An anatomically-based 3D parametric lip model to support facial animation and synchronized speech

Scott A. King, Richard E. Parent and Barbara L. Olsafsky
Department of Computer and Information Sciences
The Ohio State University

Abstract

We present work on a new anatomically-based 3D parametric lip model for synchronized speech, which also supports lip motion needed for facial animation. The lip model is represented by a B-spline surface and high-level parameters which define the articulation of the surface. The model parameterization is muscle-based to allow for specification of a wide range of lip motion. The B-spline surface specifies not only the external portion of the lips, but the internal surface as well. This complete geometric representation replaces, the possibly incomplete, lip geometry of any facial model.

We render the model using a procedural texturing paradigm to give color, lighting and surface texture for increased realism. We use our lip model in a text-to-audio-visual-speech system to achieve speech synchronized facial animation.

1 Introduction

Facial animation is becoming more important as a communicative technique between man and machine. In addition it is pivotal in the development of synthetic actors. The lips play an extremely important role in almost all facial animation. They are a significant component of expressing emotion as well as being instrumental in the intelligibility of speech. Therefore, in order to achieve realism and effective communication, a facial animation system needs extremely good lip motion with the deformation of the lips synchronized with the audio portion of speech.

Because of the demands placed on the motion of the lips, the desire for high-quality rendering and the lack of internal geometry for most input data, we have chosen to use a generic deformable lip model.

This lip model which can be used with any human-like facial model, provides:

- a sufficiently controllable model to support lip synchronization as well as supporting other motions used in expressing emotions,
- a sufficiently smooth model to support quality rendering,
- internal geometry (the part of the lips in the oral cavity) usually not provided in digitized facial models,
- support for procedural texture maps to support quality rendering

The lip model consists of a B-spline surface and high-level parameters which control the articulation of the surface. The high-level controls strive to be intuitive, precise, and as complete as possible. The controls are intuitive so they can easily and effectively be used by an animator or director. The controls are precise because of the desire to support lip synched animation. The controls are complete in the sense that most useful positions and motions of the lips can be specified.

We choose a B-spline surface for its c^2 continuity and the ease of deforming the surface by simply moving vertices of the control mesh. The drawbacks of B-splines include difficulty in placing a part of the surface exactly in $\Re^3$, preserving volume, detecting collisions and rendering. Fortunately, by polygonalizing the model, post processing after deformations can achieve volume preservation and collision detection while rendering the polygons is straight forward. Polygonalization loses the c^2 continuity of the B-spline surface, but we control the quality and with phong shading the impact is minimal. Volume preservation and collision detection are the subject of ongoing research and are not presented here.

As a preprocessing step, the user guides a process which replaces the lip region in a given facial model and grafts the generic lip model onto the rest of the facial geometry. The controls can then be used to animate the lips, the deformation of which also drives facial deformation in the surrounding area.

Our current application is a text-to-audiovisual-speech system (TTAVS) and speech-synchronized speech is its primary goal. However, we aspire to create a complete facial model that can be used for all types of facial animation. Because our system uses keyframing for the animation, we require a parameterized lip model that:

- supports speech synchronization,
- can be used for any set of lips,
- can be realistically animated,
- supports other lip motion.

Furthermore, it is desirable for the lip model to allow:

- easy definition of the action of the parameters,
- intuitive creation of desired animation.

Since our application is a TTAVS system, the lip model must support speech synchronization. Emotion is also desirable, so the lip model should be capable of general lip motion. Our TTAVS system is capable of modeling any human-like head, so the lip model must be capable of representing any set of lips. Since not all lips behave the same, the model behavior should be easy to change and define. And finally, the parameterization should be intuitive for the animator to use.

2 Previous Work

Over the last three decades many techniques have been used in an attempt to create convincing speech-synchronized facial animation. It has proven a difficult task due to the complexity of the system and the low tolerance for inconsistancies in the animation from a human audience. Concentration on the lips for the synchronization has been a theme, but only one research team has created a separate lip model. Generally the concept of visemes, the visual elements of speech, is used under the guise of phonemes, the phonetic elements of speech. The model is placed in a position that represents a phoneme which by consequence deforms the lips.

The early work in speech-synchronized facial animation involved creating animation using traditional hand-drawn animation techniques. There was also early work in the speech and hearing community involving the use of oscilloscopes to generate lip shapes. Later work involved automated techniques to synchronize the lips with the audio. In this section we discuss the early methods of generating lips and lip synchronized animation.

2.1 Speech Reading

Fromkin [17] reports on a set of lip parameters that characterize lip positions for American English vowels using frontal and lateral photographs, lateral x-rays, and plaster casts of lips. The lip parameters identified are:

1. Width of lip opening.
2. Height of lip opening.
3. Area of lip opening.
4. Distance between outer-most points of lips.
5. Protrusion of upper lip.
6. Protrusion of lower lip.
7. Distance between upper and lower front teeth.

This parameterization of the lips is very good for speech but it does not allow for other lip motions, such as expressive ones. We instead base our parameterization on muscle actions.

