



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Conceptos de modelado y  
animación digital y su  
reutilización

T E S I S

QUE PARA OBTENER EL TÍTULO DE:  
LICENCIADO EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A :

CARLOS ALBERTO GARCÍA MARTÍNEZ

DIRECTOR DE TESIS

M. EN C. ALEJANDRO AGUILAR SIERRA



2010



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Sinodales:

Propietario: Mat. María Concepción Ana Luisa Solís González Cosío

Propietario: Dr. Jorge Alberto Marquez Flores

Propietario: M. en C.C. Alejandro Aguilar Sierra

Suplente: M. en C. Manuel Cristobal López Michelone

Suplente: M. en C. María de Lourdes Guerrero Zarco

# Agradecimientos

Al igual que la mayoría de los trabajos el presente no fue realizado solo por una persona, es cierto que solo se ve el nombre de dos personas en la carátula y el presente trabajo no es la diferencia. Diversos autores no podrían hacerlo solos; de la misma manera yo tampoco pude haberlo hecho solo, es por eso que quisiera agradecer a todas las personas que estuvieron involucradas en el desarrollo de esta tesis.

Me gustaría agradecer a los Maestros Alejandro Aguilar y René Garduño por el apoyo que me brindaron en diferentes áreas, pero principalmente por haberme brindado la oportunidad de desarrollar el presente trabajo y por ayudarme a conseguir una beca del proyecto IN-501408.

Gracias a la Doctora Verónica Teixeira, a la M. en C. María de Lourdes Guerrero Zarco y por último al Doctor Pablo Barreda por la aportación del material que necesitaba para comprender el tema expuesto.

Por otro lado me gustaría agradecer a Ricardo González por los diseños de los personajes y a Ben Dansie por los modelos que fueron utilizados para comprobar la eficacia del sistema.

Finalmente quisiera agradecer a mi familia y a todos mis amigos, que siempre han estado apoyándome.

# Resumen

CGI (Computer-generated imagery o Imágenes generadas por computadora) es hoy en día una de las herramientas más utilizadas para la visualización/simulación de fenómenos científicos, pero también puede ser un método para el entretenimiento, usado principalmente en largometrajes/cortometrajes y videojuegos. Las imágenes generadas por computadora han cambiado por completo los efectos especiales, debido al increíble realismo que puede ofrecer tanto en sus ambientes como por sus animaciones y en especial el motivo de este trabajo: La animación facial.

Aunque la animación corporal puede resultar difícil (dependiendo de lo que se desee), la animación facial puede resultar aún más difícil, debido a la complejidad de simular las emociones, además los rostros pueden variar enormemente. Para lograr una excelente animación se necesita tener un *rig* facial que sea fácil de usar y le brinde a los animadores los elementos necesarios para concluir dicha tarea, sin embargo, crear un *rig* manualmente es una tarea lenta para cualquier persona.

El sistema que se presentará facilita el proceso a cualquier animador o modelador que desee realizar el *rig* facial. En términos concretos el sistema traslada de forma rápida, precisa y con la misma calidad los *blend shapes* (que forman parte de un *rig*) de un modelo hacia otro (sin importar la forma o apariencia del modelo).

Para comprobar la eficacia del sistema, se modelaron a mano las expresiones de los modelos a los que se les desee transferir los *blend shapes* y se compararon con las expresiones que genera el sistema.

# Índice general

<b>1. Introducción</b>	<b>5</b>
1.1. Esquema . . . . .	5
1.2. Comunicación por medio de las expresiones faciales . . . . .	6
1.3. La Anatomía del rostro humano . . . . .	6
1.3.1. Anatomía del esqueleto . . . . .	7
1.3.2. Anatomía muscular . . . . .	7
1.4. Aplicaciones del estudio del rostro . . . . .	11
1.4.1. FACS y su futuro . . . . .	12
1.5. Modelos computacionales . . . . .	13
1.6. <i>Rigging</i> . . . . .	14
1.7. Modelado y animación facial . . . . .	17
1.7.1. Bases para la sincronización de labios . . . . .	17
1.7.2. Expresiones de la parte superior del rostro . . . . .	18
1.7.3. <i>Landmarking</i> . . . . .	20
1.8. Cuello de botella . . . . .	21
1.9. Objetivo . . . . .	23
<b>2. Estado del arte</b>	<b>24</b>
2.1. Antecedentes . . . . .	24
2.1.1. Interpolación de <i>keyframe</i> . . . . .	25
2.1.2. Deformación geométrica . . . . .	27
2.1.3. Animación facial con bases físicas . . . . .	28
2.1.4. Captura de movimiento . . . . .	30
2.1.5. Reasignación . . . . .	31
2.2. Parametrización facial . . . . .	33
2.2.1. ¿Cuál es una buena parametrización? . . . . .	34
2.2.2. Parametrización basada en los rasgos faciales . . . . .	34
2.2.3. Parametrización muscular . . . . .	35

