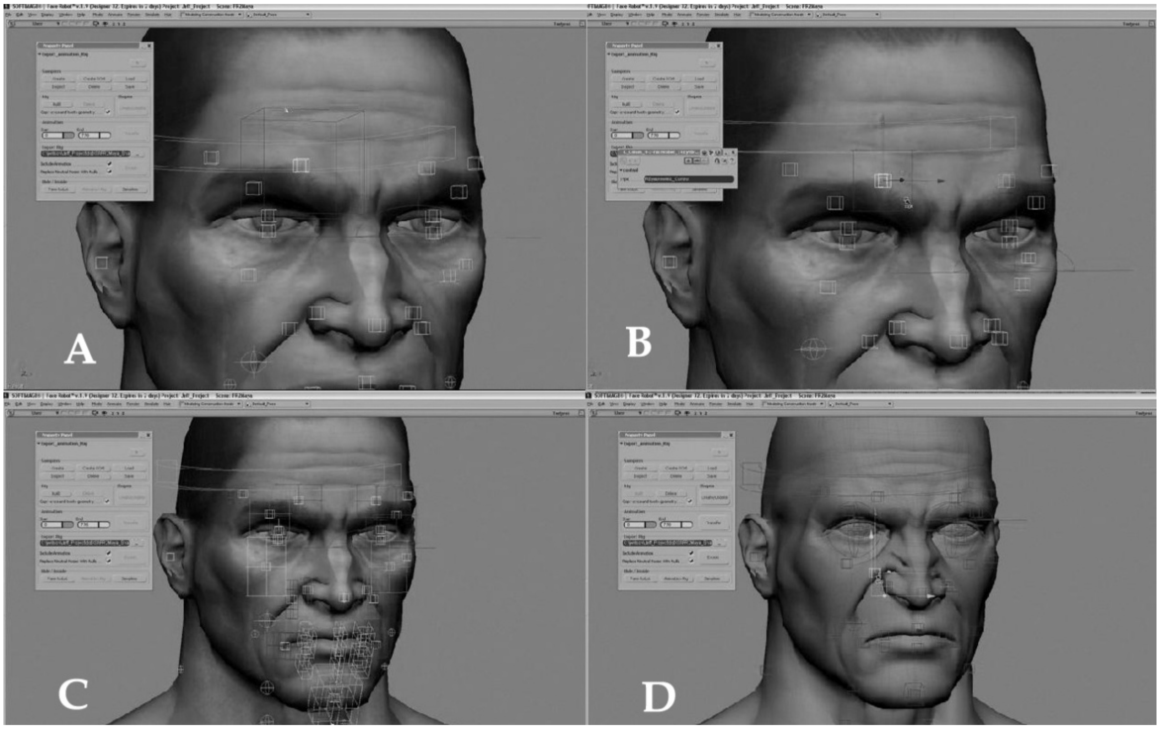


Figure 3-2 A facial rigging example in FaceRobot. Image A shows a selected control box affecting part of the eye brows. Image B shows the step of setting the region that the control box can affect. Image C shows all the control boxes have been set to some specific regions and ready to do animation. Image D shows the muscle deformation caused by one of the control boxes near the corner of the nose.

Figure 3-3 Face Robot motion capture

Accompanied with the fast development of computer hardware and the entertainment industry, not only some powerful and comprehensive systems like Face robot were invented, but also a large number of small and specialized systems. There are facial modeling systems for example Mudbox, Zbrush which are specialized for creating face models that are compatible with other animation systems. Separating modeling systems and animation systems can provide animators more flexibility and also may produce better results than doing them in the same system. These customized modeling system are designed for 3D face models particularly, hence they usually adopt the newest and fastest modeling techniques. For example, the newest version of facial studio uses a photo matching technique to model a three-dimensional head. It uses only two pictures, including the front and side, to simulate the face in Dio-matic facial studio. The system provides a matching interface that measures a head model using a number of key points. When you import the two pictures, you can match those key points to corresponding points on a real human face. Hence, users can easily and rapidly relate still pictures to the standard three dimensional models. Furthermore, it can import texture onto the model at the same time. After adjusting the settings, the system provides many parameter bars for animation. Users can select each parameter to change the appearance of every single region on the face although the resolution of the textures is not good enough to produce convincing results. Therefore, other modeling systems use different modeling methods such as Autodesk Mudbox and Zbrush. Both sculpting tools provide a more intuitive experience of high resolution digital sculpting as well as texture mapping with organic brush-based 3D modeling techniques and a highly intuitive user interface. The sculpting tool set contains an assortment of brushes with adjustable falloffs. Subdivision of models occurs using the Catmull-Clark subdivision algorithm in Mudbox [Catmull 1978]. 3D layers that can store different detail passes of current 3D model are used in Mudbox to prevent us from making permanent changes. Layers were originally designed by 2D Adobe software to organize work. By bringing them into 3D software, together with layer masks and multiplier sliders, typically is the start of a new way of designing 3D user interfaces. Consequently, the more ways to manipulate models the better and why not import some successful experiences from 2D software [Figure 3-3]. Hence, in Mudbox users can work with tools such as a lot of brushes, stamps, and stencils to quickly sculpt the geometry of models and create lifelike textures similar to their real world experiences. Similarly, Zbrush uses a proprietary "pixol" technology which stores lighting, color, material, and depth information for all objects on the screen. Besides, the sculpting brushes in Zbrush are even more advanced than Mudbox. These techniques are able to sculpt medium to high frequency details that were traditionally painted in bump maps using textures to simulate [Figure 3-4]. Both Mudbox and Zbrush are able to divide the polygon models into millions of meshes that are huge support to the development of next generation, high definition, digital modeling. They have been used by companies ranging from Electronic Arts to ILM and contributed to the making of a series of movies, such as “*King Kong”* and “*Lord of the Rings*”.

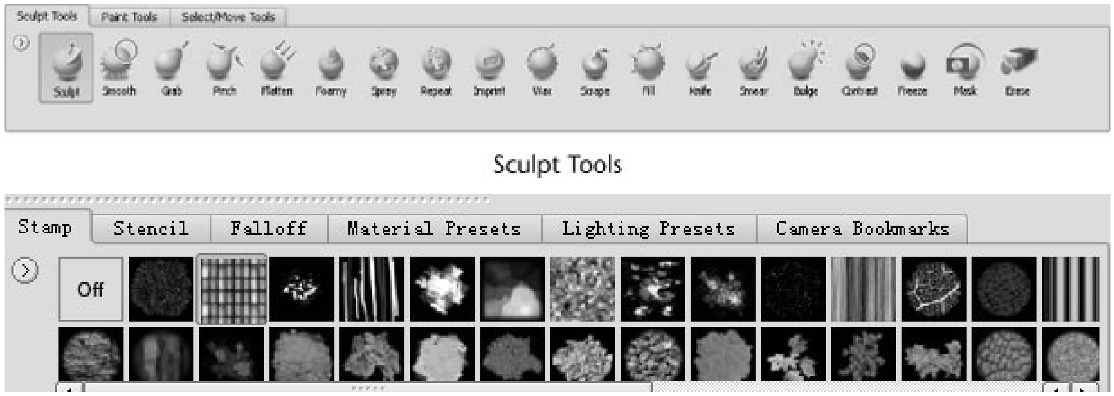


Figure 3-4 Organic sculpting brushes in Mudbox

Figure 3-5 Organic sculpting brushes in Zbrush. Brushes contain diffusion, pattern and depth information which support advanced digital sculpting and modeling.

There are also other small facial modeling systems that are used by companies and studios such as Facegen from Singular Inversions Inc., which use one or two photos to create realistic faces. The advantage of Facegen is that it takes only a few minutes to create a new head, yet it does not disappoint users in its facial details. Another commercial animation system, ImageMetrics' facial animation solutions are also based on performance-driven technology. What is unique is that there are no time-consuming markers or makeup to apply during the video capture process which definitely saves animators a large amount of effort. Usually these standalone facial animation systems all use more than one techniques combined together to provide a good effects (LifeStudio:Head; Diomatic facial studio; Lbrush; MotionBuilder; Proface; Zigntrack; Facefx; Eyematic/Neven Vision Meta Motion; Crazytalk; Spacetime; faces Magpie pro).

3.3 Research Systems (issues)

As the 3D talking head became more popular in the industry domain, it caused more academic studies to be written on the same subject in the research domain. There have been hundreds of academic papers published on various academic facial animation platforms, based on various facial animation techniques. For instance, there is Xface, which is based on the MPEG-4 standard, Efase, which is based on a data-driven approach, Pogany, which is based on FACS technology, Spacetime Faces-which is based on a performance-driven technique, etc. These systems usually are designed through one major facial animation technique because of individual ability limitations. Also, the lack of free and open sources limits further research into facial animation. Every researcher has to implement their own, new frameworks for each new system. And that is what determines why academic systems must have disadvantages and limits when compared to commercial systems. We need some fundamental and widely accepted standardized frameworks that exist as free and open sources and that can let researchers do development work easier. More and more people recognize this problem and have developed some open source platforms. For example, Facade uses a phonetic transcription of Magiepro and a lip-sync toolkit of Baldi. The biggest issue is whether or not it is well acknowledged and authentic. It is hard to take a big leap until something is claimed standard and extensible

If it is so hard, why do researchers keep exploring more and more systems? On one hand, the facial animation field does need better systems to perform tasks that are desired by current industries. On the other hand, researchers always believe their systems are better than others and thus keep designing systems that are a lot alike, but not extraordinary. One can often see similar functions and techniques among these tons of academic systems, but that does not help researchers jump to the next generation of facial animation systems. Therefore, in this chapter we will introduce some systems based on popular academic techniques and their common issues to see where they miss the mark and what should be explored.

3.3.1 eFASE

eFASE (expressive Facial Animation Synthesis and Editing system) is a comprehensive data-driven animation system that generates expressive facial animations by concatenating motion captured facial motion data, giving novel phoneme-aligned speech input and its emotion modifiers (specifications), while animators establish constraints and goals [Deng 2006b]. Because browsing and selecting among large numbers of motion databases is a challenging problem for data-driven based facial animation, a specific phoneme-Isomap based editing tool was invented to visualize the facial motion database in an intuitive way that helps users to remove contaminated sequences, to insert new ones as well as to reuse uncontaminated motion sequences efficiently. They also designed a dynamic algorithm to search the best-matched frames from a motion captured database to synthesize facial motion sequences based on given speech phoneme sequences, while the animator-specified constraints for phonemes use emotion modifiers to guide the search process. This algorithm includes a new position velocity cost for favouring smooth paths and can seamlessly synthesize animation based on an emotion mismatched penalty. Therefore animators do not have to create a separate motion database for each motion category when simplify the selecting process. Figure 3-5 illustrate the high-level components of the eFASE system.

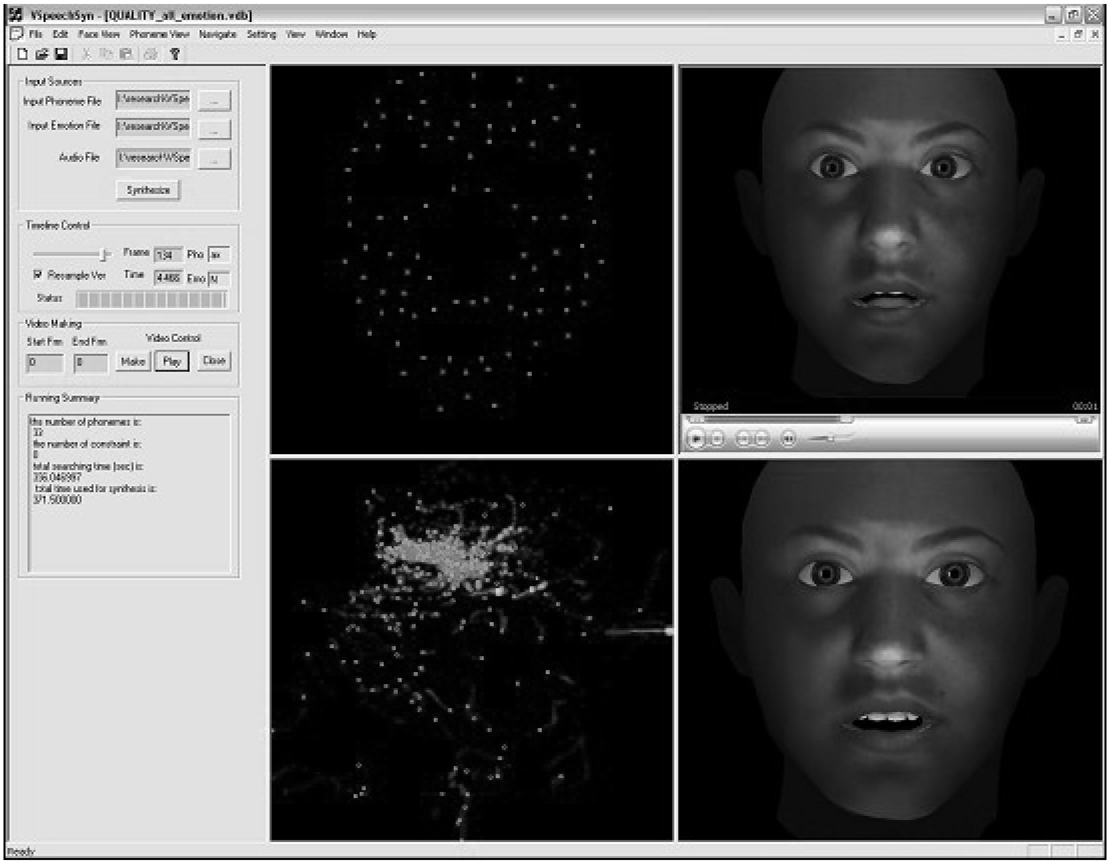


Figure 3-6 A snapshot of the running eFASE system

The distinction of eFase is that it provides motion-node constraints and emotion modifiers to intuitively control the motion captured database, which makes it easier for data selecting and synthesis. The novel phoneme-Isomap is another powerful editing tool for managing a large facial motion database. However, the limitation is that the quality of the motion synthesis requires a large motion database plus an accurate phoneme alignment. Where to find such a large database is the most common issue of all data-driven based approaches because building such a database requires lots of time and effort. Furthermore, the current eFASE system does not support real-time applications, which means it does not alleviate the difficulty of making facial animations, while also dropping behind common commercial animation systems means it is falling behind.