Research by the speech community on lip reading involved drawing lip outlines to represent speech. The results from these works are not as good as we would like, and our work is also concerned with visual realism instead of just speech intelligibility. Boston [7] reports that a lip reader was able to recognize a small vocabulary and sentences of speech represented by mouth shapes drawn on an oscilloscope.

Erber [13] also uses an oscilloscope to create lip patterns and claim to create more natural motion. Their studies show that lipreading the display gives similar results to those of reading a face directly. Later, Erber [14] drew eyes, nose and a face outline on cardboard and cut out the mouth area and placed it over the oscilloscope. This is one of the earliest examples of real-time facial animation. Erber [15] uses high speed cameras to capture speech and determine lip positions by hand.

Brooke [9] draws outlines of the face, eyes and nose with movable jaw line and lip margins. Positional data is hand-captured from a video source. A perception study [10] is made to determine if a hearing speaker can identify the utterances. The natural vowels were identified 98% of the time and there was good identification of the synthetic vowels (/u/ 97%, /a/87%). The vowel /i/28% (which is almost always confused with /a/) and medial consonants were not identified.

Montgomery [25] draws lip outlines on a CRT from data hand captured from video frames in a system designed to test lip reading ability. They augment by adding nonlinear interpolation between frames as well as forward and backward coarticulation approximation.

2.2 Computer Models

In Parke's [27] ground-breaking work, he uses 10 parameters to control the lips, teeth, and jaw during speech synchronized animation. The parameters are chosen empirically and from traditional hand drawn animation methods. These parameters give control over the lip motion, but have little consideration over lip shape, and the lip motion is limited.

Bergeron [5] reports that for the film short "Tony De Peltrie" key frames were defined then interpolated using curves. The keyframes included those for phonemes. This method of shaping the lips for each phoneme is a common approach. However, it is difficult to create a complete set of poses that encompass the entire space of motion desired. For example, to have a pose of smiling and saying /a/ at the same time requires having to create that specific pose. By defining shapes that differ from the neutral in a single way, such as smiling, these shapes can be used to define a space of possible shapes. Each base shape is calculated as a displacement [4] from the neutral. By defining a percentage of each base shape more complex animations can be achieved. Our work is similar to this concept, however, we choose our bases using muscle displacements instead of phonemes and expressions, which we believe gives a larger space of possible lip shapes. Another difference is in how we combine the displacements.

2.3 ICP Lip Model

Guiard-Marigny [20] measures the lip contours of French speakers articulating 22 visemes in the coronal plane. Assuming symmetry, the vermillion region of the lips is split into three sections and mathematical formulas are created to approximate the lip contours. From polynomial and sinusoidal equations the 14 coefficients are reduced to 3 using regression analysis. The three parameters are internal lip width, internal lip height and lip contact protrusion. With the same technique on lip contours in the axial plane Adjoudani [2] identifies two extra parameters to extend the lip model to 3D. The two new parameters are upper and lower lip protrusion. Guiard-Marigny et al. [21] describe the 3D model in English. This work was carried out at the Institut de la Communication Parlée in Grenoble, France and we refer to the model as the ICP lip model.

Guiard-Marigny et. al.[22] replace the polygonal lip model with an implicit surface model using point primitives for fast collision detection and contact surfaces. For the polygonal model, to increase the speed of computation they use a keyframe animation technique. The inbetweens are calculated as the barycenter of a set of extreme lip shapes. There are two extreme shapes per parameter and the barycentric coordinates are the parameters. They build the implicit surface from point primitives for each of the 10 key (extreme) shapes. Any lip shape is found by interpolating the point primitive positions and field functions. Implicit surfaces give an exact contact surface [18] which allows modeling the interaction of the lips with other objects (a cigarette in their examples.)

We originally used the ICP lip model using interpolation of the 10 extreme shapes, in our TTAVS system. The ICP lip model was designed by analyzing speech and is only capable of representing lip shapes used during speech production. We desire a model capable of expressing at least simple emotion such as smiling. The ICP lip model also lacked visual realism while producing rounded lip positions for phonemes such as /o/ and /r/, since the corners didn't move correctly. While it was possible to significantly modify this model, it seemed more prudent to start from scratch creating an anatomically-based model that was parameterized and deformed based on muscle actions.

3 Lip Anatomy

The face is a biological system and the lips are deformed as a result of muscle contraction. We look at the anatomy of the lips and the underlying muscles to define possible lip motion. Any good general anatomy reference [19], speech and hearing anatomy reference [12] [3] or facial anatomy reference [8] will be a good source of information. In this section we give a description of the muscles that affect the lips as well as a description of the motion of the mandible.