2.2.4.	El MPEG-4 . . . . .	36
2.3.	Industria del entretenimiento: herramientas internas . . . . .	36
2.4.	Aplicaciones libres y patentadas . . . . .	37
2.5.	Futuro de la animación facial . . . . .	37
2.5.1.	El porvenir del <i>rig</i> facial . . . . .	38
<b>3.</b>	<b>Automatización del <i>rig</i> facial</b>	<b>39</b>
3.1.	El problema a considerar . . . . .	39
3.1.1.	La singularidad de los rostros . . . . .	39
3.1.2.	Dificultades técnicas del <i>rig</i> facial . . . . .	40
3.2.	Una solución: deformación de modelos faciales . . . . .	40
3.3.	Descripción del sistema . . . . .	41
3.3.1.	<i>Thin-plate spline</i> . . . . .	41
3.3.2.	Deformación de los rostros . . . . .	43
3.3.3.	Blender y Python . . . . .	44
3.4.	Flujo de trabajo . . . . .	47
<b>4.</b>	<b>Resultados</b>	<b>50</b>
4.1.	Aplicación del sistema a un rostro joven y a un rostro maduro	50
4.1.1.	Verificación del sistema . . . . .	51
4.1.2.	Comprobación del sistema . . . . .	51
4.1.3.	Transferencia y validación del <i>rig</i> al modelo masculino maduro . . . . .	52
4.2.	Observaciones . . . . .	55
<b>5.</b>	<b>Conclusión</b>	<b>57</b>
5.1.	El sistema . . . . .	57
5.2.	Mejoras del <i>rig</i> . . . . .	58
5.2.1.	Trabajo futuro . . . . .	59

# Capítulo 1

## Introducción

*“Ver es creer. Pero sentir es estar seguro” .*

Anónimo

La mayor parte de la gente es capaz de mostrar una amplia gama de emociones y pensamientos. Cualquiera persona es experta en el reconocimiento de las emociones a través de las expresiones faciales y se nos es fácil detectar algún error al percibir las emociones de un personaje virtual. Las expresiones faciales son complejas en diversas formas. Este capítulo muestra el por qué de su complejidad y algunas técnicas que ayudan a generar un mejor resultado en las animaciones.

### 1.1. Esquema

Este trabajo se integra de la siguiente manera:

El capítulo dos presenta algunos de los antecedentes que se han presentado a lo largo de 30 años con respecto a la animación facial, además lo que se espera a futuro.

El capítulo tres da la solución al problema que sufren las producciones de animación al momento de crear el *rig* para cada uno de los personajes que están involucrados. El capítulo presenta el algoritmo y el sistema, los cuales son la base para la solución.

El capítulo cuatro da los resultados obtenidos del presente trabajo con diferentes tipos de topología.

El capítulo cinco presenta las conclusiones a las que se llegó y del por qué y cómo se puede mejorar aún más el sistema.

## 1.2. Comunicación por medio de las expresiones faciales

El antropólogo Ray Birdwristell [Bir70] concluyó que las personas aprenden a usar las expresiones faciales para transmitir información y que el tipo de información depende fuertemente de cada cultura, sin embargo, en la década de mil novecientos setenta las investigaciones [Ekm73, Ekm72] mostraron convincentemente que existen constantes en cada una de las personas que habitan el globo y que en cada una de ellas se puede apreciar la íntima relación entre las emociones y las expresiones faciales.

La expresión facial es, junto con la mirada, el medio más rico e importante para expresar emociones y estados de ánimo como felicidad o tristeza, estas emociones pueden mostrarse al sonreír o fruncir, respectivamente. Podemos utilizar las expresiones faciales en respuesta a la información que nos proporcionan otras personas, por lo que es factible ampliar o cerrar la comprensión dependiendo lo que nos comunican. Al entablar una comunicación hacemos juicios sobre la persona y otros rasgos de la gente, todo lo anterior en función de lo que vemos en sus rostros, permitiéndonos realizar juicios sobre el carácter, personalidad, inteligencia, temperamento, hábitos personales, capacidad de trabajo, etc.