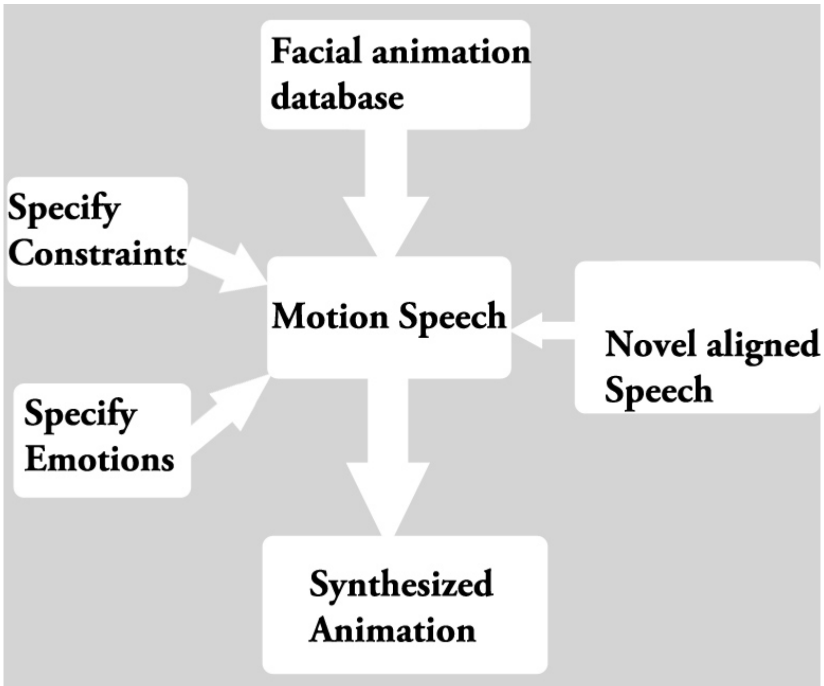


Figure 3-7 Illustration of the high-level components of the eFASE system. The pipeline starts from the top, given novel phoneme-aligned speech and specified constraints, the algorithm searches for best-matched motion nodes in the facial motion database and synthesizes expressive facial animations. Then, based on user specified motion-node constraints and emotions, the bottom illustrates the timeline for the speech.

3.3.2 Xface

Xface is a MPEG-4 based open source toolkit for developing 3D talking agents [Balci 2004]. The independent platform incorporates three major modules: the core library, the XfaceEd editor and the XfacePlayer. The architecture of the system is designed based on extensible and configurable principles. Therefore, these three major modules are all independent and can be compiled with any ANSI C++ standard compliant compiler. It also uses some other open source tools to incorporate Xface such as the apml2fap tool to parse APML [DeCarolis

2002] scripts and to generate facial animation parameters, and the Festival2 tool to synthesize facial animation speech [Festival]. Xface also supports keyframe animation for better implementation.

The core Xface library can load face models and corresponding FP and FAPU information into Xface. It is designed based on the MPEG-4 standard and is able to playback MPEG-4 FAP (Facial Animation Parameters) streams. An Xface library can stream FAP data as well as the deformation of the face mesh to help create facial animation. The 'Xface library is also responsible for the rendering process using OpenGL, which is virtually available for all desktop systems.' It is an independent framework so developers can easily embed 3D facial animation into their software; also, researchers who are focusing on related topics can utilize it into their own implementations without creating a new one from scratch. With the clean interface of the library, they can easily integrate Xface to their applications. MPEG-4 provides a standard way of encoding various facial actions by defining parameters. But before applying it on a new face model, one has to define their own FAPUs and FPs manually. Xface Ed editor provides an easy way to do this process, so that in Xface one can define FAPU and FPs and also other weights and parameters on a static face model before animating it. It simplifies the preparation of ready face meshes from existing 3D face models in order to create facial animation. The output of XfaceEd editor is a configuration file, which lets other users change the 3D face model without doing any programming. Figure 3-7 is a screenshot that illustrates the process of setting up FP regions. The last module, XfacePlayer, is the action module that demonstrates the whole toolkit functionality and how users can implement a face player using Xface library. This component is designed based on a SDL3 [Deng 2006] library to manage windows, audio controls and the user interface. It is also responsible for implementing a basic talking head by loading a MPEG-4 FAP file, an audio file for speech, and a configuration file.

One limitation is that Xface does not cover both the decoding and encoding side of creating facial animation. MPEG-4 does not provide an animation language but only a set of low-level parameters. Although Xface is a powerful system in facial animation, it lacks higher levels of abstraction, timing control and event management. As well, MPEG-4 facial animation framework usually is used on web or mobile applications, so a lot of other MPEG-4 implementations and applications have been proposed [Deng 2006].

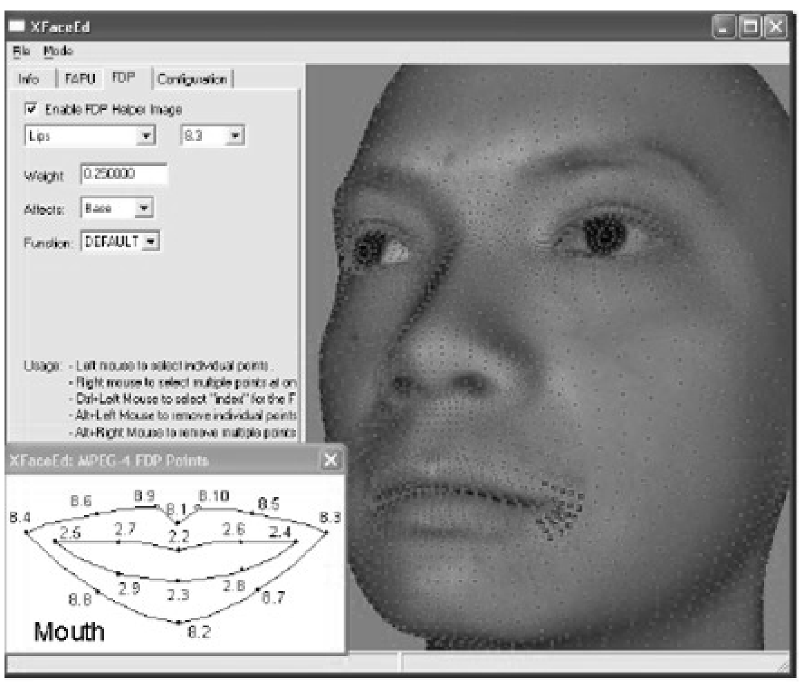


Figure 3-8 Sample shot from XfaceEd editor. It illustrates the process of defining mouth parameters in Xface.

Other research systems like Medusa use physics-based simulation methods to explore more anatomically correct models and synchronizations. The techniques they use range from scanning devices to subdivision surfaces, to mimic facial skin and its corresponding subtle muscle actions [Haber 2001]. Pogany designed a special interface for generating facial expressions with a sensing device. Users can touch each region of the device to control the corresponding region of the virtual head. This intuitive interface significantly reduces the painstaking process of manually keyframing mesh deformations to create expressions [Jacquemin 2007]. FaceEMOTE is another MPEG-4 based facial animation system which uses different parameterizing methods from Xface. It selects Effort parameters, attributed to the expressiveness or qualitative aspects of movements, as the high level parameters controlling low level parameters that define local deformations [Byun 2002]. One last example is Realface, a photogrammetric based application that delivers realistic textured 3D facial models from different camera views [Pighin 1998].

3.4 Other Facial Animation Systems

3.4.1 2D System

The most common technique in two-dimensional facial animation is commonly dependant on the transformation of images, including images from both still photography as well as sequences of video known as image morphing, and its variations. Traditional 2D animation is gained from a sequence of still images that are considered “keyframe” poses. Then, a camera may be used to shoot the sequence of papers. Because the images are keyframe poses, when human eyes watch them move quickly, because of the retentivity of vision, these images seem to blend with each other, and by this way are “animated”. Hence, the image sequences become a fluent video with a smooth transition using morphing in-between some frame techniques.

Recently, producing 2D animations has become easier with the help of computer graphic applications. CharToon is a system by which one can interactively construct parameterized 2D drawings and a set of time curves to animate the drawings [Hangen 1999]. The independent platform has been implemented by JAVA to produce 2D faces, and also is suitable for web applications. The system is composed of two major parts: animation parameter editor, and a 2D drawing package, which are a face editor and a movie player. The face editor is used to stall the geometry of the face and all the animation parameter description files. It also is responsible for the design of the drawings. Then the animation editor controls all the time-behaviour of the drawing’s animation parameters through their description file and saves the output as a script. In the end, the face player renders the movie from the face description input, with the script together with the drawings and exports everything to the final output video.

3.4.2 Web-based system

In the facial animation system fields, there are some unusual applications that are applied on slightly different works. For instance, a web-based facial animation system named WebFace, based on a reduced set of features for facial animation creatures, is commensurate for both commercial systems and research systems [Marri 2005]. This system uses a Candide wireframe model for 3D face modeling. It also provides texture mapping functions by registering images through a set of control points for mapping over 3D models. This performance can by delivered by the user over the web, thus making the web interface very important during the task. It is different from other standalone systems because it depends on the web, and it does not need a high competency requirement for the hardware. Developers can also optimize the entire graphical design and coding system to enable users to manipulate it over the web by remote access. Hence, users do not need to worry about the lag while using WebFace on the internet. The system has a demonstrated ability of real time performance as well as producing realistic web-based facial expressions.

Although there have been quite a few research systems, a disconnection between industry and academics still exist. For example, most commercial systems are based on motion capture methods, however, few research systems use that and use that properly. The animation industry is omitting some important issues like eye animation synchronizations, even though researchers are exploring this field more than ever. All these differences require more communication and connection between the two fields. That is why this survey is trying to clear some uncertainty that exists for each field and to find techniques that are working well and will also work well.

4: OUR FACIAL ANIMATION SURVEY

Facial animation appears to be starting to grow in techniques from simple morph targets and parameterization approaches to more face-centric sophisticated techniques. Due to the unclear situation of current facial animation systems, we want to contribute some valuable suggestions regarding both the current and future trends of facial animation techniques and systems. There are few surveys focusing on the computer facial animation domain and we hope this survey could 1) connect the research world and the industry world, 2) let the two worlds work together to gain incentive to build better systems, 3) develop more customized, face-centric and sustaining systems. Following, we present a survey of all the techniques and systems in the facial domain, and based on our responses, we discuss the characteristics and states of our current facial animation field to help evaluate the most desired techniques and methods for building future facial animation tools. See the survey question in the Appendix. We noted all techniques and types of facial animation and we devised a survey that we could give out to facial animators, facial animation tool builders, animation managers and research in the areas of films, games, tool companies and schools. The goal of which was to document what the common issues are that they currently have and what they would like to see improved. Hopefully this comprehensive survey can benefit all facial animators and tool makers to clear up some problems and help with better facial animation systems.

4.1 Sample Selection

We built a facial animation survey that aimed to find out the problems and hopes for facial animation techniques and systems. The whole process lasted nearly half of a year from building questions, applying for ethic approvals, and collecting data. Usually researchers try to involve a large number of subjects in a survey to assure that the sample is representative of the population they want to generalize in. But in our situation, those experts in computer facial animation are a small group in general, are spread out in industries that prefer to keep trade secrets, or are part of a animation group where it is hard or impossible to delineate the facial experts from the general characters or all around animation experts. So, the attempt to deliver a large sample survey is not applicable in such a narrow and idiographic discipline. Hence, our sample is not strictly a desirable random sample or a large size that is ideal in all survey studies. However, since North America is believed to be ahead of the world in the computer graphic field, we believe our sample, which is known to be biased, is still useful on some levels if we involve comparatively sufficient populations plus professional approvals. Therefore, we tried every possible sample source to cover a wide range of people who are making facial animations, including academic researchers, advanced university students that are learning and using facial animation tools, animation instructors, individual advanced users, facial animation experts in game companies and animation studios (both large and small) as well as facial animation tool makers. They are almost all experienced users that are very familiar with the present conditions in computer facial animation.

We contacted more than 200 computer facial animation experts in total. The methods of collecting resources included working with large 3D game and film companies, advertising notices on industry level animation and motion capture expert forums, checking the contact methods of facial animation researchers who have published papers and books, contacting significant game companies and animation studios to get permission to send out the survey and a few students who are in advanced animation classes. Out of the 200 hundreds experts we contacted, 51 subjects finished taking the survey. All the submissions were anonymous; however each of the subjects was given a unique password for possible future contact. The sample is composed of 8% of students; 8% of facial animation tool makers; 10% of animation studio users who work for higer game, film or advertisemen companies; 12% of animators from animation forums (where we are unsure of where they come from); 18% of animators from game companies; 18% of academic researchers; 26% of animators from film companies.

4.2 Survey

With the constant new developments in the computer facial animation area, F.A researchers, animators and tool makers all seem to be interested in more significant information and connections within the field. However, there is little comprehensive information to answer whether current techniques are working well or what are the most concerning issues in the future of computer facial animation – this type of information needs a rigorous comprehensive survey. But we were not able to find surveys done with facial animation. Hence, Prof. Steve DiPaola and I started this idea of delivering a rigorously planned and implemented facial animation survey including modeling techniques to find out what the state of facial animation tools are currently, what are current bottlenecks and issues in current facial animation systems, as well as what might be the most powerful techniques in the future and what kind of techniques need to be developed for better animations. This facial animation survey was created over a long period from building questions, making adjustments of questions based on professional opinions from the animation industry and animation research fields, applying for ethics approval, and building a strong online survey system for a diverse group of experts in the facial animation field.

Basically the survey has three aims:

- to understand the subjects' most used and preferred facial animation techniques and system types;
- to understand their most common issues on the techniques and systems they use;
- To gather their thoughts on future trends in facial animation techniques, functions and interfaces.

All checkbox, rating and text box questions are combined to balance the potential negative motions of iteration. On the experimental design side, we convened a committee meeting to iterate a strong survey setup; we then went to the Chair of SIAT, SFU, John Bowes and re-iterated the design and survey questions based on his strong expertise.

On the content side, questions, content lists and categories were vetted through many expert resources. Including:

Early work:

- Dr. Frederic Parke, Professor of Architecture and Co-Director Visual Studies Unit, Visualization Sciences Program Coordinator at the Department of Architecture, Texas A&M University. Dr. Parke is considered the father of facial animation with his pioneering parametric work at the Univ. of Utah and later as research Director of the Computer Graphics Lab at New York Institute of Technology. Dr. Parke is the author of the facial animation Bible, “Computer Facial Animation” now in its second edition [Parke 2008].
- Rex Grignon, Senior Supervising Animator at PDI/Dreamworks Animation and-founder the character animation group at PDI. He was the lead character or character animation supervisor on such CG movies as *Toy Story, Antz, Shrek and Madagasgar.*

Final Work:

- Laurent M. Abecassis is the Emmy award winning Technical Head and President of Di-O-Matic, Inc. (Montreal) which is a world leader in the development of high-end character animation software including Facial Studio, a major set of software for 3D facial animation. He is also known for his significant R& D work in the Computer Graphics industry, recognition of which was receiving a personal Emmy award.

- David Coleman, Senior Computer Graphics Supervisor at Electronic Arts, Canada. He managed the development of EA's internal facial animation solutions and has immeasurable experience in facial animation tools and real world facial animation issues. We had numerous conversations with David, culminating in a research lab site visit and a lengthy evaluation of all the survey questions, categories and technique lists.