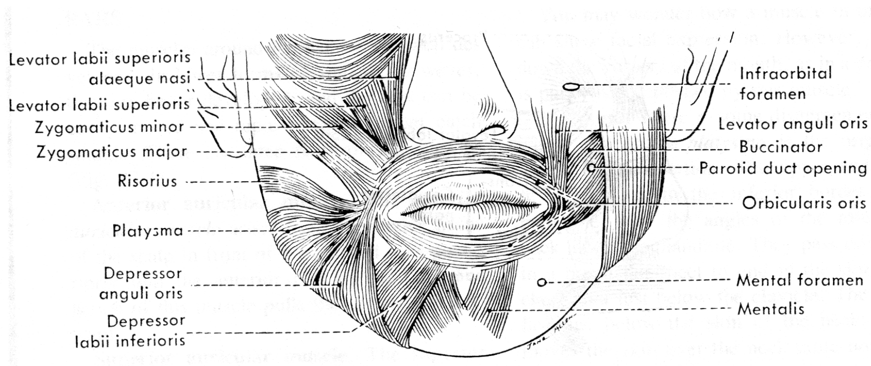


Figure 1: Muscles relating to the lips. The deeper muscles are pictured on the left side of the face and the superficial muscles on the right side. All but the orbicularis oris have a corresponding muscle on the left and right side of the face.

All muscles that affect the lips, except the orbicularis oris, are paired with a corresponding muscle on the right and left side of the face. Figure 1 depicts the muscles of the face with the superficial muscles shown on the right and the deep muscles on the left. Table 1 lists the muscles that affect the lips with a brief description of their actions.

3.1 The Mandible

The lips are tightly connected to muscle, with some muscles attached to the mandible, such as the depressor labii inferioris, so when the mandible moves so do the lips. In order to properly control lip deformations, the movement of the mandible must also be considered.

The temporomandibular joint is a diarthrodial ginglymous [26] (sliding hinge) joint that allows the mandible a large scope of movements. The mandible acts mostly like a hinge, with two separate joints acting together, and each of the joints having a compound articulation. The condyle of the mandible is nested in the mandibular fossa of the temporal bone. The upper part of the joint enables a sliding movement of the condyle and the articular disk, moving together against the articular eminence. In the lower part of the joint, the head of the condyle rotates beneath the undersurface of the articular disk in a hinge action between the disk and the condyle.

The mandible can be thought of as a 3 DOF joint, with the jaw moving in (retraction) and out (protraction), side to side (lateral movements), and opening and closing. The mandible lowers in a hinge-like manner except in extensive openings when the movement is downward and forward. During protraction and retraction the condyles slide on the articular eminences and the the teeth remain in gliding contact. Lateral movement is achieved by fixing one condyle and drawing the other condyle forward.

Name	**Action**
Buccinator	Compresses the cheek against the teeth, retracts corner of mouth.
Depressor anguli oris	Draws the corners of the mouth downward and medialward.
Depressor labii inferious	Depresses the lower lip.
Levator anguli oris	Moves the corner of mouth up and somewhat medially.
Levator labii superioris	Raises the upper lip and carries it a little forward.
Mentalis	Raises and protrudes the lower lip.
Orbicularis oris	Closes the lips, compresses the lips against the teeth, and protrudes the lips.
Platysma	Pulls the corner of the mouth down and back.
Risorius	Pulls the corner of the mouth back.
Zygomaticus minor	Draws the outer part of the upper lip upward, laterally and outward.
Zygomaticus major	Draws the angle of the mouth laterally and upward, as in laughing.
Incisive superior	Pulls the upper lip in towards teeth.
Incisive inferior	Pulls the lower lip in towards teeth.

Table 1: Muscles that directly affect the lips and their actions.

4 Lip Parameterization

Parameterizing the motion of the lips allows us to reduce the number of degress of freedom of the system. The goal is to minimize the number of degrees of freedom while still providing flexibility and generality. Besides a minimal set, we need a parameterization for the lip motion that is intuitive to use, easily defined and modified for different mouths and supports speech synchronization and the wide range of other lip motions needed for facial animation.

The lips deform due to the contraction of the connected muscles and the movement of the mandible. We use the muscles that affect the lips as the basis for our parameterization resulting in anatomy-based deformations. The parameterization must also include the movement of the mandible, which when moved affects the position of the lower lips, and thus the position of the upper lips. The muscles are a good choice because their action is mostly along a vector, so their effect on the lips can be easily defined. This works for all muscles but the orbicularis oris which actually constricts and protrudes the lips. In most cases we have a parameter for each muscle, one for the left and one for the right side. Exceptions are made for the depressor inferioris, depressor oris and mentalis since individual control is rare. The incisive inferior and incisive superior are also treated as a single muscle. Some texts consider these part of the orbicularis oris, but the action of pulling the lips in tight against the teeth is needed as a seperate degree of freedom from that of the orbicularis oris. And lastly we treat the levator labii superioris and the zygomaticus minor as a single muscle since the zygomaticus minor is usually not well developed and their actions are very similar.