Los humanos pueden adoptar alguna expresión facial a voluntad o incluso intensificarla para resaltar aún más la emoción; sin embargo, debido a que las expresiones están íntimamente ligadas a las emociones, regularmente son involuntarias e incluso pueden ser sumamente difíciles de evadir ante algún estímulo. La relación puede incluso presentarse en la otra dirección; se ha observado que asumir voluntariamente alguna expresión puede causar la emoción asociada.

La gama de expresiones es muy amplia, pero hay un número limitado (Tabla 1.1) de emociones que la mayoría de nosotros puede reconocer con cierta fiabilidad.

## 1.3. La Anatomía del rostro humano

La dinámica del rostro nos brinda los medios para entender su comportamiento e incluso sus funciones; no sólo puede revelar parte de la personalidad de la persona, sino también sus sentimientos.

Cuadro 1.1: Gama de Expresiones Faciales.

Expresión	Significado
Ira	Es la pasión que causa indignación y enfado. Se suele enseñar los dientes. fruncir el ceño o boca y levantar la voz.
Miedo	Es la respuesta emocional ante amenazas y peligros. Se demuestra con los ojos, levantado el labio, etc.
Felicidad	Es el estado de la mente que se caracteriza por alegría. Se expresa con una sonrisa y con los ojos.
Tristeza	Esta emoción es provocada por un decaimiento moral. La forma de demostrarse puede ser por el llanto, las cejas pueden bajar sutilmente, al igual que las comisuras.
Sorpresa	Estado emocional, resultado de un evento inesperado Se expresa mediante las cejas, párpados y al bajar la mandíbula.

### 1.3.1. Anatomía del esqueleto

El cráneo humano está formado por ocho placas y por 14 huesos faciales, los cuales son la base para los tejidos suaves del rostro y cabeza; las formas y características del cráneo determinan las apariencias estéticas del rostro y proveen la base para las características de la fisonomía, por lo que en un mundo donde existen más de seis millardos de rostros, todos y cada uno de ellos son únicos ( a excepción de los gemelos). Los huesos de la cabeza están asociados con muchas más características (áreas de proceso, líneas, forámenes y concavidades); algunas de ellas afectan la fisonomía del rostro en conjunto, con otras variaciones en tamaño, grueso, locación y forma.

En [Cha96] la ciencia forense ha utilizado estas características para reconstruir la apariencia superficial del rostro; "Kennewick Man" (ver Figura 1.1).

La reconstrucción del cráneo ha resultado muy importante para comprender nuestra evolución y redefinir la historia; justo como lo hizo Kennewick Man, quien es un descendiente de una migración más antigua de Asia que la de los ancestros de los Indios Americanos.

### 1.3.2. Anatomía muscular

Al estar los músculos detrás de los cambios de apariencia, la anatomía muscular está directamente relacionada con las expresiones (Figura 1.2).

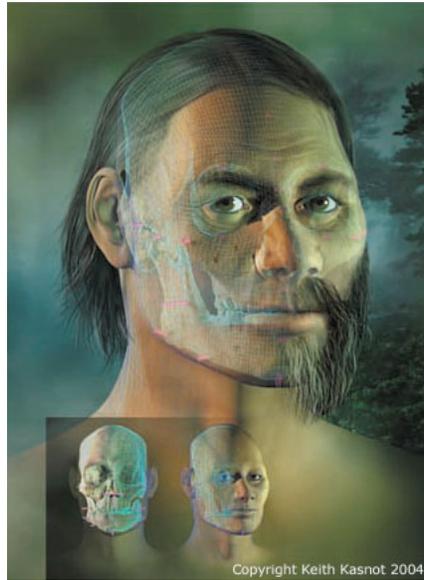


Figura 1.1: Kennewick Man. Construcción de un aborigen americano [Col08]

La Tabla 1.2 muestra la lista de los grupos más significativos que están relacionados con la apariencia y las expresiones.

A pesar de que se han agrupado los músculos, su número no ha sido definido por los científicos (no hay un acuerdo en común al categorizar los músculos), debido a que los músculos se enciman unos con otros, haciendo difícil identificarlos. Diferentes publicaciones han mostrado que se utilizan menos músculos para poder sonreír que para fruncir el ceño, pero han dado diferentes números para cada acción (vea Figura1.3).

Los músculos faciales o también llamados músculos miméticos cuentan con varias propiedades [Kel]:

1. Son un grupo de músculos estriados inervados por el nervio facial.
2. Una propiedad muy particular es que no tienen fascias o fundas fibrosas, en comparación con otros músculos.