Given this level of experimental design and facial animation technique expertise, we feel strongly that our survey design and questions are as optimal as possible given a chaotic knowledge area. Hopefully this survey can give some useful results and hints regarding the three concerns, after the study is accomplished.

4.3 Data Analysis

4.3.1 Research questions

• What kind of 3D tools and techniques people are using to generate facial animations?

• What are their big issues in the world of facial animation techniques?

• What techniques are believed to be the dominant among all facial animation techniques?

4.3.2 Ethic approval

The University and those conducting this research study subscribe to the ethical conduct of research and to the protection at all times of the interests, comfort, and safety of participants. This research is being conducted under permission of the Simon Fraser Research Ethics Board.

4.3.3 Descriptive statistics

The statistical package for social science (SPSS for windows 13.0) was used for data analysis. The study used descriptive statistics and nonparametric tests to analyze the data. Frequency data about each single question that can be coded were calculated.

First, we found out who took the survey by asking how many years they used facial animation and for what uses. Then we asked what they mainly used for tools and techniques. From this we would build a fundamental profile of our subjects and probably the type of people that we wanted to generalize in, as well as know the current state of facial animation. Second, we established the current issues facing facial animation in all areas as well as advanced facial animation areas that need better tools. Last, we established the structure of future facial animation techniques and tools by asking each technique's future development and software's future interface and solutions. As well, we included a few questions from free thoughts about facial animation five years in the future.

We analyze the raw data by defining a set of scores to numerically represent the categories and rankings. We then treat the scores as numerical measurements of the techniques and compare mean scores across groups (kind of like an ANOVA).

For example, with a rating level question “what techniques do you use for facial animation?” we could use

Scores / techniques	5 (Primary)	4(Secondary)	2(Not preferred)	0 (Never used)
A	23	13	8	7
B	13	8	10	20
C	12	7	8	24
:				

Table 4-1 Example of coding categories in this study.

In this case, we replace each technique’s ranking responses with the numerical scores and test to see whether the mean scores are different in any of the groups. Because we assume that the space difference between “secondary” and “not preferred but used before” is twice as big as the difference between “primary” and “secondary”; the space difference between “not preferred but used before” and “never used” is twice as big as the difference between “primary” and “secondary”, we believe that a 5-4-2-0 set of scores could represent the space between each groups better than ordinal scores without spaces. Also, because we give the option of *used but not like*, we could assume that missing responses could mean never used it and is represented by a zero score. Other ways of coding will be clarified later.

Within the sample, more than 80% have used three dimensional computer animation tools over five years, which means the sample is mostly composed of experienced animators [Figure 4-1] [Table 4-2]. We have several major subject types: people who do research and study facial animation, people who use facial animation to create games and films for work, people who study facial animation as students or because of an interest for many years, and people who develop new techniques and frameworks for facial animation. That assures our subjects are widely distributed among all kinds of animators, which increases the validity of our test results to represent a broad group of animators. Since we did not design any experiments or use any treatments on our subjects, there is no need to do a pre-test and post-test reliability test.

	Frequency	Percent	Valid Percent	Cumulative Percent
1 year or less	2	3.9	3.9	3.9
2-3 years	2	3.9	3.9	7.8
3-5 years	5	9.8	9.8	17.6
over 5 years	42	80.4	80.4	100.0
Total	51	100.0	100.0	

Table 4-2 Descriptive statistics value for question 4: In total, how long have you been involved in general 3D computer animation? (All types of 3d animation not just facial)

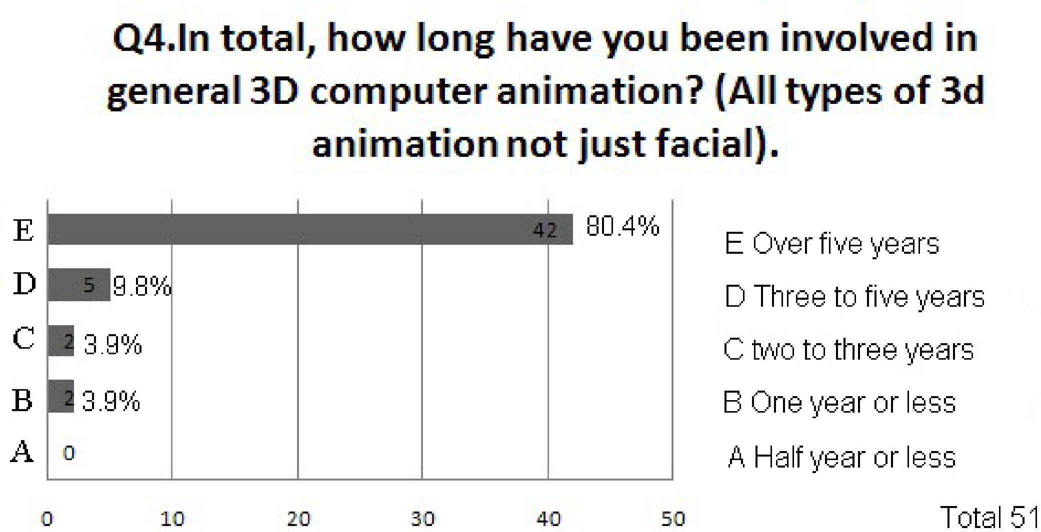


Figure 4-1 Data illustration of question four from facial animation survey.

Different types of facial animation application areas of these subjects are distributed quite evenly in this survey [Figure 4-2]. Both facial animation tools related and film related works are their main uses [Table 4-3]. Professional games and internet uses is another favourite choice. Comparably, not many people use facial animation to make independent games, or others such as medical and educational uses. Clearly, the main application of facial animation is still in its own field and public entertainment areas.

	Independent games	Professional games	Short or academic (student/research) films	Feature films	Internet/web/educational uses	Facial animation software tools uses	Other (medical, ...)
no	46	37	32	30	37	31	39
yes	5	14	19	21	14	20	12

Table 4-3 Descriptive statistics for question 6: What do you typically create facial animation for? Check all that apply

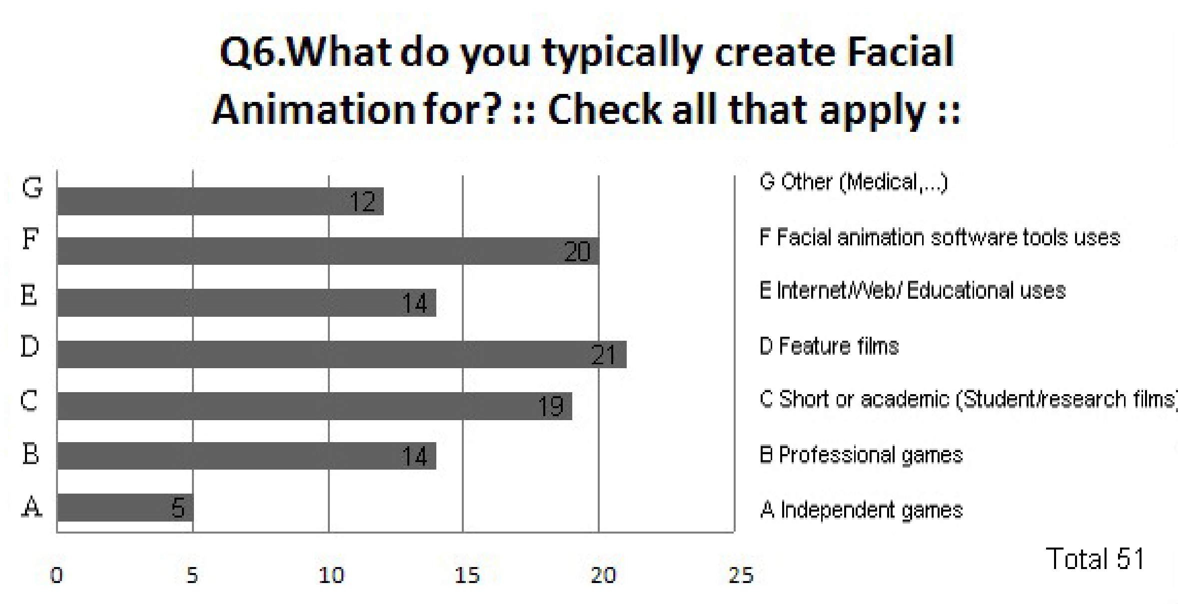


Figure 4-2 Data illustration of question six from facial animation survey.

The subjects claimed general large 3D packages are their primary choice for creating facial animation [Figure 4-3] [Table 4-4]. These packages generally adopt traditional techniques such as shape interpolation or bone based methods to create facial animation. A number of subjects prefer in-house, proprietary systems as well. Not many subjects are using stand-alone facial animation systems or facial plug-ins within large systems.

	Frequency	Percent	Valid Percent	Cumulative Percent
General large	25	49.0	49.0	49.0
Plug-in	9	17.6	17.6	66.7
Stand alone	4	7.8	7.8	74.5
In house	13	25.5	25.5	100.0
Outside service	0	0	0	100.0
Total	51	100.0	100.0	100.0

Table 4-4 Descriptive statistics for question 7: What type of software process do you use for creating facial animation?

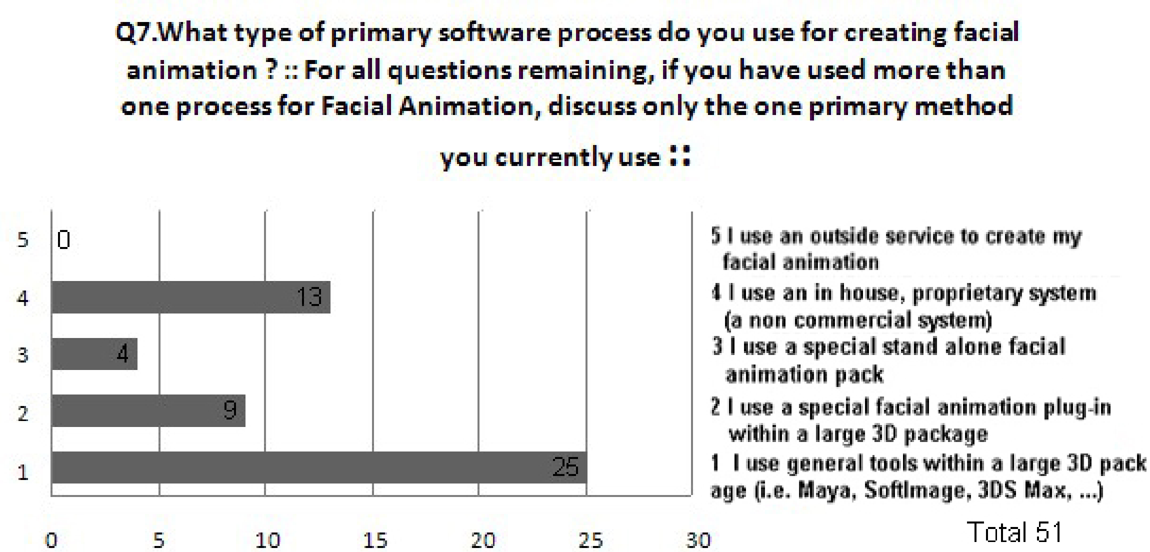


Figure 4-3 Data illustration of question seven from facial animation survey.

Knowing the current popular tools people are using, the current techniques listed correspondingly tells the same story. The largest mean of animator's primary facial animation technique is shape interpolation or blend shapes. The second largest is bone based technique [Figure 4-4] [Table 4-5]. These are the techniques that are adopted by general large 3D packages and are widely used. Hence, people are still using general large 3D packages and traditional techniques not specific for faces when creating facial animation. Also by looking at the raw data, among all the subjects, no matter if they are very experienced or less experienced users, all use more than one technique, and usually more than three techniques, with a high percentage using a bone based technique as well as motion capture. Within all motion capture based techniques, the most used is motion capture with markers. Very few are using motion capture without markers. Almost everyone using motion capture techniques also uses bones or blend shapes or both together.

	Blend shape or shape interpolation	Performance-driven techniques	Parameterization based	Deformation-based approaches	Bone-based system	Physics-based muscle modeling systems
Mean	3.55	2.29	1.94	2.51	3.47	1.53
Std.deviation	1.836	2.119	2.130	2.130	1.869	1.880
Variance	3.373	4.492	4.536	4.535	3.494	3.534

Expression cloning	Facial action coding system	Audio based lip-sync	Motion capture with keyframing	Motion capture with markers	Motion capture without markers	Motion capture libraries
1.20	1.86	2.45	1.67	2.02	1.22	1.00
1.778	2.088	1.983	1.946	2.186	1.803	1.442
3.161	4.361	3.933	3.787	4.780	3.253	2.080

Table 4-5 Descriptive statistics for question 9: What technique(s) do you use for facial animation?

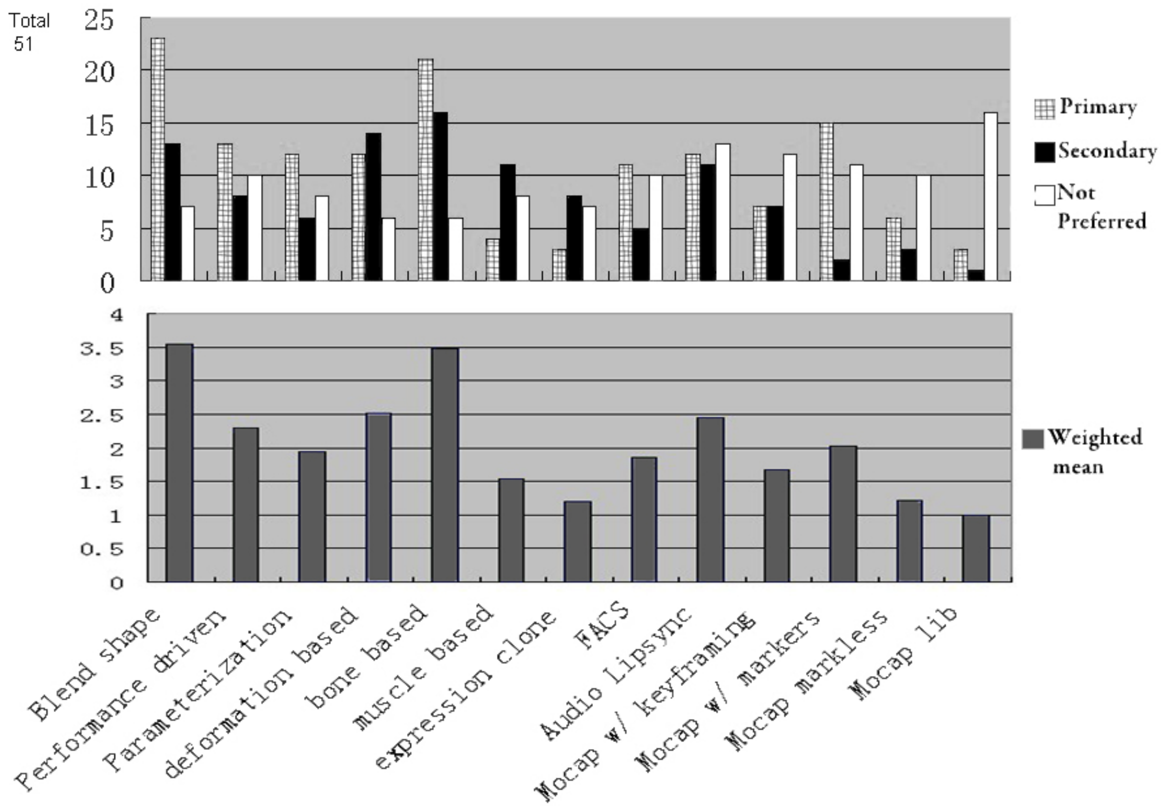


Figure 4-4 Data illustration of question nine from facial animation survey. What technique(s) do you use for facial animation?