The 21 parameters we use for our lip model along with their definitions are:

- Open Jaw - The mandible rotating open which lowers the lower lip and the corners of the mouth causing the lower lip to rotate out slightly.
- Jaw In - Movement of the mandible superiorly and inferiorly causing the lower lip to move inward or outward.
- Jaw Side - Lateral movement of the jaw causing the lower lip to skew to one side or the other.
- Orbicularis Oris - Contraction of the orbicularis oris muscle causing the lips to pucker.
- Left Risorius - Contraction of the left risorius muscle pulling the left corner of the mouth back.
- Right Risorius - Contraction of the right risorius muscle pulling the right corner of the mouth back.
- Left Platysma - Contraction of the left platysma muscle pulling the left corner of the mouth down and back.
- Right Platysma - Contraction of the right platysma muscle pulling the right corner of the mouth down and back.
- Left Zygomaticus - Contraction of the left zygomaticus muscle pulling the left corner of the mouth up and back.
- Right Zygomaticus - Contraction of the right zygomaticus muscle pulling the right corner of the mouth up and back.
- Left Levitator Superior - Contraction of the left levator labii superioris muscle raising the left part of the upper lip.
- Right Levitator Superior - Contraction of the right levator labii superioris muscle raising the right part of the upper lip.
- Left Levitator Nasi - Contraction of the left levator labii superioris alaeque nasi muscle raising the left part of the upper lip as well as the wing of the left nostril.
- Right Levitator Nasi - Contraction of the right levator labii superioris alaeque nasi muscle raising the right part of the upper lip as well as the wing of the right nostril.
- Depressor Inferious - Contraction of both depressor inferious muscles depressing the lower lip.

- Depressor Oris - Contraction of both depressor anguli oris muscles drawing the corners of the mouth downward and medial-ward.
- Mentalis - Contraction of both mentalis muscles raising and protruding the lower lip, a paired muscle but treated as a single muscle.
- Left Buccinator - Contraction of the left buccinator muscle retracting the left corner of mouth and keeping cheeks taut against teeth.
- Right Buccinator - Contraction of the right buccinator muscle retracting the right corner of mouth and keeping cheeks taut against teeth.
- Incisive Superior - Contraction of both upper incisive muscles pulling the upper lip in towards teeth.
- Incisive Inferior - Contraction of both lower incisive muscles pulling the lower lip in towards teeth.

An added benefit of using a muscle based parameterization is that these muscles also effect other parts of the face and can be used to deform these other parts. Examples are nose wrinkling, platysma affecting the neck, mentalis affecting the chin, the zygomaticus affecting the lower eyelid, and soforth. As well, when the muscles contract they bulge, which affects the surface of the face. Higher level parameterizations that use the basic parameters can be built to allow for easier use by an animator or director, such as smile, sadness, etc.

5 Implementation

We represent the lips as a B-spline surface with a 16x9 control grid. The parameters itemized above are mapped to changes in the positions of the control grid vertices. In this way, the definition of the lips can be replaced by another model as long as the parameter values are mapped correctly to changes in shape of the underlying model.

The geometry contains all of the vermillion zone (the red area of the lips) as well as the part of the mucous membrane that covers the lips internally. The geometry also contains a little extra of the mucous membrane to avoid observing an edge when looking at the lips from the outside. Figure 2 shows the control points of the lip model along with a polygonalization of the B-spline surface.

All of the muscles except the orbicularis oris are treated as a vector displacement acting upon its insertion points. The orbicularis oris constricts the shape of the lips into an oval while also extruding them. The parameters for the jaw articulate a virtual mandible based on the three jaw parameters and its resulting transform is used to move the lower lips. The lower lip is rotated outward with the opening of the jaw as well.

For each control point we calculate its position based on the parameters by the following:

$$p_i' = \hat{p}_i + L_i + J_i$$

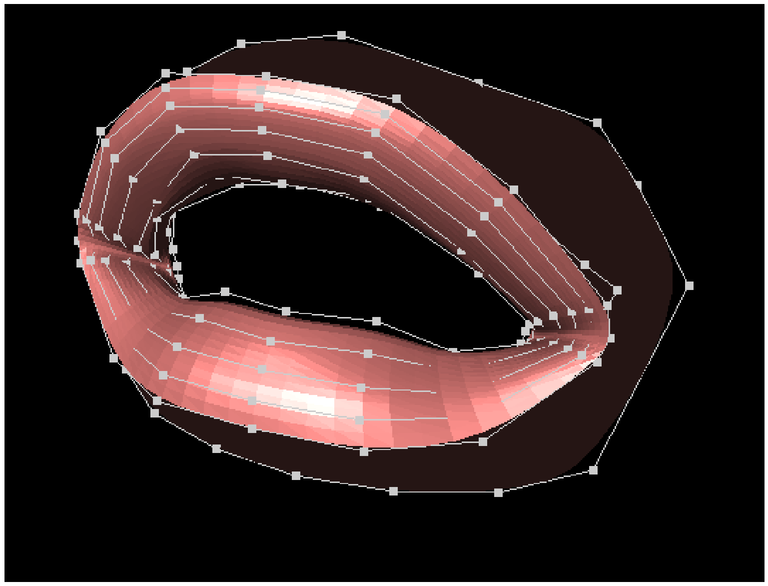


Figure 2: The B-spline control mesh and the surface used for the geometry of our lip model.

where $\hat{p}_i$ is the starting value for control point i. L_i is the sum of the displacements from the linear muscles and J_i is the contribution of the jaw.

$$L_i = \sum_{j=0}^{m} \rho_j M_j \delta_{ij}$$

where m is the number of linear muscles, $\delta_{ij} = 1$ if muscle j inserts into p_i and is 0 otherwise, m_j is a vector representing the maximum displacement caused by muscle j, and ρ_j is the parameter value for muscle j.