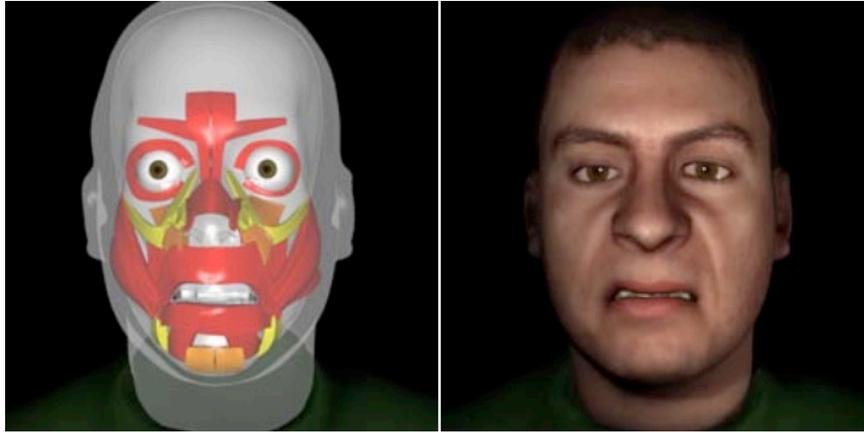


Figura 1.2: Los músculos y las expresiones [Sif05].

Cuadro 1.2: Grupos Faciales y su acción o expresión [Kel].

Grupos	Expresión Facial o Acción muscular
Orbicular de los labios	Estos músculos permiten que los labios se cierren y abran.
Borla, triangular y cuadrado de la barba.	Estos pequeños grupos son los encargados del sollozo.
Risorio	Este grupo ayuda a la creación de expresión feliz, es decir crea la sonrisa.
Cigomático mayor y menor.	El movimiento de estos músculos ayuda a crear una la sonrisa. Pero un movimiento hacia abajo demostraría una sonrisa despectiva.
Superciliar, piramidal y orbicular de los ojos	Grupo de músculos que rodean los ojos, los cuales son los responsables de cerrar los párpados, el uso de estos músculos puede intensificar el aburrimiento o el enojo.
Occipitio-Frontal	Estos grupos eleva la cejas, el uso de estos puede provocar una apariencia de sorpresa o relajación
Mirtiforme, transverso elevadores, canino y dilatador.	Grupos que se encuentran entre la nariz y la cejas es el responsable de crear una emoción tan fuerte como la ira.