More than 60% of our subjects are somewhat satisfied with the systems they are currently using. Whatever systems or techniques they are using, they are neither fulfilling their tasks perfectly nor unable to complete their work with them [Figure4-5].

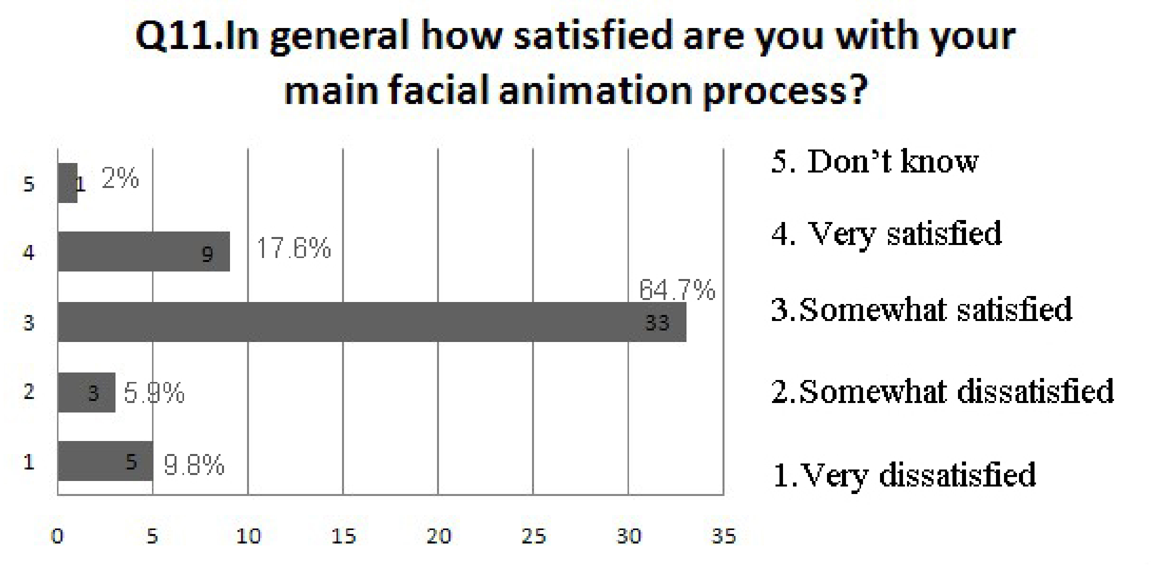


Figure 4-5 Data illustration of question eleven from facial animation survey.

To find out what areas people are unsatisfied specifically, we used a five point rating scale: "biggest issue", "big issue", "middle-level issue", "small issue" and "not an issue" representing areas in facial animation where most animators have issues. We choose a 5-4-3-2-1 set of scores to represent *biggest issue-big issue-middle level issue-small issue-not an issue* for statistical analysis. Hence, the higher the mean, the bigger the issue is. Most participants consider achieving realism as their biggest issue and ease of setup as the next issue to getting their work done. Similarly, few subjects choose these two areas as not an issue. They also prefer higher production speed besides realism and easy setup [Table4-6] [Figure 4-6].

	Achieving realism	Ease of Setup/rigging	Flexibility of authorship	Flexibility of animation style
Mean	3.75	3.50	2.69	2.78
Std. Error of Mean	.172	.165	.152	.152
Std. Deviation	1.230	1.165	1.065	1.075
Variance	1.514	1.357	1.134	1.155
Using any head model	Production speed	Lip Sync	Animation reuse	Full preview
3.14	3.35	3.10	2.70	2.64
.163	.168	.171	.170	.156
1.167	1.197	1.195	1.199	1.102
1.361	1.433	1.427	1.439	1.215

Table 4-6 Descriptive statistics for question 12: What are your biggest issues with current facial animation?

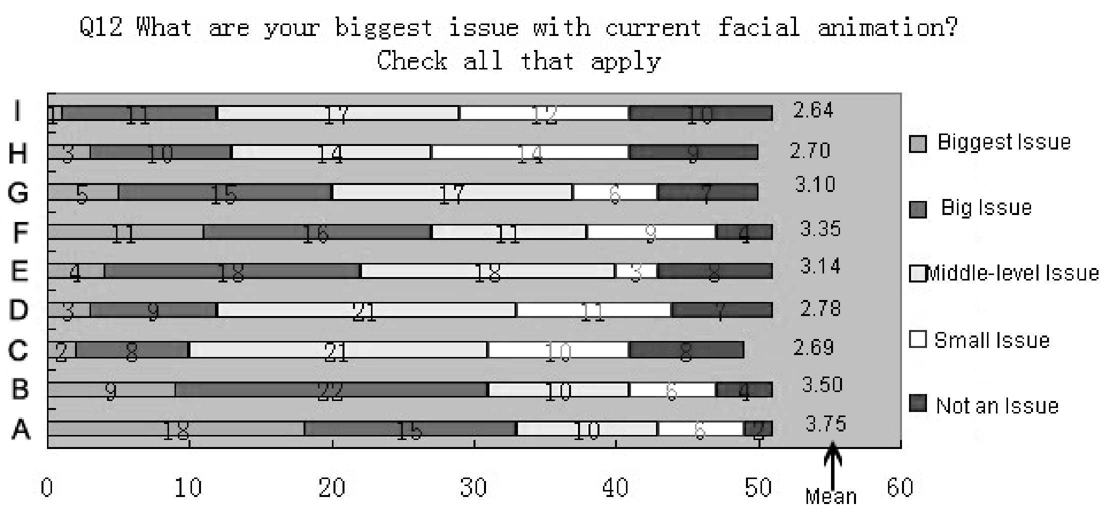


Figure 4-6 Data illustration of question twelve from facial animation survey. A represents achieving realism; B represents ease of Setup/rigging; C represents flexibility of authorship; D represents flexibility of animation style; E represents using any head model; F represents production speed; G represents lip Sync; H represents animation reuse; I represents full preview.

We also used a five point rating scale to represent the future developments of each facial animation techniques and see which ones are going to be dominant. There are varying opinions about where developers should concentrate on developing future techniques, although markerless motion capture and performance-driven techniques are where most people see the need, with many still interested in better bones based systems. Motion capture libraries was rated the lowest as a future direction [Table 4-7] [Figure 4-7].

	Shape interpolation	Performance-driven techniques	Parameterization based	Deformation-based approaches	Bone-based system	Physics-based muscle modeling systems
Valid	40	42	38	39	39	38
Missing	11	9	13	12	12	13
Mean	3.38	3.81	3.39	3.56	3.87	3.50
Std. Error of Mean	.202	.171	.212	.151	.173	.222
Std. Deviation	1.275	1.110	1.306	.940	1.080	1.371
Expression cloning	Facial action coding system	Audio based lip-sync	Motion capture with keyframing	Motion capture with markers	Motion capture without markers	Motion capture libraries
36	37	39	35	35	44	34
15	14	12	16	16	7	17
3.06	3.14	3.21	3.20	3.00	3.86	2.68
.199	.226	.208	.224	.213	.194	.201
1.194	1.378	1.301	1.324	1.260	1.287	1.173

Table 4-7 Descriptive statistics for question 14: What techniques do you think deserve development to be the dominant ones?

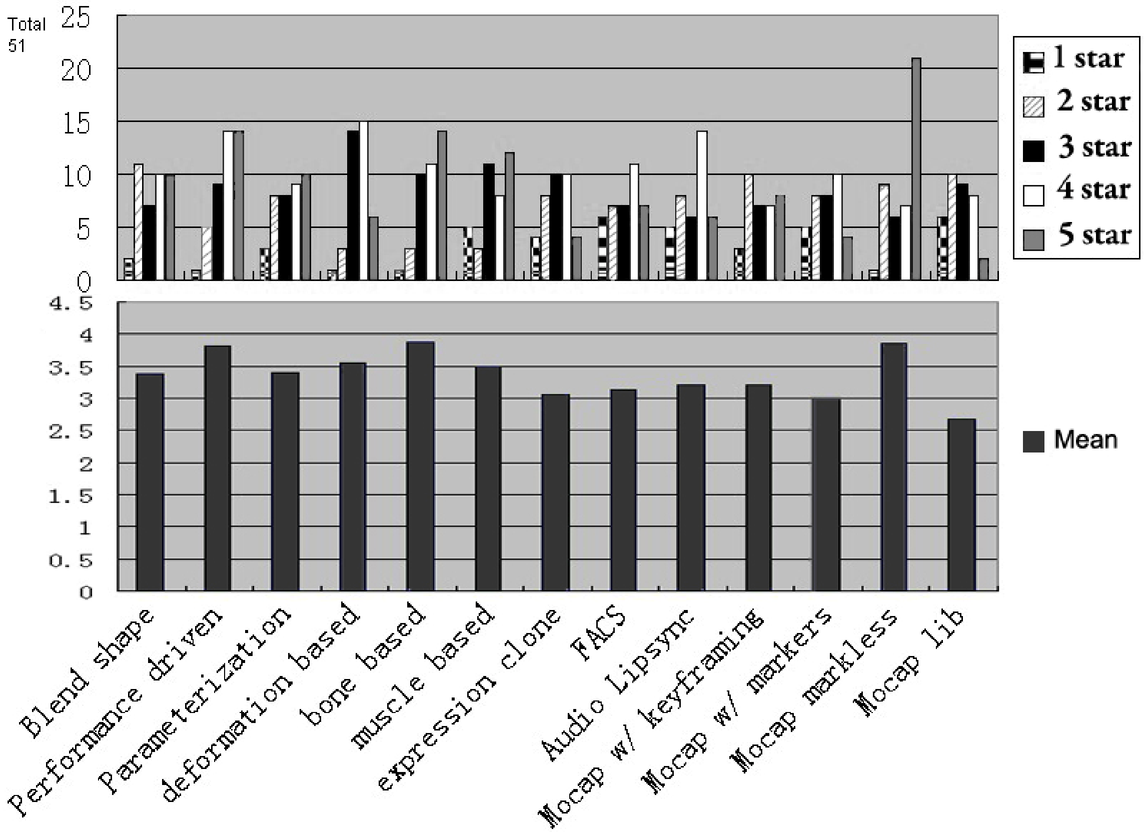


Figure 4-7 Data illustration of question fourteen from facial animation survey. What techniques do you think deserve development to be the dominant ones?

Overall, a high percentage of our survey takers thought there was a need for better tools in all categories, especially in lip sync, eye gaze and face muscle control. The three categories can all fit into achieving the realism issue in previous questions about their current biggest issues, which correspondingly enhances the desire of solving realism problems in doing facial animation. Also, a number of people prefer better tools on real time editing and motion capture cleanup, where motion capture cleanup can also fit achieving realism, again, backup the story [Table 4-8] [Figure 4-8].

	Lip-sync	Motion capture data clean up	Interface design	Real time editing	Face muscle control	Sync face and body gestures	Non emotion based face deformation	Eye gaze animation	Eye brow animation	Head movemen ts	Hair animation
Valid	51	51	51	51	51	51	51	51	51	51	51
Missing	0	0	0	0	0	0	0	0	0	0	0
Mean	.65	.49	.43	.51	.57	.45	.29	.57	.27	.24	.39
Std. Error of Mean	.068	.071	.070	.071	.070	.070	.064	.070	.063	.060	.069
Std. Deviation	.483	.505	.500	.505	.500	.503	.460	.500	.451	.428	.493
Variance	.233	.255	.250	.255	.250	.253	.212	.250	.203	.184	.243

Table 4-8 Descriptive statistics for question 16: What areas in facial animation do you feel need better tools?

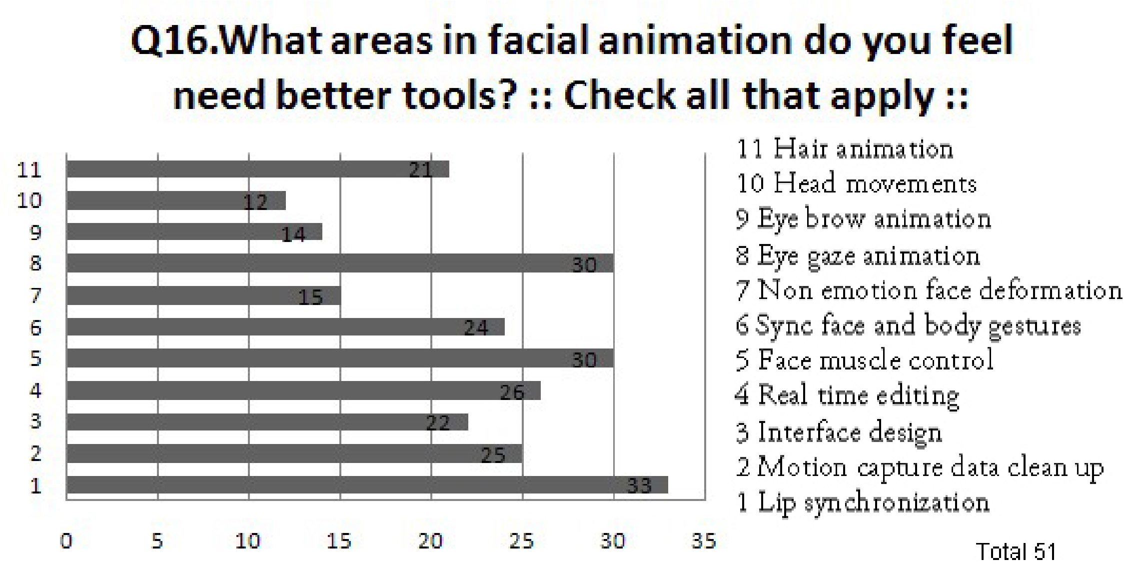


Figure 4-8 Data illustration of question sixteen from facial animation survey.

We asked specific questions about motion capture, focusing on the large concerns of this technique. Overall, more than 60% of subjects have used one or two motion capture systems when doing facial animation [Table 4-9] [Figure4-9]. About 37% of subjects have never used any motion capture systems; conversely, many subjects do use more than one motion capture system to accomplish their works. From the survey we can say that a large percent of animators are using motion capture systems, and they are more likely to see this technique continue rather than any other technique in the future regarding of their future developments.