J_i is the movement of the lips due to the articulation of the mandible and it is calculated as:

$$J_i = J_{open} + J_{in} + J_{side} + LD\alpha_i + Open\gamma_i$$

where J_{open} is the effect of opening the mouth using the mandible, and is a rotation about the axis running through the condyles. J_{in} is the movement of the mandible in or out and J_{side} is the lateral movement of the jaw. LD is the motion vector for the lower lip and α_i is a constant which represents how much pulling the lower lip down will affect the upper lip point. $Open$ is how much the mouth is being open (which tightens the lips) and γ_i is how much opening the mouth will pull control point p_i towards the center of the mouth. $\alpha_i = \gamma_i = 0$ for the lower lip points.

The orbicularis oris pulls the mouth shut like a draw string. The result of its contraction is dependent on the mandible and contraction of the other muscles.

Combining the displacements of the other muscles with those of the orbicularis oris are difficult, so instead we calculate the effect of the other muscles first then use that to calculate the effect of the orbicularis oris. The function O_i calculates the motion resulting from contraction of the orbicularis oris on control point i. The final location of control point p_i is

$$p_i = O_i(p'_i)$$

where

$$O_i(p) = o(\theta_i + e_i(p) + \chi_i)$$

where o is the value of the orbicularis oris parameter, θ is the maximum rotation due to the puckering of the lips (we use 20°, remember the upper and lower lips rotate in opposite directions), and χ_i is the maximum extrusion from contraction of the orbicularis oris. $e_i(p)$ calculates an ellipse shape for the mouth and returns a motion vector for moving the control point p to a point on the ellipse created when contracting the orbicularis oris.

An option would have been to calculate the forces of each muscle and using a Newtonian physics model, numerically solve the ODEs to find the new locations of the control points. This would have allowed us to constrain the lip shape using springs, but we would have had to solve the ODEs. We wanted a closed-form solution that would avoid the rubbery look of spring-based systems. And using implicit integration would add unnecessary complexity.

5.1 Grafting

Grafting of the lip model geometry onto the input face geometry is done interactively. First an interactive tool is used to align the lip model with the input geometry. The tool allows the lip model to be shaped like the input lips in a neutral position. A convex outline surrounding the lips of the input geometry is created. After applying a cylindrical projection, all vertices, and thus all triangles, inside the convex hull are removed thereby removing the input lips. The fitted lip model is polygonalized and the outline of the lip model is triangulated along with the remaining input geometry using a delauney traingulation method [30]. The new triangles are added, along with the lip model, to the input facial geometry, effectively replacing the input lips with the lip model geometry.

6 Rendering

In order to create realism, the rendering of the lips is important. A common method to improve realism is the use of texture maps. The same problems associated with gathering the geometry of the lips also exist for gathering color information. Incomplete texture information will leave visible artifacts. We could use methods to warp what texture information is obtained, but there is no clear cut way to do this. This would also exacerbate the problems associated

Figure 3: Rendering of the lip model using our custom shaders. The left image is of wrinkled dry lips, and the right of wrinkled wet lips.

with texture maps, such as limited resolution and lighting inherent in texture acquisition.

We instead choose a different approach. We use a procedural texture shader to increase realism. Beside color information, we can also add surface detail with a bump shader. We implement our shaders using Pixar's renderman language and render it with prman. Figure 3 shows examples of wrinkled lips, both dry and wet. This method only works for offline generation of animations since it is too slow for our real-time version, where we choose a single color for the lips.

Lips are covered with very thin skin that tends to wrinkly easily. Besides the constant fine to medium wrinkles, when the lips are compressed (as in a pucker) there are large undulations of the surface. We currently ignore the finer wrinkles and instead concentrate on the larger wave-like wrinkles created during compression. Wrinkles are implemented as a bump shader.

Another shader determines the color of the lips. We can simulate natural lip colors as well as lipsticks and lipgloss. When the lips are licked, this results in differing thicknesses of saliva across the lips. We model this affect by creating a second layer, using a noise function, which represents the wetness pattern. This pattern is then mixed with the current lip color to increase the specular component. Lipstick and lipgloss are implemented as a uniform color change across the lips with transparency and glossiness components controlling matte versus glossy. Flecked lipsticks are modeled by adding a flecked silver pattern to the lipstick color.

7 Animation

Our parametric lip model can be animated in any standard method such as procedurally modeling the lip movement using forces and masses [31], a finite element approach [32] or keyframing [23] [24] [28] .