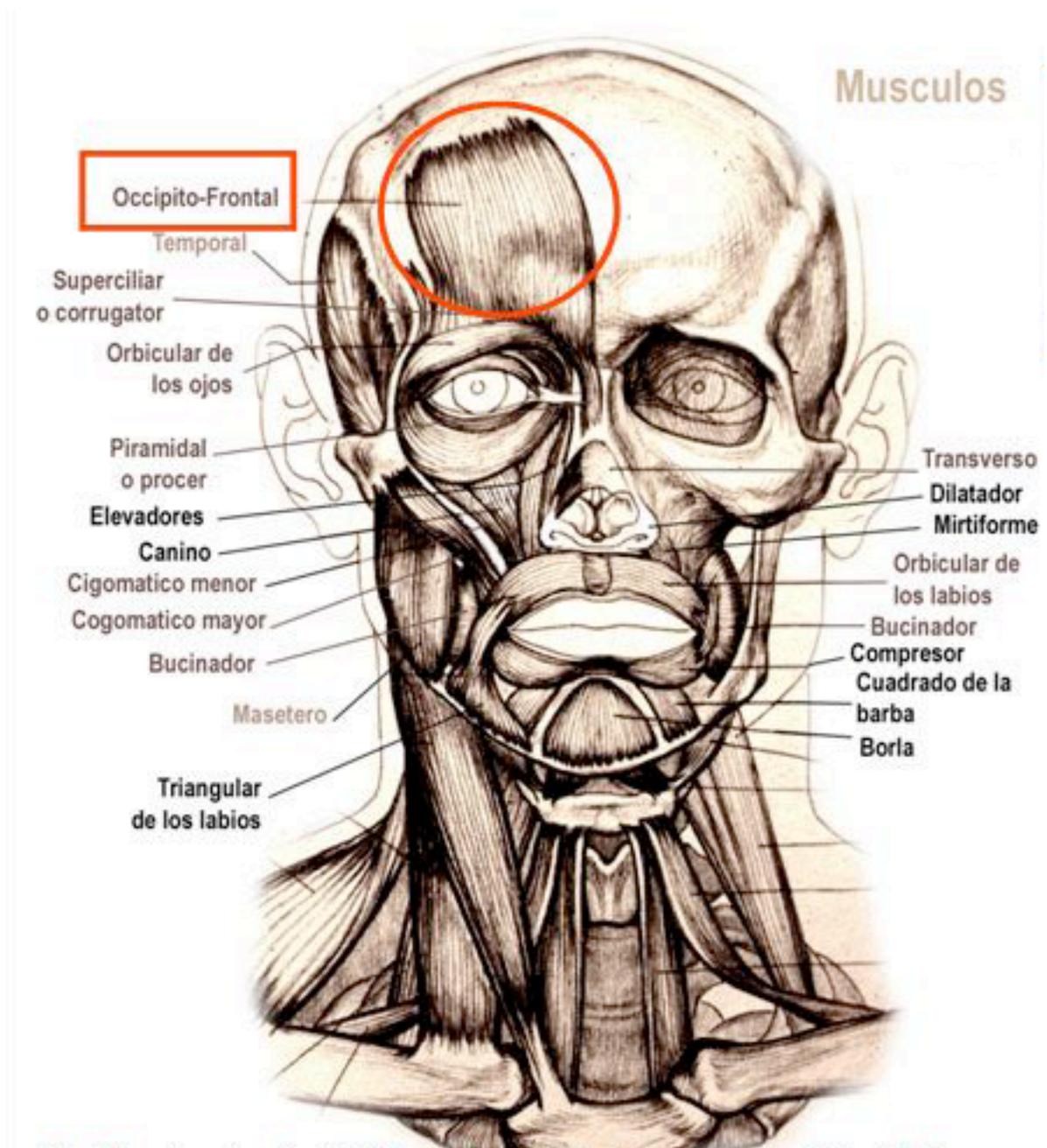


Figura 1.3: Grupos Musculares (imagen creada por Carlos Plascencia, Facultad BBAA, departamento de dibujo)

3. Los músculos se distribuyen entorno a los orificios faciales y en sentido longitudinal.
4. Excitabilidad: es la sensibilidad que presentan los músculos ante un estímulo nervioso que llega a la fibra muscular. También se pueden excitar por un pinchazo, una corriente eléctrica, un golpe, etc.
5. Elasticidad: es la capacidad que tienen los músculos de aumentar su longitud y recuperar su dimensión inicial.
6. Contractilidad: es el acortamiento del músculo sin que varíe su volumen. Cada fibra muscular se contrae del todo o no se contrae. Como el músculo está formado por muchas fibras, se contraerá más o menos según el número de fibras que lo integren.
7. Tonicidad: en reposo, los músculos no están relajados del todo. Siempre tienen cierto grado de contracción o tono muscular, responsable de que mantengamos la forma del cuerpo.

En el pequeño resumen se puede apreciar la complejidad de los músculos faciales humanos y su relación con el cráneo. Al día de hoy la creación de un modelo y la animación para cada uno de los músculos, sus propiedades y su relación con las expresiones, no es una tarea trivial (Figura 1.4 el modelo fue hecho por Ken, B. para su estudio de los musculos faciales<sup>1</sup> ).

## 1.4. Aplicaciones del estudio del rostro

Diversas herramientas han sido creadas para facilitar el conocimiento del rostro, las cuales permiten comprenderlo fácilmente como un sistema que brinda mensajes, ideas y emociones; su rol en la comunicación no verbal, etc. Algunas de estas herramientas reconstruyen un cráneo (reconstrucción forense) [Cha96]; algunas pueden identificar la identidad de las personas por medio de las características más significantes de su rostro (aplicaciones legales); algunas pueden envejecer los rostros (utilizada para la búsqueda de niños desaparecidos); avances en las investigaciones de la psicología, entre algunas otras.

---

<sup>1</sup><http://www.brilliant-creations.com>



Figura 1.4: Simulación de las estructuras que conforman el rostro humano (modelo hecho por Ken, B).

Un método que ha estado presente desde hace tres décadas es FACS (Facial Action Coding System), el cual es un sistema de codificación de acciones faciales que clasifica las expresiones del rostro humano en base a los músculos que las generan.

#### 1.4.1. FACS y su futuro

FACS fué desarrollado por Ekman y Friesen [Ekm78], convirtiéndose en el estándar de clasificación sistemática de las expresiones físicas de las emociones. FACS es un método versátil y muy utilizado que mide y describe los comportamientos faciales. La primer versión de FACS determina cómo es que la contracción de músculo facial (uno por uno y en combinación con otros músculos) transforma la apariencia del rostro.