	No	Yes-using a general motion capture system	Yes-using a video camera setup	Yes-sending the work out to a motion capture service
Frequency	19	21	16	7
Percent	37.3%	41.2%	31.4%	13.7%

Table 4-9 Descriptive statistics for question 17: Have you used motion capture in your facial animation work?

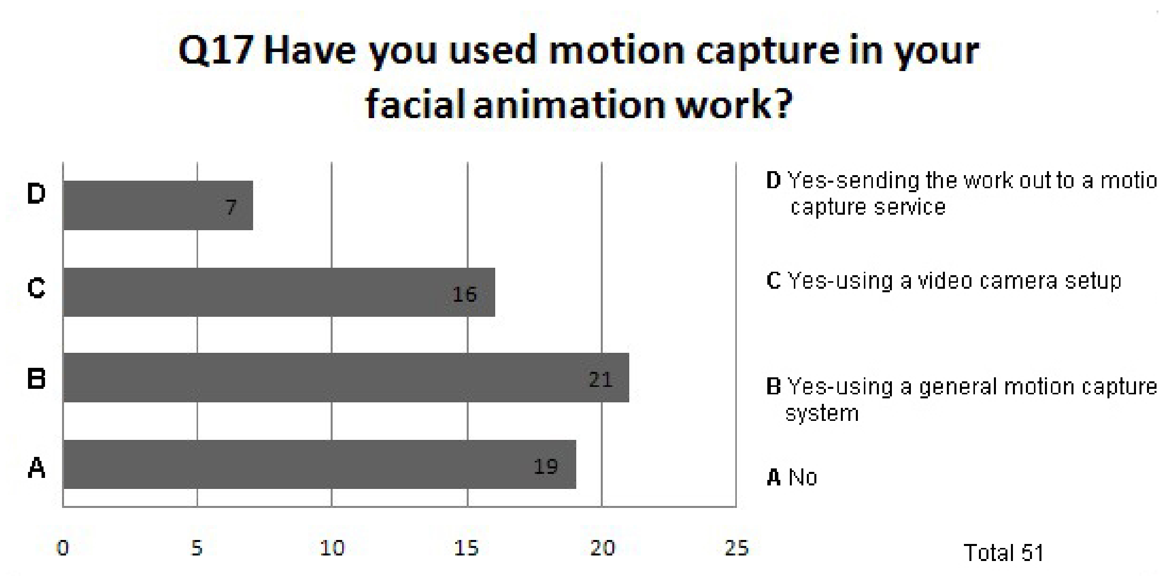


Figure 4-9 Data illustration of question seventeen from facial animation survey.

As facial animation is developing independently in the academic world, we built a question asking people's opinions about whether face tools should be separated with general 3D tools. The confirmed response of separating them is slightly higher than "it doesn't matter". Only a few people still think they should still be combined together [Table 4-10] [Figure 4-10]. This might be because they not only are doing facial animation but also are doing body animation and other texturing and lighting works, and they are willing to do them in one package rather than importing models from different software. As well, sometimes the different formats or object files can cause severe problems transferring between different systems.

	Frequency	Percent	Valid Percent	Cumulative Percent
Combine	10	19.6	19.6	19.6
Separate	22	43.1	43.1	62.7
Doesn't matter	19	37.3	37.3	100.0
Total	51	100.0	100.0	

Table 4-10 Descriptive statistics for question 20: Which do you agree?

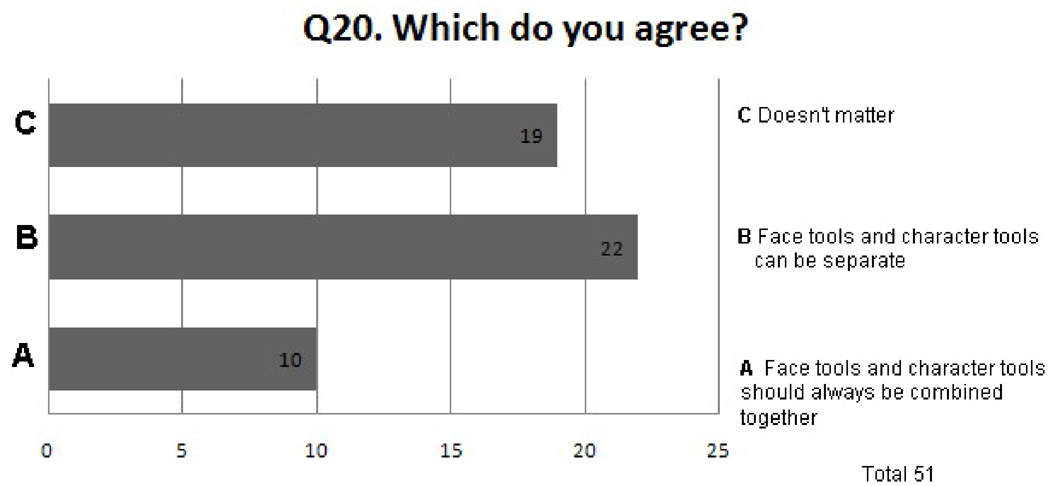


Figure 4-10 Data illustration of question twenty from facial animation survey.

4.3.4 Comparison test

In this research, we only analyzed parts of the questions in the survey because there could be many possible ways and possible directions to take in the data. We have proposed a few selection questions on the data, and proposed descriptive statistics previously. Furthermore, we also tested comparisons to get more possible results from the survey, such as if there are any differences between game users and film users on their current using techniques and how much they like them. If there are any different attitudes between motion capture users and non motion capture users on future developments of facial animation techniques, in each case, we divide the subjects into two groups and compare the means using non-parametric statistical methods such as the Two-Sample Kolmogorov-Smirnov test and the Mann-whitney U test, which are the most useful and general non parametric methods for comparing two samples. The Kolmogorov-Smirnov test is more reliable and suitable for small sample size studies that are under 40 subjects. The Mann-whitney U test is better for large sample size studies.

4.3.4.1 Gamer users versus film users study

In this study, unlike you are either motion capture users or non- motion capture users, subjects might use facial animation to create both games and films or just for internet and web uses. Those cases are not suitable for the following comparison tests because this test only compares film people versus gaming people, so we omit these answers and just look for subjects that choose games only or films only so that we can make two independent groups. First, we want to know if the primary techniques of film and game users are the same. Taking an alpha risk as .05, the null hypothesis is:

There is no difference between film users' primary technique and game users' primary technique.

The variable for making groups is defined as V6 which can be 0 or 1. There are 8 subjects in the game group representing by 0 and 25 subjects in the film group representing by 1 [Table 4-11].

No technique's significant value is less than .05, thus the null hypothesis is correct [Table4-12]. There is no significant difference between film users' mainly used techniques and game users' mainly used techniques. But there are some different trends regarding each technique between the two groups [Table 4-13]. By looking at the means, we can say that most game users do not use the blend shape technique which is the traditional facial animation technique, but many film users do. Film people use deformation based techniques a lot more than gaming people. Gaming people use performance driven techniques a lot more than film people, specifically motion capture with keyframing.

V6		Number
Game group	0	8
Film group	1	25
Total		33

Table 4-11 Each and total sample sizes of film vs. game study.

	A	B	C	D	E	F	G	H	I	J	K	L	M
Kolmogorov-Smirnov Z	1.268	.640	.320	.849	.640	.480	.283	1.243	.443	.160	.542	.320	.775
Asymp. Sig. (2-tailed)	.080	.807	1.000	.466	.807	.975	1.000	.091	.989	1.000	.931	1.000	.585
game Mean	3.13	2.75	2.00	1.75	2.88	.75	.75	3.00	1.63	1.38	2.63	1.25	.50
film Mean	3.92	1.84	1.56	3.08	3.56	1.64	1.16	1.12	2.08	1.44	1.68	.92	1.12

Table 4-12 **Comparison test of film and game group of question 9 from facial animation survey. A represents Blend shape or shape interpolation; B represents Performance-driven techniques; C represents Parameterization based; D represents Deformation-based approaches; E represents Bone-based system; F represents Physics-based muscle modeling systems; G represents Expression cloning; H represents Facial action coding system; I represents Audio based lip-sync; J represents Motion capture with keyframing; K represents Motion capture with markers; L represents Motion capture without markers; M represents Motion capture libraries.**

Second, we test if there is difference between film user' current issues and game users' current issues. Taking an alpha risk at .05, the null hypothesis is there is no difference between game and film users' current issues.

Table 4-13 shows the results of the Kolmogorov test. No significant value of any issue is less than .05, thus the null hypothesis is correct, and there is no significant difference between film and game users' attitudes regarding current facial animation issues. However, by looking at the means we can also hypothesize that game users have a tendency for achieving realism more than others. While film users want faster production speed as well as ease of setup/rigging process. Full preview and flexibility of authorship are not very big issues to both film and game users.

	Achieving realism	Ease of Setup/rigging	Flexibility of authorship	Flexibility of animation style	Using any head model	Production speed	Lip Sync	Animation reuse	Full preview
Kolmogorov-Smirnov Z	.566	.561	.320	.227	.689	.591	.281	.357	.535
Asymp. Sig. (2-tailed)	.906	.911	1.000	1.000	.729	.876	1.000	1.000	.938
game Mean	3.88	3.43	2.63	2.57	3.13	3.00	3.00	2.88	2.71
film Mean	3.36	3.52	2.56	2.72	2.84	3.56	3.12	2.56	2.64

Table 4-13 Comparison test of film and game group of question 12 from facial animation survey. A represents achieving realism; B represents ease of Setup/rigging; C represents flexibility of authorship; D represents flexibility of animation style; E represents using any head model; F represents production speed; G represents lip Sync; H represents animation reuse; I represents full preview.

Next we test if there is any difference between film and game user's attitudes on facial animation areas need better tools to solve different problems. The null hypothesis is there is no difference between film and game user's attitudes on facial animation areas need better tools. The Kolmogorov test shows no significant value is less than .05, hence the hypothesis is correct. However, by looking at the means, we can also know game users Film users want, mostly, better tools for lip synchronization and eye gaze animation. Besides, game users want better tools on real time editing and face muscle control. Eye brow animation and head movements are currently not in animators most want list.

		A	B	C	D	E	F	G	H	I	J	K
Kolmogorov-Smirnov Z		.775	.049	.455	.960	1.169	.652	.234	.258	.222	.726	.665
Asymp. Sig. (2-tailed)		.585	1.000	.986	.315	.130	.788	1.000	1.000	1.000	.667	.769
game	Mean	.88	.50	.63	.75	.88	.63	.38	.63	.25	.38	.25
film	Mean	.56	.52	.44	.36	.40	.36	.28	.52	.16	.08	.52

Table 4-14 Comparison test of film and game group of question 16 from facial animation survey. A represents lip synchronization; B represents motion capture data cleanup; C represents interface design; D represents real-time editing; E represents face muscle control; F represents Sync face and body gestures; G represents non emotion face deformation; H represents eye gaze animation; I represents eye brow animation; J represents head movements; K represents hair animation

4.3.4.2 Motion capture users versus non-motion capture users study

This study is to test differences between motion capture and non-motion capture groups. The first test is if there are any differences between motion capture users' mainly used techniques and non-motion capture users' mainly used techniques. The variable to make groups is defined as V17, which values 0 or 1 [Table 4-15]. Twenty non-motion capture users are represented by 0 and thirty-one motion capture users are represented by 1. The null hypothesis is there is no difference between motion capture users and non-motion capture users' mainly used techniques. Taking an alpha risk as .05, the results of Mann-Whitney U test show that there are six techniques that have significant values smaller than .05 and five of them are performance-driven related techniques [Table 4-16]. It is easy to understand because we are performing this analysis on motion capture groups and non-motion capture groups. The other technique that is not motion capture related is expression cloning. Motion capture users are using expression cloning a lot more than non-motion capture users. Based on the results, we could reject the null hypothesis.

V17		Number
Non-motion capture	0	20
Motion capture	1	31
Total		51

Table 4-15 Each and total sample sizes of motion capture vs. non-motion capture study.

	A	B	C	D	E	F	G	H	I	J	K	L	M
Z	-1.712	-3.656	-.819	-.081	-.397	-.681	-2.553	-1.411	-1.848	-4.450	-5.254	-3.780	-3.421
Asymp. Sig. (2-tailed)	.087	.000	.413	.936	.691	.496	.011	.158	.065	.000	.000	.000	.001

Table 4-16 Comparison test of motion capture and non-motion capture group of question 9 from facial animation survey. A represents Blend shape or shape interpolation; B represents Performance-driven techniques; C represents Parameterization based; D represents Deformation-based approaches; E represents Bone-based system; F represents Physics-based muscle modeling systems; G represents Expression cloning; H represents Facial action coding system; I represents Audio based lip-sync; J represents Motion capture with keyframing; K represents Motion capture with markers; L represents Motion capture without markers; M represents Motion capture libraries.

The next aim is to test if there is any difference between current issues that motion capture users have and that non-motion capture users have. Taking an alpha risk .05, the null hypothesis is there is no difference between motion capture users' current issues and non-motion capture users' current issues. The results of the Mann-Whitney U test shows that there are two significant values smaller than .05 [Table 4-17]. We can reject the null hypothesis by saying that the attitudes regarding current facial animation issues on achieving realism and ease of setup/rigging are significantly different between motion capture groups and non-motion capture groups. Apparently, motion capture users want more realism in facial animation and more control of the setup and rigging section while creating facial animation.

	Achieving realism	Ease of Setup/rigging	Flexibility of authorship	Flexibility of animation style	Using any head model	Production speed	Lip Sync	Animation reuse	Full preview
Z	-2.315	-2.126	-.160	-.115	-1.863	-.956	-1.095	-.448	-.259
Asymp. Sig. (2-tailed)	.021	.033	.873	.908	.062	.339	.273	.654	.796
nonmocap Mean	3.25	3.05	2.70	2.74	2.75	3.15	3.32	2.60	2.58
mocap Mean	4.06	3.77	2.69	2.81	3.39	3.48	2.97	2.77	2.68

Table 4-17 **Comparison test of motion capture and non-motion capture group of question 12 from facial animation survey.**

The next test is aiming to test the future developments of facial animation techniques and the differences between the two groups. The null hypothesis: the rating levels of future developments on each technique between motion capture users and non-motion capture users are the same. Taking an alpha risk as .05, the Mann-Whitney U test shows that the attitudes of these two groups on future developments of motion capture with keyframe and motion capture with markers are significantly different [Table 4-18]. The motion capture group rates these two techniques much higher than the non-motion capture group. Also by looking at the means, even though the non-motion capture group does not use motion capture at all, they agree with the motion capture group on the future development of markerless motion capture. Many say it is the future technique to use. Apparently the non-motion capture users do not like motion capture with keyframe and with markers as much as motion capture users, but they do agree with the trends of markerless motion capture to be dominant in the future.