In the most common method, keyframing, visemes (visually distinct units of speech) are identified, with the number of visemes used differing widely. Commonly, phonemes (audibly distinctive speech units) are mapped into visemes using a many-to-one relation. However, a phoneme does not truly map into a static position of the vocal tract, but rather a dynamic shaping of the vocal tract.

Once the visemes have been identified they must be converted into a deformation of the model, and the transitions between visemes must be found. Unfortunately, it isn't quite that simple since during speech the same phoneme does not always visually look the same but instead depends on the phonemes before and after. This effect, called coarticulation, is a byproduct of the laws of physics and human anatomy. The vocal tract parts do not move and stop instantaneously so we must anticipate or lag behind, blurring the lines between phonemes. For example, when saying "how" the lips are rounded in anticipation of the /ow/ during production of the /h/, but when speaking "hat" there is no rounding. Coarticulation has been tackled with look ahead [29], triphones [16], nonlinear interpolation and masses [31] and using a coarticulation model such as the Lofqvist model [11] . In addition to coarticulation affects there are differences due to prosody (stress and intonation).

In our TTAVS system [23] we use keyframing. Text is input to Festival [6], which converts the text into phonemes. The phonemes are sent to MBROLA [1] to generate a waveform. The phonemes are also sent to the viseme generator, which produces a series of visemes (keyframes) from the phonemes. The viseme specifies the parameters for the lips, tongue and jaw. The facial model then takes the visemes and the waveform and generates a synchronized animation. The waveform is simply a sound track, and using t, the time from the beginning of the waveform, along with the visemes, the facial model is deformed to produce the correct shape that corresponds to the audio.

The facial model parameters associated with each phoneme are determined, thus creating a viseme and the definition of the Festival voice is modified to contain this extra information. We do this by interactively setting the facial model to the keyframe position for each phoneme. When text is parsed into phonemes, it is also parsed into visemes with the same timing as the phonemes that make up the waveform. Playing the waveform and using the time t to interpolate the visemes, lip synchronization is achieved. Our system currently uses linear interpolation without a coarticulation model; however, we are actively researching coarticulation methods.

8 Results

We have successfully incorporated this lip model into the facial model used by our TTAVS system. Our TTAVS system creates animations from text creating a stream of visemes, or keyframes, to be interpolated between. Figure 4 shows our lip model grafted onto an input geometry and displayed in our real-time

Figure 4: The lip model grafted onto face geometry. This is from our real-time TTAVS system using OpenInventor to render the frames.

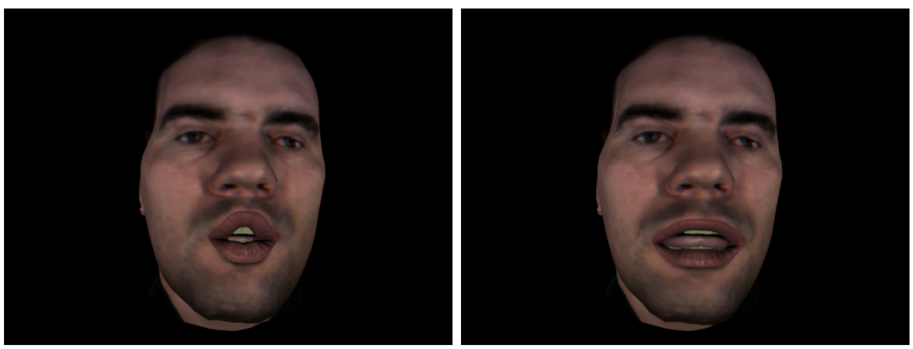


Figure 5: Frames from an animation rendered using our lip shaders.

system. Flat shading is used to more easily see the graft.

Figure 5 made up of frames from an off-line rendering using our rendering process for the lips. With our rendering technique we can achieve wrinkled and wet lips for increased realism. Figure 6 depicts frames from an off-line rendering and demonstrate motion blur of the lips, which can move extremely fast during speech.

Figure 7 shows closeups of the mouth area from our TTAVS system. Figure 7a is the viseme /aw/, while figure 7b is the viseme /aw/ while also activating the zygomaticus major muscle creating a happy /aw/. Figure 7c is a half smile, created by activating only the right zygomaticus major.

9 Conclusions and Future Work

Our anatomically-based lip model improves our ability to create realistic speech-synchronized facial animation with more realistic deformations of the lips. Be-