Ekman y Friesen examinaron algunos videos que muestran el comportamiento facial, la finalidad fue poder identificar cómo es que los cambios en las contracciones musculares afectan al rostro, además de conocer la mejor forma de diferenciarlas. Al estudiar la anatomía, reproducir la apariencia y palpar sus rostros, asociaron los cambios en la apariencia con las acciones de los músculos que las producen. La meta fue crear medios confiables para que las personas entrenadas determinaran la categoría o categorías en que se ajusta cada comportamiento facial.

Las unidades de medición no fueron los músculos sino que se definieron Unidades de Acción (o AUs). La elección de estas unidades se debe a que para algunas formas del rostro, más de un músculo fue combinado dentro de una sola AU (los cambios en la apariencia que algún músculo genera, probablemente no podría distinguirse). Otra razón, es que los cambios en la apariencia que produce un solo músculo fueron separadas en dos o más AUs para representar acciones relativamente independientes de diferentes partes del músculo.

Un codificador de FACS observa una expresión y la descompone en AUs específicos que producen el movimiento. Los registros de cada expresión facial consisten en una lista de AUs que producen la expresión. Un detalle importante es que se puede reajustar la duración, la intensidad y la asimetría.

Los registros en FACS son sólo descriptivos y no proveen ninguna implicación acerca del significado del comportamiento. Existen varias alternativas que superan esta limitante, una de ellas es FACSAID (diccionario interpretativo de sentimientos del sistema de codificación de acciones faciales), proyecto que relaciona las expresiones faciales con su interpretación psicológica; por otro lado el programa gráfico CANDIDE<sup>2</sup> permite plantear artificialmente las expresiones faciales mediante el establecimiento de AUs.

El nuevo FACS fue desarrollado en el año 2002 por Paul Ekman, Wallace V. Friesen y Joseph C. Hager. En esta nueva versión la simultaneidad entre normas fueron removidas, se agregaron y eliminaron AUs; fue agregado un novedoso índice sobre definición de la intensidad.

FACS, hoy en día es una herramienta benéfica para los científicos, animadores CG, psicólogos y computólogos interesados en los programas de reconocimiento de patrones; el uso FACS en su trabajo profesional es necesario para saber los movimientos exactos que el rostro desempeña; los músculos que son parte de esos movimientos y la relación estrecha que existe con las emociones que un individuo puede ofrecer.

## 1.5. Modelos computacionales

Simular las expresiones se convierte en un objetivo atractivo (por el poder expresivo del rostro), pero difícil de alcanzar debido a la complejidad del desarrollo de modelos computacionales del rostro. Por otro lado existe una necesidad imperiosa para su desarrollo, no solo en la animación, sino también

---

<sup>2</sup>[www.bk.isy.liu.se/candide/](http://www.bk.isy.liu.se/candide/)

en teleconferencias de ancho de banda baja, cirugía plástica, cirugía asistida por computadora, hasta la investigación criminal, pero el primer paso es representar el rostro computacionalmente.

Una malla poligonal (polygon mesh) es una colección de vértices, aristas y caras que definen la forma de un poliedro. Las caras usualmente constan de triángulos, cuadriláteros u otros simples polígonos convexos, pero también puede estar compuesto de polígonos cóncavos más generales, o polígonos con agujeros (Figura 1.5). Puede observar por la Figura 1.5 una aplicación que se le da a esta malla poligonal es el modelado de la cabeza, más aún es el estándar para crear animaciones en las industrias de videojuegos y cine.

## 1.6. *Rigging*

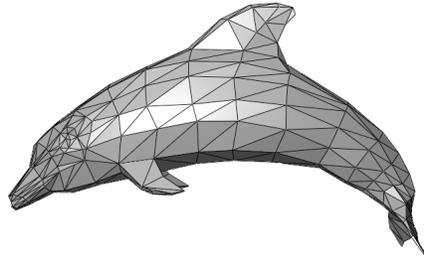
El proceso que transforma un modelo (malla) estático e inanimado en un personaje que se puede manipular por un animador y crear movilidad, se conoce como “*rigging*” [Kla06]. El resultado es una estructura (o *rig*) que puede ser manipulada por un conjunto de controladores. La Figura 1.6 muestra el proceso de *rigging* que está presente en cualquier producción.