	A	B	C	D	E	F	G	H	I	J	K	L	M
Mann-Whitney U	159.000	151.000	153.000	170.000	141.000	164.500	127.500	147.000	130.500	82.000	61.500	156.000	97.500
Z	-.491	-1.537	-.597	-.155	-1.177	-.109	-.746	-.100	-1.477	-1.998	-2.571	-1.453	-1.424
Sig. (2tailed)	.624	.124	.550	.877	.239	.913	.456	.921	.140	.046	.010	.146	.155
nonmocap Mean	3.54	3.44	3.53	3.50	4.13	3.43	2.85	3.17	3.53	2.58	2.18	3.36	2.38
mocap Mean	3.30	4.04	3.30	3.60	3.71	3.54	3.17	3.12	3.00	3.52	3.38	4.10	2.86

Table 4-18 Comparison test of motion capture and non-motion capture group of question 14 from facial animation survey. A represents Blend shape or shape interpolation; B represents Performance-driven techniques; C represents Parameterization based; D represents Deformation-based approaches; E represents Bone-based system; F represents Physics-based muscle modeling systems; G represents Expression cloning; H represents Facial action coding system; I represents Audio based lip-sync; J represents Motion capture with keyframing; K represents Motion capture with markers; L represents Motion capture without markers; M represents Motion capture libraries.

The last we tested the difference between motion capture users and non-motion capture users' attitudes on current facial animation areas that need better tools. Taking an alpha risk at .05, the Mann-Whitney U test shows that one significant value is smaller than .05: motion capture data cleanup. By looking at the means we can know that motion capture users overwhelming think there should be better tools for motion capture data clean up [Table 4-19]. The non-motion capture users are not concerning with this problem. It is easy to understand because they are not using motion capture hence they are not familiar with this area. However, this can reveal the significant problem of data clean up in motion capture techniques once again.

	A	B	C	D	E	F	G	H	I	J	K
Z	-.035	-2.161	-.360	-1.816	-.787	-.560	-.550	-.214	-.948	-1.142	-1.836
Asymp. Sig. (2-tailed)	.972	.031	.719	.069	.431	.576	.582	.831	.343	.253	.066
nonmocap Mean	.65	.30	.40	.35	.50	.50	.25	.55	.20	.15	.55
mocap Mean	.65	.61	.45	.61	.61	.42	.32	.58	.32	.29	.29

Table 4-19 Comparison test of motion capture and non-motion capture group of question 16 from facial animation survey. A represents lip synchronization; B represents motion capture data cleanup; C represents interface design; D represents real-time editing; E represents face muscle control; F represents Sync face and body gestures; G represents non emotion face deformation; H represents eye gaze animation; I represents eye brow animation; J represents head movements; K represents hair animation

4.3.5 Discussion

In this survey we performed a comprehensive design of questions to help expound current states, current issues and future trends in computer facial animation. We gathered all professional responses to build a trustworthy database and a reliable statistical analysis. This section attempts to analyze both the descriptive frequency of many questions and by the comparison testing of them to gain as much information as possible needed in this thesis from the survey.

Currently, about 50% of people are still using general large 3D packages and corresponding traditional facial animation techniques to create facial animations. Our results show that there are mainly two dominant techniques, shape interpolation and bone based techniques. Game experts tend to use more performance-driven techniques than film experts, meanwhile film experts prefer shape interpolation, bone based techniques and deformation based methods. By a margin of 20%, game experts are more open to new and fast techniques, such as performance-driven techniques, than film experts. There are some academic techniques that have not been accepted by most industrial users, and thus scored lowest in this survey, such as expression cloning, motion capture libraries and physics muscle based techniques.

Generally a large percentage of experts are looking forward to see better tools on all aspects of facial animation with more than 70% agreeing there are issues for achieving realism and difficulties in the setup process. Those who are using motion capture overwhelmingly think (over 75%) that setup and achieving realism are very big issues in motion capture, while those that never used motion capture are less concerned about these areas and are more concerned with the issues of lip synchronization. The most concern by those that use motion capture are with the data cleanup process, this is the area that most motion capture experts want better tools. Overall, users mostly want more realism in their facial animation (about 75%), much easier setting-up and rigging control over the facial model (about 70%), and a higher production speed when creating facial animation (about 67%).

As for future developments, the distribution of their attitudes is uneven. Both game experts and film experts have a tendency towards seeing more work on marker-less motion capture. Although few experts are using marker-less motion capture at this time, over 70% of experts say that is the future they believe facial animation production will go. This is true for those who currently use motion capture and is also true for current game experts and film experts. There is a significant group of non-motion capture users that have doubts about a markerless motion capture future (over 30%) as well as about motion capture for character and facial animation in general. Because of cost, lack of control over the system, and the quality of directly recorded actions, they are not satisfied yet with motion capture techniques. This is an area we recommend needs more research and development in the future to improve these cost and performance issues. Another future trend from our analysis is bone-based methods. Bone-based methods are currently the most used technique in authoring computer facial animation (about 70%). Apparently people are content with this approach – one reason why is because most general tool systems use bones for full character animation and have rather sophisticated tools for using bone-based approaches. It allows animators to use similar techniques for the body as well as the face – and allows tool makers to concentrate on one technique to constantly improve. Since almost 70% of facial animation experts agree on either one of the motion capture marker-less and bone based approaches, we believe it is a strong future direction to combining motion capture marker-less with bone rigging techniques. And it is feasible to do in reality. A commercial company ImageMetrics (www. image-metrics.com) at times working with university researchers, has used this combination technique benefiting the work on several significant film and game projects.

Another problem with current facial animation is the lack of attention on eye movement/gaze animation and real time editing as revealed in our survey. Over half of the experts want better eye gaze animation and real time editing and agree there should be better tools in these areas, as well as in addition to areas of achieving realism, involving better face muscle control, or better lip synchronizations. Currently the difficulty of mimicking lifelike eye animation has not been completely conquered. Only a few techniques are focusing on this issue. However, there is a great desire regarding it among users and we recommend more research in this area.

The limitation of our survey is that the subjects are not randomly selected. This survey is different from others due to a lack of source subjects. Not everyone is familiar with facial animation and can understand the questionnaire. Even in the expert areas, many animators stick to the technique they use or are provided for by their tools, and have limited knowledge of the tools of other facial animation techniques. There is also a problem of terminology, where most techniques can go by different names and are in fact a culmination of several techniques, or in some cases a limited version of the technical definition of a technique. Working with our advisors and experts, we have made a significant effort in better defining technical, adhoc, and archaic naming conventions and in fact, are even re-categorizing techniques and other facial animation descriptions/lists into more appropriate categories. While this effort was one of the significant contributions of the research, facial animation still suffers from blurred boundaries where one technique starts and another ends – in fact many named “techniques” are a mix of several. Additionally, the ‘computer facial animation experts’ population is small, not even mentioning only English speaking countries and those who are willing to do the survey. Hence we could not ideally perform a perfect random statistical test and that limits the reliability of our survey results. However, the subjects who took the survey are all experienced and the fact that they do not need pre-test training strengthens our confidence. Also the subjects are not constrained to a small group of people so the diversity of our samples is great. Another small issue is because we assume people's common sense (a five rating level is from one to five), question 14 is not clear in five star rating, with 5 as the highest in the rating scale, which could influence and mislead the subjects on a very small level.

5: RECOMMENDATIONS FOR FUTURE SYSTEMS

5.1 Recommendations for Academia

Previous chapters have described the state of current facial animation techniques and systems both from an academic view and an industry view. Based on the survey results we can see there is a gap between academic facial animation and industrial facial animation regarding techniques and systems. This chapter attempts to provide an overview of the techniques that should be explored further and those where less effort should be put from an academic view. Since many techniques flow from academic research to research and development at major companies to finalized techniques in major applications or small plug-ins, what we mean by an academic view is fostering completely new techniques or maturing newer techniques for better adoption or transference to commercial developers. This differs from our recommendations for industry in our next section where streamlining or maturing known techniques and dealing with user issues, speed, user interface and production quality are the main issues.

Our subjects have designated eye gaze animation, lip synchronization, face muscle control, and real time editing as the advanced areas most needed for improvement. These are areas that the industry is looking for more academic and research support with; to provide better solutions and techniques in these emerging aspects, while at the same time remain easy to use. Sometimes, academic techniques are too obtuse and immature and are therefore not widely accepted by industry users. Mostly, given the small adoption rate from the industry of these tools as shown by our survey, researchers must be aware that these newer tools, while providing new benefits, have shortcomings such as poor setup, a confusing interface, not integrating well with animators' overall systems, and noisy data. The survey results shines a light on some of the reasons for low adoption including ease of setup, which along with achieving realism, are some of the most significant complaints in facial animation. This problem will typically decrease as researchers naturally perfect their emerging systems. However it is our recommendation that researchers be more vigilant about the importance of ease of setup. From the survey results, the most unknown or unused techniques are expression cloning, motion capture libraries, parameterization and physics muscle based techniques – techniques that are more well known and published within academia. Therefore, it is obvious to us that the benefits of these approaches need to be better transmitted to the industry if they indeed so they can be better evaluated and if appropriate more universally adopted. worth.

Many early manual based techniques such as shape interpolation and deformation based techniques were implemented without significant knowledge of to facial anatomy. While these methods (especially shape interpolation) are still popular according to our study, they have a limited growth potential because they lack this inherent facial knowledge. However, some researchers are still developing these techniques with improved algorithms or better mapping models, trying to make the interpolation realistic or at least informed by facial knowledge. Considering emerging technical trends, future and better facial animation techniques can be better served by incorporated data and knowledge of how human faces move and emote. There is increased data on how to synthesize motions from the real human face, or based on physical models that have the same structures and muscles as anatomical human faces. Nothing is more natural than the actual expressions of real people. Techniques will only have the same function or in the future, surpass what our faces can do when they are built like us. Hence, from our perspective, researchers should be focusing more on realistic simulations of real human faces, both internally and externally; on better ways of importing data files recorded from human faces, and on problems like data transferring and data cleaning during the process for achieving better quality and realism.

This implies getting better data acquisition in the first place, then transferring this recorded data footage to geometry meshes that can be applied on models, and finally cleaning non useful or noisy data, in other words post-processing. These are the needed keys for automating the procedure: by enhancing the production speeds we also achieve greater facial and emotional realism, (as revealed in areas where people need better tools in the survey).

Beside the push for more work and achievements in furthering motion capture, the data results from our survey shows there are some subjects who still prefer traditional methods rather than motion captured facial animation methods. A typical reason why they do not believe in motion capture (from our fill-in question: why are you not using it) is because they feel the most expressive and realistic facial animations are not made by motion capture, but by skilled animators through traditional methods. So it should be noted that there are facial animation experts who will always prefer the skill of an animator over data recorded or transferred techniques like motion capture. It is our belief that this group can be persuaded to see the benefit if they had more significant creative control as well as the ability for exaggeration and subtle refinements. There are other problems with motion capture adoption, as described by our subjects, including cost, clean-up, and especially complex and slow setup. These are areas that researchers need to look into along with the problems of creative control and expression as mentioned earlier.

If these problems are overcome and we expect that current research will begin to conquer most of them, we expect to see emerging automatic acquisition methods to be designed for new general motion capture systems. While there is much experimentation in motion capture solutions, markerless motion capture was identified in our survey more than any other technique as the most wanted future technique. There appears to be a clear expectation that markerless methods will alleviate some of the major current problems, most notably long and complex setup issues. Markerless systems, while being favoured as a strong future contender for facial animation, must overcome the problems of noisy and unreliable data acquisition, since this issue is also a major negative factor for our subjects. Markerless motion capture is much farther along for facial animation than it is for character animation, mainly because there are consistent areas of the face (i.e., middle of eyes, corners of mouth) that can be recognized and used as control points. While much research effort is ongoing with motion capture libraries, our subjects rated it the lowest for possible future developments. There is a real schism between academia and industry in this area, which could point to libraries being in the very early research stage or could point to something more significant that researchers should take note of. Again, if it can be shown that motion capture libraries will add more flexibility and customized control, then we believe it will be more readily adopted since it addresses the issues of why some do not like motion capture. However, it is obvious from our data that this possible benefit has not been communicated well to the industry.

Another technique our subjects are looking for easer and better implementation of is lip synchronization. It is an important topic in current and future facial animation. Researchers have continued to develop phonemes, text based, and other methods to match the lip movements to the given speeches. However, we think due to the complex interactions between phonemes and visemes, it is not enough to reproduce default lipsync positions. Speech synchronization must be customized and adjusted for an avatar's expressions and mental states. Techniques that can combine expression understanding from psychology with speech synthesis will achieve better results. These systems are just emerging from research labs.

We would be remiss to not bring up one of the most significant needs of our subjects, which is to achieve better visual realism. This is true in all areas but is especially important within certain fields, including the medical industry. Another recommendation then is while continuing research in visual facial realism, (shaders, hair, wrinkles, …) it is important to note that these need to go hand in hand with animated realism – that is how a more realistic face moves.

5.2 Recommendations for Industry

With the continuous growth of hardware, the facial animation industry is leading to high quality, better realism and faster speeds. Ongoing improvements in computer graphic technology help to generate more sophisticated models and animations. However, on the other side, great limitations still exist that stifle future improvement. To predict which technology is better than the other and will take its place in the future is neither easy nor accurate. This chapter gives general recommendations of possible facial animation trends and attempts to provide an overview of the broad and interwoven areas that should be explored more from the industry view – where ease of use, repeatability, and controlled high quality results using speedy production cycles are paramount.

In many facial animation industries, specific groups are using many different approaches, in interwoven and adhoc ways, during projects as a way to achieve desire facial animation results. Some subjects have hoped this mix and match process can be supplanted with one smart plug-in for all the major 3D software needed in the survey. Whether it is a plug-in or a standalone solution, it is obvious from the survey that the current industrial process needs a more integrated solution that supplies intuitive control, and ease of use with quality results. Since no one technique will be acceptable for all facial animators, at least such a solution saves the iterative process of importing and exporting as well as readjusts itself in a new application for better compatibility. Associated components such as hair and eye animation are being requested to also be combined to provide a comprehensive system. Our recommendation to solve many of the frustrations of the current systems discussed in the survey is to provide one integrated framework (possibly standards based) , which can not only allow strong interconnectedness from a data flow perspective, but can allow animators to enable modules for the techniques they are comfortable with. This deals with the current issue of piecemeal solutions in a way that still gives flexibility to animators' preferences for authoring facial animation. In the short term, this solution still allows companies to use varied techniques to get the job done but at least it will be in one integrated system with standard UI, and with training and acceptance of the same input and output formats. It could theoretically provide all the popular techniques such as bone rigging, motion capture or shape interpolation. Users could choose different modules to use and compose them to make their own applications. In the long term, this system can accept new techniques as they become available and can allow animators to become familiar with and comfortable with the new recommended techniques we discussed earlier without significant retraining.