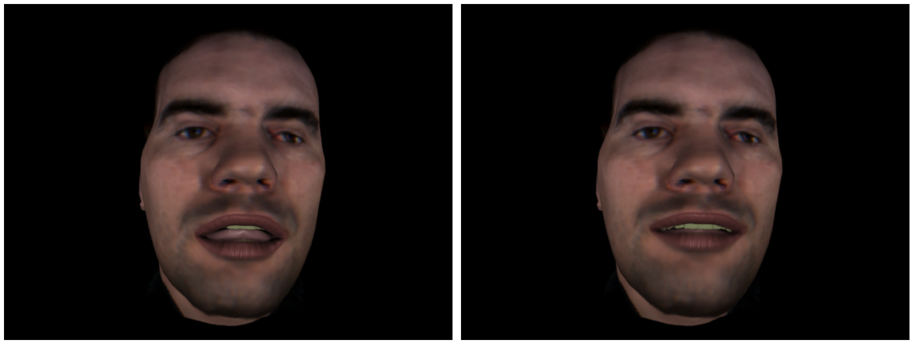
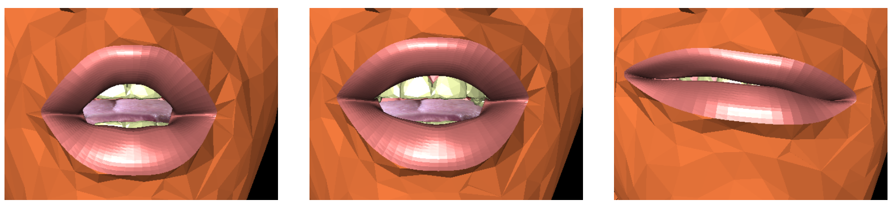


Figure 6: Frames from an animation rendered using our lip shaders.



(a) (b) (c)

Figure 7: /aw/, a happy /aw/ and a halfsmile

cause it is muscle-based, the effects of contraction of the mucles that affect the lips on other parts of the face are more easily calculated. Our lip model has both internal and external lip geometry, and by replacing the input lip geometry with the lip model's geometry we are guaranteed to have the internal geometry. The internal geometry is often missing, especially when the input geometry is acquired via a laser scan of the subject. This internal goemetry is important to have when the mouth opens to avoid loss of realism. With our generic lip model that is fitted to the subject, we also do not need to redefine the insertion of the muscles for each new subject.

Our lip model is a step in the right direction, but a lot of research remains to be done. Our lip model is muscle-based and allows for realistic deformations, but how to control the lip model is still in question. The lips help shape the vocal tract to produce the desired sound. With computer generated speech the lips are not needed for that role. However, a human audience is expecting the lips to perform that function and is highly sensitive to the position of the lips. A computer generated animation with speech must learn how to give the lips appropriate motion. Our simple method of using visemes and key positions in the speech is not at all adequate. For one, visemes actually take a segment of speech, which involves moving parts of the vocal tract, and reduces that segment to a single static position. Also, these static positions are used reqardless of the neighboring visemes. However, the neighboring visemes play an important part in shaping the current viseme, an effect know as coarticulation.

Our current focus has been on the motion of the lips due to muscle contractions, however, we also need to consider deformations due to collisions between the lips and other parts of the face. The lips must flow around the teeth and not penetrate them. Further, when the tongue presses against the lips for creating sounds or when wetting them, there is a slight deformation that is needed to improve realism. Finally, when the upper and lower lips come into contact with each other there are subtle changes that need to be shown. However, these deformations can be done without collision detection between the lips. And the spacial relationship between the uppor and lower lips makes interpenetration hard to notice.

Because the lips don't have a concept state, that is they don't know what came before, certain positions cannot be distinguished without further informaiton. For example, to rotate the lower lip outward into a pout the lower lip is pushed upward toward the upper lip which is tensed, causing the lower lip to slide over the upper lip and outward. However, if the upper lip is not tensed it will be pushed upward by the lower lip. So for these two final positions they can have the same parameters with two distinctly different results.

Our rendering technique gives us improved realism by allowing control of the surface detail as well as lighting, particularly highlights due to wetness of the lips. Unfortunately, lip wetness is currently applied to the entire lip as we do not have a way to specify to the shader how far the tongue has moved across the lip, how high it has gone on the lip as it moved, and in what direction it is moving. Changing this parameter to a function which tracks the aforementioned state changes and leaves a trail of saliva behind would allow better control. We would also like to be able to represent chapped lips as well as finer wrinkles. A better model of the mucous membrane is also desirable. This section of the lip is constantly moist and has significantly fewer wrinkles than the external section. The rima oris (point of contact of the lips during closure) in our model is simply a line in texture space, but in reality should be a curve over which there is a smooth transition between the two sections.

10 Acknowledgments

This work was partially funded by a grant from Texas Instruments.

References

[1] The MBROLA project. http://www.tcts.fpms.ac.be/synthesis/.

[2] Ali Adjoudani. Élaboration d'un modèle de lèvres 3d pour animation en temps réel. Master's thesis, Mémoire de D.E.A. Signal–Image–Parole, Institut National Polytechnique, Grenoble, France, 1993.

[3] Hugh E. Bateman and Robert M. Mason. *Applied Anatomy and Physiology of the Speech and Hearing Mechanism*. Charles C. Thomas, Springfield, IL., 1984.

[4] Philippe Bergeron. 3-d character animation on the symbolics system. In *SIGGRAPH '87 course notes: 3-D Character Animation by Computer*. jul 1987.

[5] Philippe Bergeron and Pierre Lachapelle. Controlling facial expressions and body movements in the computer-generated animated short 'tony de peltrie'. In *SIGGRAPH '85 Advanced Computer Animation seminar notes*, pages 1–19. jul 1985.