Esta labor (sin ningún estándar al día de hoy) puede ser difícil, extensa e inclusive intensa hasta para cualquier artista experimentado; esto puede deberse a que los animadores necesitan características adicionales del *rig* y que no están presentes en paquetes de software, quizá los controles no son fáciles de usarse o difíciles de entender, entre algunas otras razones.

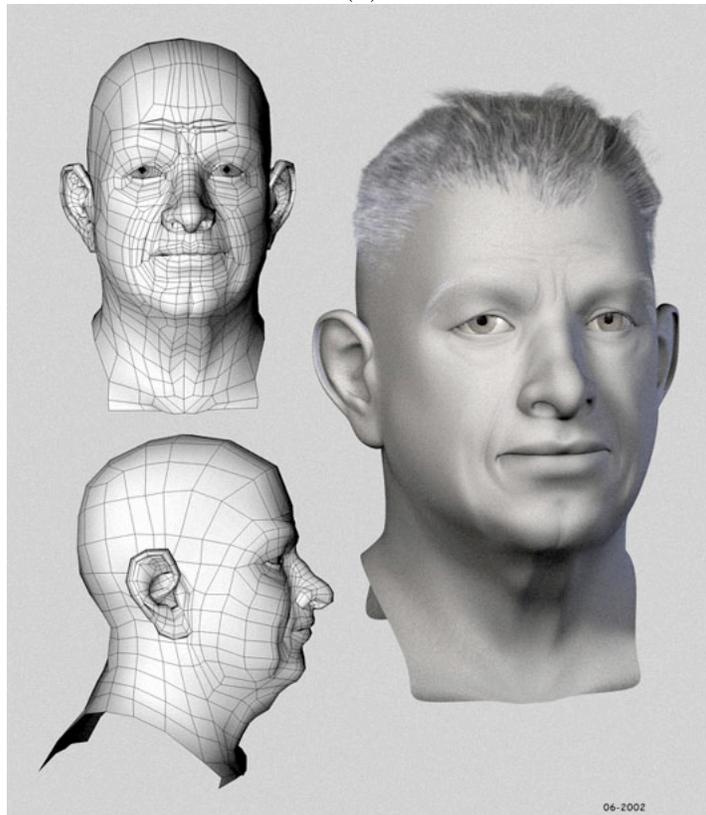
Las tres vertientes más comunes para crear el *rig* son: “*blend shapes*”, “*bone-driven*” o la combinación de ambas. Se define como *shape* cierta pose/expresión (de la topología inicial), entonces el resultado de una animación se debe de interpolar diferentes *shapes* en un periodo de tiempo (incluso pueden usarse más de dos *shapes* en un momento del tiempo, siempre y cuando el rostro resultante esté bien “combinado”), esta técnica se conoce como *blend shape* y es sumamente utilizada por los artistas. Por ejemplo interpolar el *shape* que representa el rostro al abrir la boca y la malla que representa la boca cerrada, crearía la ilusión del habla (Figura 1.7, modelo hecho por Ben Dansie<sup>3</sup>).

---

<sup>3</sup><http://www.bendansie.com/>



(a)



(b)

Figura 1.5: Ejemplos de mallas poligonales:(a) un modelo del delfín; (b) cabeza de un hombre maduro.

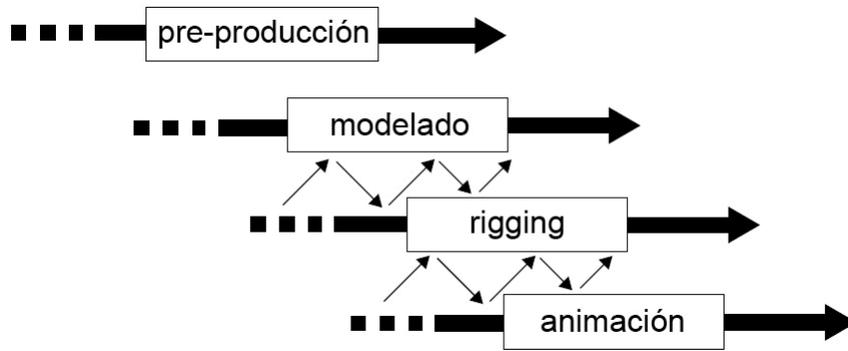


Figura 1.6: Un proceso tradicional de animación, los procesos de modelado, *rigging* y animación se trabajan en paralelo.



Figura 1.7: La forma al abrir la boca, expresión realizada por Carlos A..