Performance driven approaches have proved in real world scenarios to be extremely strong from a time efficiency point of view. Their future depends on the progress of how much emotional and gestural realism they can achieve. Although we can capture high quality expressions with fine details of a human inputted face, more work needs to be done to efficiently use this data, with less cleanup manipulations and better coordination, and with an avatar's emotions and expressions, to get exactly what we captured without still being a problem. This is where a big division occurs with our survey subjects, some still want to use the animator controlled techniques (i.e. bone-based, deformation –based) to get the exact expressiveness they demand. Others, based on production issues, have moved to motion capture and more data acquisition techniques. We see a way to combine these groups and solve their control and time efficiency issues by moving to “smart” performance and data acquisitions techniques that are then intelligently controlled and manipulated to better deal with controls and quality demands. We see this as a strong integrated approach.

The survey data results seem to indicate that tool makers should develop motion capture marker-less based systems that will easily record the face of an actor or actress. Another important issue is the data of facial muscle movements is very different from the data of eye animation, therefore a significant area that subject see need improvements is dealing with eye movements. These issues may be solved with an advanced understanding of the eye movements and signals sent by the eyes. We expect significant developments in the field of eye animations.

Contemporary industrial facial animation techniques are centered on motion capture and bone based approaches. Because of the sophisticated characteristics and flexible muscles and joints on the human face, bone based techniques applied on the face is much harder to do than if it is applied on the human body. The higher quality and realistic the animation is, the more complex the bone rigging needs to be. This can very complicated due to its rigging setup and use. We feel bone based techniques will be better served by being combined with emerging motion capture techniques. In this way, bone-based animation would be considered as a post processing method of the motion capture data, where it can adjust and set data captured by cameras. Using this integrated technique, high quality recorded facial actions acts upon a fully rigged model according to facial animation coding systems, so we can match the 3D models with our data.

Another issue of our subjects' concerns is the interface design of current facial animation systems. There is very little, if any, published papers in terms of user interface specific to 3D facial authoring. Good user interface design is now universally recognized as a crucial part of the software and hardware development process [Bowman 2004]. In our survey, most subjects agree that user interface design is a very important area where better tools will emerge.

Subjects were asked to discuss where they see facial animation interfaces going in the future (5 years out). By looking at the short text-based answers generally, many subjects want intuitive interaction with real time editing and less controllers and manipulations. Many felt the setup and adjusting process needs to be quicker, simpler and more intuitive. Some prefer large amounts of preset actions and expressions; others were more interested in tool flexibility and the ability to change tools to fit individual needs. Almost all motion capture subjects prefer “work simple” with their one computer where they can accomplish the data recording work with a single video camera. This is far from the current reality of many multiple cameras with a team of workers, and days of clean up in a typical motion capture process. At the same time, they would like to have real time retargeting so the actors can view their character performance in a “mirror” for greater feedback. Many subjects look forward to new emerging interface techniques such as “controller free manipulation in which you simply grab the section of face to move” or “touch screen - multi-touch manipulation”. The results show the significant role of interface design in the development of facial animation tools. Both researchers and companies will benefit from making a better interface with less controllers but better quality tools.

To sum up, our subjects generally want more customized and intuitive controls but with less initial setup. User interface control is a complicated issue, the details of which are beyond the scope of our survey and this thesis, but we feel it is important to mention that no techniques based solution, including those we have recommended, will be achievable without a strong pairing with new user interface designs that are intuitive to the user. This should be the work of a further facial animation tool study.

6: CONCLUSION

This thesis has provided a summary of human facial modeling and animation techniques and systems. We then conducted with a comprehensive survey relating current issues through to future trends of facial modeling and animation.

Facial animation techniques include shape interpolation, bone based approach, parameterization methods, performance driven techniques, lip synchronization, MPEG-4 standard based, FACS based, deformation based, muscle mimic based, eye animation, head animation and expression cloning methods. Facial modeling as a component of these techniques was accomplished by scanning, digitizers, or photo composition. There are no well defined boundaries of facial animation technology because often a new technique is designed partly based on an old one. We gave particular attention to the orderly categorization of all the techniques and tools, and how they were formed to the current domain state of facial animation. Hence, for future use, there will be a reference for future researchers to study this intriguing and complex area.

Facial animation is widely used in the entertainment industry, telecommunication and medicine, etc. The future development of facial animation depends mostly on computer graphic technologies. Some techniques appear to be developing as the dominant approaches, performing fast, flexible and realistic animations such as new forms of motion capture. On the other hand, advanced visualization and cognitive knowledge techniques are informing these systems to be more realistic and precise for medical fields, as well as more emotionally believable overall.

Quite a number of animators are still using traditional methods to create facial animation, but they appear to be willing to adopt more developed and comprehensive methods. Performance driven approaches have a potential to deliver more high quality animations within efficient time frames, but need better cleanup and setup solutions. Meanwhile, bone based techniques are still attracting a large proportion of users and could be another area of development. Given the survey results, we also believe that markerless motion capture could be part of a more intuitive integrated solution when used to capture the basic facial movements and transfer them onto a fully rigged 3D face model. Markerless motion capture still has technical hurdles to conquer before it is a viable commercial technique but within facial animation as opposed to full body animation, the process is more understood and attainable as shown by new companies such as ImageMetrics, MaMoCa (www.mamoca.com) and Inition (www.inition.co.uk/) as well as work being research at institutions world-widely [Furukawa 2009] [Kaimakis 2004] [Hasler 2009] [Rosenhahn 2007, 2008] [Sundaresan 2005].

Based on the survey results, this thesis also gives recommendations to both academia and industry for ways to improve techniques and tools, as well as provides some perspectives on the future direction of facial animation that we see are worth exploring.

7: REFERENCES

[Allen 1987] J. Allen, M.S. Hunnicutt and D. Klatt, From Text to Speech, Cambridge University Press, Cambridge, 1987.

[Amresh 2003] A. Amresh, G. Farin and A. Razdan. Adaptive subdivision schemes for triangular meshes, Hierarchical and geometrical methods in scientific visualization, Mathematics and Visualization, 319-327. 2003.

[Arai 1996] K. Arai, T. Kurihara and K. Anjyo, Bilinear Interpolation for Facial Expression and Metamorphosis in Real-Time Animation, The Visual Computer, 12 pp. 105–116, 1996.

[Ayra 2003] A. Arya and B. Hamidzadeh, Personalized Face Animation in ShowFace System, Int. Journal of Image and Graphics, Special Issue on Virtual Reality and Virtual Environments, vol. 3, no. 2, pp 345-363, 2003.

[Ayra 2003b] A. Arya and B. Hamidzadeh, FIX: Feature-based Image Transformations for Face Animation, IEEE Conf IT in Research and Technology (ITRE), Newark, NJ, August 12-14, 2003.

[Ayra 2007] A. Arya and S. DiPaola, Multi-Space Behavioural Model for Face-based Affective Social Agents, Journal of Image and Video Processing, Special Issue on Facial Image Processing, Vol. 2007, 12 pages, 2007.

[Balci 2004] K. Balci, Xface: MPEG-4 based open source toolkit for 3D Facial Animation, Proceedings of the working conference on advanced visual interfaces, May 25-28, 2004, Gallipoli, Italy

[Bowman 2004] D. Bowman, E. Kruijff, J. LaViola and I. Poupyrev, 2004 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc.

[Busso 2005] C. Busso, Z. Deng, U. Neumann, and S. Narayanan, Natural head motion synthesis driven by acoustic prosody features, Computer Animation and Virtual Worlds, 16(3-4):283–290, July 2005.

[Busso 2007] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, Rigid head motion in expressive speech animation: Analysis and synthesis. IEEE Transaction on Audio, Speech and Language Processing, March 2007.

[Byun 2002] M. Byun and N. Badler, FacEMOTE: qualitative parametric modifiers for facial animations, Proceedings of the 2002 SIGGRAPH Eurographics symposium on Computer animation, July 21-22, 2002.

[Catmull 1978] E. Catmull and J. Clark, Recursively generated b-spline surfaces on arbitrary topological meshes, Computer-Aided Design, 10(6):350–355, September 1978.

[Coquillart 90] S. Coquillart, Extended free-form deformation: A sculpturing tool for 3D geometric modeling, Computer Graphics, 24:187–193, 1990.

[Cyberware Laboratory Inc. 1990] Cyberware Laboratory Inc. 4020/RGB 3D Scanner with Color Digitizer, Monterey, CA, 1990.

[deGraf 1989] B. deGraf, Notes on Facial Animation, In State of the Art in Facial Animation, SIGGRAPH ’89 Tutorials, pp. 10-11. New York: ACM, 1989.

[DeCarolis 2002] N. DeCarolis, V. Carofiglio, and C. Pelachaud, From Discourse Plans to Believable Behaviour Generation, In International Natural Language Generation Conference, New York, July 2002.

[Deng 2003, 2004, 2005, 2006, 2006b] Z. Deng, J.P. Lewis and U. Neumann, Practical eye movement model using texture synthesis, In Proc. of ACM SIGGRAPH 2003 Sketches and Applications, San Diego, 2003.

[Deng 2004] Z. Deng, C. Busso, S. Narayanan and U. Neumann, Audio-based head motion synthesis for avatar-based telepresence systems, In Proc. of ACM SIGMM 2004 Workshop on Effective Telepresence (ETP 2004), pages 24–30, New York, Oct. 2004.

[Deng 2005] Z. Deng, J.P. Lewis and U. Neumann, Automated eye motion synthesis using texture synthesis, IEEE Computer Graphics and Applications, pages 24–30, March/April 2005.

[Deng 2006] Z. Deng, Data-driven facial animation synthesis by learning from facial motion capture data, Ph.D. Thesis, University of Southern California, 2006.

[Deng 2006b] Z. Deng and U. Neumann, eFase: Expressive facial animation synthesis and editing with phoneme-level controls, In Proc. of ACM SIGGGRAPH/Eurographics Symposium on Computer Animation, pages 251–259, Vienna, Austria, 2006.

[DiPaola 2002] S. DiPaola, Investigating Face Space, In ACM SIGGRAPH 2002 Conference Abstracts and Applications, SIGGRAPH '02. ACM, New York, NY, 207-207. 2002.

[DiPaola 2002b] S. DiPaola, FaceSpace: A Facial Spatial-Domain Toolkit, In Proceedings of IEEE Information Visualization, London, 105-109, 2002.

[Doo 1978] D. Doo and M. Sabin, Behaviour of recursive division surfaces near extraordinary points, Computer-Aided Design, 10(6):356–360, September 1978.

[Dyn 1990] N. Dyn, D. Levin and J. Gregory, A butterfly subdivision scheme for surface interpolation with tension control, ACM Transactions on Graphics, 9(2):160–169, April 1990.

[Efase] Z. Deng and U. Neumann, eFASE: expressive facial animation synthesis and editing with phoneme-isomap controls, Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation, September 02-04, 2006.

[EKman 1978] P. Ekman and W.V. Friesen, Facial action coding system: A technique for the measurement of facial movement, Palo Alto, Calif.: Consulting Psychologists Press.1978.

[Elisei 2007] F. Elisei, G. Bailly, A. Casari and R. Stephan, Towards eye gaze aware analysis and synthesis of audiovisual speech, Auditory-visual Speech Processing Conference, Hilvarenbeek, The Netherlands, 2007.

[Ezzat 1998] T. Ezzat and T. Poggio, MikeTalk: A Talking Facial Display Based on Morphing Visemes, IEEE Conf Computer Animation, 1998.

[Festival 2002] T. C. for Speech Technology Research University of Edinburgh. The Festival Speech Synthesis System, 2002. http://www.cstr.ed.ac.uk/projects/festival/.

[Freeman 2000] E. G. Freedman and D. L. Sparks, Coordination of the eyes and head: movement kinematics, Experimental Brain Research, 131(1):22–32, 3 2000.

[Furukawa 2009] Y. Furukawa and J. Ponce, Dense 3D motion capture for human faces, Computer Vision and Pattern Recognition, IEEE Conference on 20-25 June 2009 Pages1674-168.

[Gouraud 1971] H. Gouraud, Continuous Shading of Curved Surfaces, IEEE Trans on Computers 20:6 (1971), 623-629.

[Graf 2002] H.P. Graf, E. Cosatto, V. Strom, and F.J. Huang, Visual prosody: Facial movements accompanying speech, In Proc. of IEEE Int'l Conf. on Automatic Face and Gesture Recognition (FG'02), Washington, May 2002.

[Haber 2001] J. Haber, K. Kähler, I. Albrecht, H. Yamauchi and H.P. Seidel, Face to Face: From Real Humans to Realistic Facial Animation, Proceedings Israel-Korea Binational Conference on Geometrical Modeling and Computer Graphics 2001, 11-12 Oct 2001, pp. 73-82.

[Hanrahan 1985] P. Hanrahan and D. Sturman, Interactive Animation of Parametric Models, The Visual Computer 1:4 (1985), 260-266.

[Hasler 2009] N. Hasler, B. Rosenhahn and T. Thormahlen, Markerless Motion Capture with unsynchronized moving cameras, Computer Vision and Pattern Recognition, IEEE Conference on 20-25 June 2009 Pages 224-231.

[Homles 2009] R. Holmes, A Quick Introduction to Subdivision Surfaces, http://www.holmes3d.net/graphics/subdivision/, 2009.

[Jacquemin 2007] C. Jacquemin, Pogany: A Tangible Cephalomorphic Interface for Expressive Facial Animation, Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, September 12-14, 2007, Lisbon, Portugal.

[Lee 2002] S.P. Lee, J.B. Badler, and N. Badler, Eyes alive, ACM Trans. Graph. (Proc. of ACM SIGGRAPH' 02), 21(3):637–644, 2002.

[Lee 1993] Y.C. Lee, D. Terzopoulos, and K. Waters, Constructing physics-based facial models of individuals, In Proc. of Graphics Interface'93, 1993.

[Lee 1995] Y.C. Lee, D. Terzopoulos, and K. Waters, Realistic face modeling for animation, In SIGGRAPH Proceedings, pages 55–62, 1995.

[Lewis 1987] J.P. Lewis and F.I. Parke, Automatic Lip-Synch and Speech Synthesis for Character Animation, In Proceedings Human Factors in Computing Systems and Graphics Interface '87, pp. 143-147. Toronto, ON: Canadian Human-Computer Communications Society, 1987.

[Lewis 2005] J.P. Lewis, J. Mooser, Z. Deng, and U. Neumann, Reducing blendshape interference by selected motion attenuation, In Proc. ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3DG), pages 25–29, 2005.

[Loop 1987] C. Loop, Smooth subdivision surfaces based on triangles, Master's thesis, Department of Mathematics, University of Utah, 1987.

[Kalra 1992] P. Kalra, A. Mangili, N. Magnenat-Thalmann and D. Thalmann, Simulation of Facial Muscle Actions Based on Rational Free Form Deformations, In Proc. Eurpgraphics 92, pp. 59-69, Aire-la-Ville, Switzerland: Eurographics Association, 1992.

[Kaimakis 2004] P. Kaimakis and J. Lasenby, Markerless motion capture with single and multiple cameras, Image Processing, International Conference on Volume 4, 24-27, Pages 2607-2610, Oct. 2004.