[6] Alan W Black, Paul Taylor, Richard Caley, and Rob Clark. The festival speech synthesis system. http://www.cstr.ed.ac.uk/projects/festival/.

[7] D. W. Boston. Synthetic facial communication. *British Journal of Audiology*, 7:95–101, 1973.

[8] Richard W. Brand and Donald E. Isselhard. *Anatomy of orofacial structures*. C.V. Mosby, St. Louis, 1982.

[9] N. M. Brooke. Development of a video speech synthesiser. In *Speech communication, recent developments in audiology, noise in industry : Autumn conference 1979 : Windermere Hydro Hotel*, pages 41–44, Windermere Hydro Hotel, 1979. Institute of Acoustics.

[10] N. M. Brooke and Quentin Summerfield. Analysis, synthesis and perception of visible articulatory movements. *Journal of Phonetics*, 11(1):63–76, jan 1983.

[11] Michael Cohen and Dominic Massaro. *Modeling coarticulation in synthetic visual speech*, pages 139–156. Springer-Verlag, Tokyo, 1993.

[12] David Ross Dickson and Wilma M. Maue. *Human Vocal Anatomy*. Charles C Thomas, Springfield, Illinois, 1970.

[13] Norman P. Erber. Real-time synthesis of optical lip patterns from vowel sounds. *Journal of the Acoustical Society of America*, 62:S76, 1977.

[14] Norman P. Erber and Carol Lee De Filippo. Voice/mouth synthesis and tactual/visual perception of /pa, ba, ma/. *Journal of the Acoustical Society of America*, 64:1015–1019, 1978.

[15] Norman P. Erber, Richard L. Sachs, and Carol Lee De Filippo. Optical synthesis of articulatory images for lipreading evaluation and instruction. In D. L. McPhearson, editor, *Advances in Prosthetic Devices for the Deaf: A Technical Workshop*, pages 228–231, Rochester, NY: NTID, 1979.

[16] Tony Ezzat and Tomaso Poggio. Miketalk: A talking facial display based on morphing visemes. In *Computer Animation '98*, pages 96–102, Philadelphia, University of Pennsylvania, jun 1998. IEEE Computer Society.

[17] Victoria Fromkin. Lip positions in american english vowels. *Language and Speech*, 7:215–225, 1964.

[18] Marie-Paule Gascuel. An implicit formulation for precise contact modeling between flexible solids. In *SIGGRAPH '93*, pages 313–320, aug 1993.

[19] Henry Gray. *Anatomy, Descriptive and Surgical.* Gramercy Books, New York, 1977.

[20] Thierry Guiard-Marigny. Animation en temps réel d'un modèle paramétrique de lèvres. Master's thesis, Mémoire de D.E.A. Signal–Image–Parole, Institut National Polytechnique, Grenoble, France, 1992.

[21] Thierry Guiard-Marigny, Ali Adjoudani, and Christian Benoit. A 3-d model of the lips for visual speech synthesis. *Proceedings of the Second ESCA/IEEE Workshop*, September 1994.

[22] Thierry Guiard-Marigny, Nicolas Tsingos, Ali Adjoudani, Christian Benoit, and Marie-Paule Gascuel. 3d models of the lips for realistic speech animation. In *Computer Graphic '96*, Geneve, 1996.

[23] Scott A. King and Richard E. Parent. Talkinghead: A text-to-audiovisual-speech system. Technical Report OSU-CISRC-2/80-TR05, Computer and Information Science, Ohio State University, 2000.

[24] Jeff Kleiser. A fast, efficient, accurate way to represent the human face. In Frederic I. Parke, editor, *SIGGRAPH '89 Course Notes 22: State of the Art in Facial Animation*, pages 36–40. jul 1989.

[25] A. A. Montgomery. Development of a model for generating synthetic animated lip shapes. *Journal of the Acoustical Society of America*, 68:S58(A), 1980.

[26] Ramond J. Nagle and Victor H. Sears. *Denture Prosthetics Complete Dentures.* The C. V Mosby Company, Saint Louis, 1962.

[27] Frederic I. Parke. *A parametric model for human faces.* PhD thesis, University of Utah, Salt Lake City, Utah, dec 1974.

[28] Frederic I. Parke. A model for human faces that allows speech synchronized animation. *Journal of Computers and Graphics*, 1(1):1–4, 1975.

[29] J A Provine and L T Bruton. Lip synchronization in 3-d model based coding for video-conferencing. In *Proc. of the IEEE Int. Symposium on Circuits and Systems*, pages 453–456, Seattle, May 1995.

[30] Jonathan Richard Shewchuk. Triangle. http://www.cs.cmu.edu/~quake/triangle.html.

[31] Keith Waters and Thomas M. Levergood. Decface: An automatic lip-synchronization algorithm for synthetic faces. Technical Report CRL 93/4, Sep 1993.

[32] Reiner F. Wilhelms-Tricarico and Joseph S. Perkell. *Biomechanical and Physiological Based Speech Modeling*, pages 221–233. Springer, 1997.