[Khullar 1999] S.C. Khullar and N. Badler, Where to look? Automating visual attending behaviors of virtual human characters, In Proc. of Third ACM Conf. on Autonomous Agents, pages 16–23,1999.

[Klatt 1987] D. Klatt, Review of text-to-speech conversion for English, J. Acoustical Soc. Am., 82, 737–793,1987.

[Kobbelt 2000] L. Kobbelt. √K Subdivision. Computer Graphics Proceedings, 103-112, ACM SIGGRAPH, 2000.

[Marri 2005] M. Al-Marri, A. Al-Qayedi and R. Benlamri, WEBFACE: A Web-Based Facial Animation System, Proc. Int. Conf. on Web Technologies, Applications, and Services (WTAS), Calgary, Canada, 4-6 July, 18-22, 2005.

[McNeill 1998] D. McNeil, The face: a natural history, 1998, Little, Brown and Company, 1998.

[MPEG4 1997] I. J. N1901. Text for CD 14496-1 Systems, Fribourg Meeting, November 1997.

I. J. N1902. Text for CD 14496-1 Visual, Fribourg Meeting, November 1997.

I. Pandzic and R. Forchheimer. MPEG-4 Facial Animation: The Standard, Implementation and Applications, Wiley, 2002.

[Ned 1998] L.P. Nedel and D. Thalmann, Real Time Muscle Deformations using Mass-Spring Systems, cgi, pp.156, Computer Graphics International 1998 (CGI'98), 1998.

[Noh 2001] J.Y. Noh and U. Neumann, Expression cloning, Proc. of ACM SIGGRAPH’01, pages 277–288, 2001.

[Hagen 1999] P. Hagen, H. Noot and Z. Ruttkay, CharToon: a system to animate2D cartoon faces, EUROGRAPHICS ’99. The Netherlands, 1999.

[Osipa 2007] J.Osipa, Stop Staring: Facial Modeling and Animation Done Right. SYBEX Inc, 2007.

[Pandzic 2003] I. S. Pandzic, Facial motion cloning, Graphical Models, v.65 n.6, p.385-404, November 2003

[Parke 1972] F.I. Parke, Computer generated animation of faces, In Proceedings of the ACM Annual Conference - Volume 1. ACM'72. ACM, New York, 451-457. 1972.

[Parke 1974] F.I. Parke, A model for human faces that allows speech synchronized animation, In Proceedings of the 1st Annual Conference on Computer Graphics and interactive Techniques, SIGGRAPH '74. ACM, New York, 1974.

[Parke 1996] F.I. Parke and K. Waters, Computer Facial Animation. A. K. Peters, Ltd. 1996.

[Parke 2008] F.I. Parke and K. Waters, Computer Facial Animation. Second. AK Peters Ltd. 2008.

[Pearce 1986] A. Pearce, B. Wyvill, G. Wyvill and D.R. Hill, Speech and Expression: A Computer Solution to Face Animation, In Proc. Graphics Interface ’86, pages 136–140, May 1986.

[Perez 1994] J. Perez and M.F. Catedra, Application of Physical Optics to the RCS Computation of Bodies Modeled with NURBS surfaces, IEEE Trans. on AP, Vol.42, No.10: 1404-1411, 1994.

[Perrett 2008] M. Androvich interviews N. Perrett, Chaning the face of animation, 2008. http://www.gamesindustry.biz/articles/changing-the-face-of-animation.

[Pighin 1998] F. Pighin and J. Hecker, Synthesizing realistic facial expressions from photographs, In ACM SIGGRAPH 2006 Courses, SIGGRAPH '06. ACM, New York, 1998.

[Platt 1980] S. Platt, A system for computer simulation of the human face, MSc thesis, The Moore School, University of Pennsylvania, August, 1980.

[Platt 1981] S. Platt and N. Badler, Animating facial expression, Computer graphics. 15(3):245–252, 1981.

[Polhemus 1987] Polhemus Navigations Sciences. 3Space Isotrack Users manual, Colchester, VT: Polhemus Navigations Sciences, 1987.

[Rosenhahn 2007] B. Rosenhahn, Scaled Motion Dynamics for Markerless Motion Capture, Computer Vision and Pattern Recognition, IEEE Conference on 17-22, Pages 1-8, June 2007.

[Rosenhahn 2008] B. Rosenhahn, Markerless motion capture of man-machine interaction, Computer Vision and Pattern Recognition, IEEE Conference on 23-28, Pages 1-8, June 2008.

[Schleifer 2004] J. Schleifer, Character setup from rig mechanics to skin deformations: A practical approach, 2004. ACM SIGGRAPH'

[Sederberg 1986] T.W. Sederberg and S.R. Parry, Free-form Deformation of Solid Geometry Models, SIGGRAPH' 86, Computer Graphics 20:4, 151-160, 1986.

[Sundaresan 2005] A. Sundaresan and R. Chellappa, Markerless Motion Capture using Multiple Cameras, Computer Vision for Interactive and Intelligent Environment, 17-18, Pages 15-26, 2005.

[Sung 2003] H. Pyun, Y. Kim, W. Chae, H. Kang and S. Sung, A Blend-shape Approach to Facial Expression Cloning, SIGGRAPH/EUROGRAPHICS Symposium on Computer Animation 167-176, 2003.

[Terzopoulos 1990] D. Terzopoulos and K. Waters, Physically-based facial modeling, analysis, and animation, Journal of Visualization and Computer Animation, 1(4), 73-80, 1990.

[Valle 1994] L. Valle, F. Rivas and M.F. Catedra, Combining the Moment Method with Geometrical Modeling by NURBS Surfaces and Bezier Patches, IEEE Trans. on AP., Vol.42, No.3: 373-381, March 1994.

[Vertegaal 2000] R. Vertegaal, G.V. Derveer and H. Vons, Effects of gaze on multiparty mediated communication, In Proc. of Graphics Interface’00, pages 95–102, Montreal, 2000.

[Vertegaal 2001] R. Vertegaal, R. Slagter, G.V. Derveer, and A. Nijholt, Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes, In Proc. of ACM CHI Conference on Human Factors in Computing Systems, pages 301–308, 2001.

[Vlasic 2005] D. Vlasic, M. Brand, H. Pfister, and J. Popovic, Face transfer with multilinear models, ACM Trans. Graph., 24(3):426–433, 2005.

[Waters 1987] K. Waters, A muscle model for animation three-dimensional facial expression, ACM SIGGRAPH Computer Graphics, v.21 n.4, p.17-24, 1987.

[Waters 1993] K. Waters and T. Levergood, DECface: An automatic lip-synchronization algorithm for synthetic faces, Technical Report 93/4, Digital Equipment Corp, Cambridge Research Laboratory, 1993.

[Williams 1990] L. Williams, Performance-driven facial animation, Proceedings of the 17th annual conference on Computer graphics and interactive techniques, p.235-242, 1990.

[Wilson 2005] M. Hanlon, Face Robot facial animation technology pioneers new approach to believable, emotional digital acting. http://www.gizmag.com/go/4364/ 2005.

[Zorin 1996] D. Zorin, P. Schroder and W. Sweldens, Interpolating subdivision for meshes with arbitrary topology, SIGGRAPH ’96 Proceedings, pages 189-192, 1996.

APPENDIX "ONLINE SURVEY QUESTION FORM"

Thank you for taking the time to participate in this survey!

Please read the informed consent information before proceeding to the survey.

Informed consent by participants in a research study

The University and those conducting this research study subscribe to the ethical conduct of research and to the protection at all times of the interests, comfort, and safety of participants. This research is being conducted under permission of the Simon Fraser Research Ethics Board. The chief concern of the Board is for the health, safety and psychological well-being of research participants.

Should you wish to obtain information about your rights as a participant in research, or about the responsibilities of researchers, or if you have any questions, concerns or complaints about the manner in which you were treated in this study, please contact the Director, Office of Research Ethics by email at hal_weinberg@sfu.ca or phone at 778-782-6593.

Your completion and submission of this online survey will signify that you have read the description of the procedures, whether there are possible risks, and benefits of this research study, that you have received an adequate opportunity to consider the information in the documents describing the study, and that you voluntarily agree to participate in the study. Your completion and submission of this online survey signifies that you are either a student of Simon Fraser University, or are 19 years of age or older.

Any information that is obtained during this study will be kept confidential to the full extent permitted by the law. Responses gathered in the online survey will remain confidential through the use of an encrypted, secure website. Knowledge of your identity is not required. You will not be required to write your name or any other identifying information on research materials. Materials will be maintained in a secure location.

Risks to the participants, third parties or society:
This study involves the completion of online surveys regarding usage experiences of facial animation tools. Because this activity shall occur in a setting of the participants choosing, this research poses no risks to psychological or physical health.

Procedures:
Participants will be asked to complete an online survey where investigators will inquire about their previous experience, use habits and expertise with facial animation tools. The information you have contributed may be used in future studies that may be similar and may required future contact with you.

Benefits of study to the development of new knowledge:
The study will help to understand the problem areas and speed along improvements in future computer facial animation tools. Your participation in this survey will be a important contribution to the furtherance of the research.

You may withdraw your participation at any time. You may register any complaint with the Director of the Office of Research Ethics.

Director, Office of Research Ethics
8888 University Drive
Simon Fraser University
Burnaby, British Columbia Canada
V5A 1S6
+1 778-782-6593
email: hal_weinberg@sfu.ca

You may obtain copies of the results of this study, upon its completion by contacting:
Chen Liu (Principal Investigator) ani.chenliu@gmail.com or
Steve DiPaola sdipaola@sfu.ca

By continuing with this survey, you are agreeing that you have been informed that the research will be confidential, you understand the risks and contributions of your participation in this study, and you agree to participate. By continuing to participate, you are confirming that you are either a student of Simon Fraser University or are 19 years of age or older. By filling out this survey, you are complying to participate.

Q1 . Start here

- [ ] I confirm that I am 19 years of age or older, OR
- [ ] I confirm that I am a Simon Fraser University student.
- [ ] I agree to future contact.

Q2 . To continue, please insert your password that we sent in your invitation email. Note: this is good for one survey, if you need to retake or have others who want to take the survey, please email us for a new invite email.

password : ______________________ *

Q3 . Do you have experience with the 3D computer facial animation?

- ( ) Yes
- ( ) No

Q4 . In total, how long have you been involved in general 3D computer animation? (All types of 3d animation not just facial).

- ( ) 1/2 year or less
- ( ) 1 year or less
- ( ) 2-3 years
- ( ) 3-5 years
- ( ) over 5 years

Q5 . In total, what % (out of 100) of that time in the above question is specifically for 3D facial animation?

Percentage: : ______________________ *

Q6 . What do you typically create Facial Animation for? :: Check all that apply ::

- [ ] Independent games

- [ ] Professional games
- [ ] Short or academic (student/research) films
- [ ] Feature films
- [ ] Internet/web/educational uses
- [ ] Facial animation software tools uses
- [ ] Other (medical, ...)

Q7 . What type of primary software process do you use for creating facial animation? :: For all questions remaining, if you have used more than one process for Facial Animation, discuss only the one primary method you currently use ::

- ( ) I use general tools within a large 3D package (i.e. Maya, SoftImage, 3DS Max, ...).
- ( ) I use a special facial animation plug-in within a large 3D package.
- ( ) I use a special stand alone facial animation package.
- ( ) I use an in house, proprietary system (a non commercial system).
- ( ) I use an outside service to create my facial animation.

Q8 . Describe the software system you use for creating 3D Facial Animation.

Software name(s) and process : ____________________ *

Q9 . What technique(s) do you use for facial animation? :: If it is a combination of techniques, check off primary and secondary techniques that you mix together -- check as many in either category as apply ::

	Primary	Secondary	Not preferred but used before
Blend shape or shape interpolation :	( )	( )	( )
Performance-driven techniques :	( )	( )	( )

Parameterization based :	☐	☐	☐
Deformation-based approaches :	☐	☐	☐
Bone-based system :	☐	☐	☐
Physics-based muscle modeling systems :	☐	☐	☐
Expression cloning :	☐	☐	☐
Facial action coding system :	☐	☐	☐
Audio based lip-sync :	☐	☐	☐
Motion capture with keyframing :	☐	☐	☐
Motion capture with markers :	☐	☐	☐
Motion capture without markers :	☐	☐	☐
Motion capture libraries :	☐	☐	☐

Q10 . If you did not see the (primary or secondary) techniques you use, please describe them.

Q11 . In general how satisfied are you with your main facial animation process?

☐ Very dissatisfied

☐ Somewhat dissatisfied

☐ Somewhat satisfied

☐ Very satisfied

☐ Don't know

Q12. What are your biggest issues with current facial animation? :: Check all that apply ::

	not an issue	small issue	middle-level issue	big issue	biggest issue

Achieving realism :					
Ease of Setup/rigging :					
Flexibility of authorship :					
Flexibility of animation style :					
Using any head model :					
Production speed :					
Lip Sync :					
Animation reuse :					
Full preview :					

Q13 . What is your favorite facial animation in a game and movie; why (optional)?

Game: :

Movie: :

Q14 . Which techniques do you think deserve development to be the dominant ones?::check all that apply::

	1 star	2 star	3 star	4 star	5 star
Shape interpolation :					
Performance-driven techniques :					
Parameterization based :					
Deformation-based approaches :					
Bone-based system :					
Physics-based muscle modeling systems :					
Expression cloning :					
Facial action coding system					
Audio based lip-sync :					

Motion capture with keyframing :	◻	◻	◻	◻	◻
Motion capture with markers :	◻	◻	◻	◻	◻
Motion capture without markers :	◻	◻	◻	◻	◻
Motion capture libraries :	◻	◻	◻	◻	◻

Q15 . Are there other techniques you think deserve development to be the dominant ones?

Answer : ____________________

Q16 . What areas in facial animation do you feel need better tools? :: Check all that apply ::

- ☐ Lip-sync
- ☐ Motion capture data clean up
- ☐ Interface design
- ☐ Real time editing
- ☐ Face muscle control
- ☐ Sync face and body gestures
- ☐ Non emotion based face deformation
- ☐ Eye gaze animation
- ☐ Eye brow animation
- ☐ Head movements
- ☐ Hair animation

Q17 . Have you used motion capture in your Facial Animation work?

- ☐ No
- ☐ Yes-using a general motion capture system
- ☐ Yes-using a video camera setup
- ☐ Yes-sending the work out to a motion capture service

Q18 . What is the biggest issue(s) stopping you from doing any or more motion capture Facial Animation work?

Q19 . What would you like to see in the future Facial Animation software? Say 5 years from now. (Feel free to list as points)

Q20 . Which do you agree?

- ☐ Face tools and character tools should always be combined together.
- ☐ Face tools and character tools can be separate.
- ☐ Doesn't matter

Q21 . What do you imagine would be a great User Interface for future systems, that is a great way for you to interact with your facial model to intuitively create quality animation?