Obtener el *rig* por medio de *bone-driven* consiste en utilizar una estructura ósea altamente articulada sobre el modelo facial. Este tipo de construcción permite tener movimientos mucho más pulidos, sin embargo, obtener un *rig* de esta manera requiere considerar que un vértice es afectado por algún hueso de la estructura ósea, por consecuencia incrementa su dificultad.

## 1.7. Modelado y animación facial

Jason Osipa [Jas02] menciona: “Al igual que la animación, la pantomima es un medio de comunicación de cuerpo completo que toma años en dominarse. Dominar la comunicación a través del rostro (así como las sutilezas de la actuación) puede ser una tarea larga y más difícil, incluso que el dominio de la pantomima”. Jason muestra otra forma de construir y animar un rostro con base en el significado de las expresiones, al descomponer de forma sencilla, paso por paso, cualquier expresión que se desea o requiera en el 99% de cualquier producción. La base de su método inicia al comprender lo importante de cualquier discurso e ignorar lo que no lo es y posteriormente construir los aspectos más complicados en el desarrollo de las expresiones.

### 1.7.1. Bases para la sincronización de labios

En lugar de modelar rostros especializados para cada fonema o expresión, sólo se modelan rostros más amplios en su aplicación, usando combinaciones para crear otros rostros más especializados; todos esos modelos se derivan de un modelo que puede realizar la mayoría de los movimientos que cualquier persona expresa, sin necesidad de modelar los cientos o miles de expresiones (*blend shapes*) que llegan a utilizarse en una animación.

Su análisis radica en que no es necesario recrear cuadro por cuadro el movimiento de cada letra o sílaba, lo único que se necesita es crear la ilusión del habla; por ejemplo las palabras mamá y papá se escuchan diferente, pero en cualquier conversación el movimiento de los labios se vería exactamente igual. El abrir/cerrar de la boca, además de su forma ancha/estrecha son las formas más importantes que realiza la boca. Los titiriteros dan la ilusión de habla a sus marionetas, el cerebro realiza la conexión entre el movimiento de la boca de la marioneta justamente cuando el actor hace el sonido. Existen ciertos sonidos que hacemos con la boca y que es absolutamente necesario representar visualmente, sin importar que: fonemas visuales o que Osipa llama

“visimes” . Estos sonidos sólo pueden lograrse con la boca y con la expresión bucal. Existen tres tipos de sonidos que podemos hacer, algunos sonidos son hechos por los labios, otros por la lengua y los restantes se realizan en la garganta y cuerdas vocales. En la animación los únicos sonidos principales son los que se realizan con los labios, pero en realidad son combinaciones de los tres ya mencionados.

Por último existen las formas relativas que dependen de los visimes que se encuentran a su alrededor. Por ejemplo los sonidos de OOs u OHs, no necesitan ser muy estrechos, sólo necesitan ser estrechos a comparación de los visimes que le anteceden y preceden.

Es destacable que cualquier palabra puede ser representada por algún *visime* (la Figura 1.8 muestra el total de los *visimes* bucales) o por alguna combinación de éstos y por los movimientos básicos de la boca, aunque se debe tener cuidado al mezclar correctamente los *shapes* para obtener animaciones precisas y creíbles.

### 1.7.2. Expresiones de la parte superior del rostro

El segundo apartado, Osipa centra su atención en las emociones que se expresan por medio de las cejas, ojos y párpados, concentrándose principalmente en los dos últimos debido a que en una conversación la atención se centra en los ojos, sin embargo, muestra en qué tipo de emociones la presencia del movimiento de las cejas ayuda a enfatizar las emociones, las cuales cuentan con dos movimientos básicos: abrir/cerrar y contraerse.

Es esencial hacer notar que las emociones siempre están presentes antes de cualquier diálogo (se piensa antes de actuar), por lo que comprender las sutilezas de las combinaciones entre cejas, ojos y párpados facilita la creación de emociones creíbles. Dichas sutilezas pueden observarse en las formas relativas que ofrecen los párpados con respecto al iris y la pupila; por ejemplo se puede observar que si los párpados se encuentran más abiertos dejando entrar más luz se puede mostrar un estado de alerta, caso contrario si se encuentran casi cerrados denotarían algo de sueño.

El último tema a destacar es lo que nos dicen los ojos, lo cual es bastante simple: indican el objeto al cual miramos (a pesar de que se cree que al mirar hacia abajo se puede mostrar cierta timidez o miedo). Sin embargo, no se debe confundir el hecho de mover los ojos y tener que dirigir la cabeza por completo al objeto que se desea mirar, en el caso anterior no se sabe

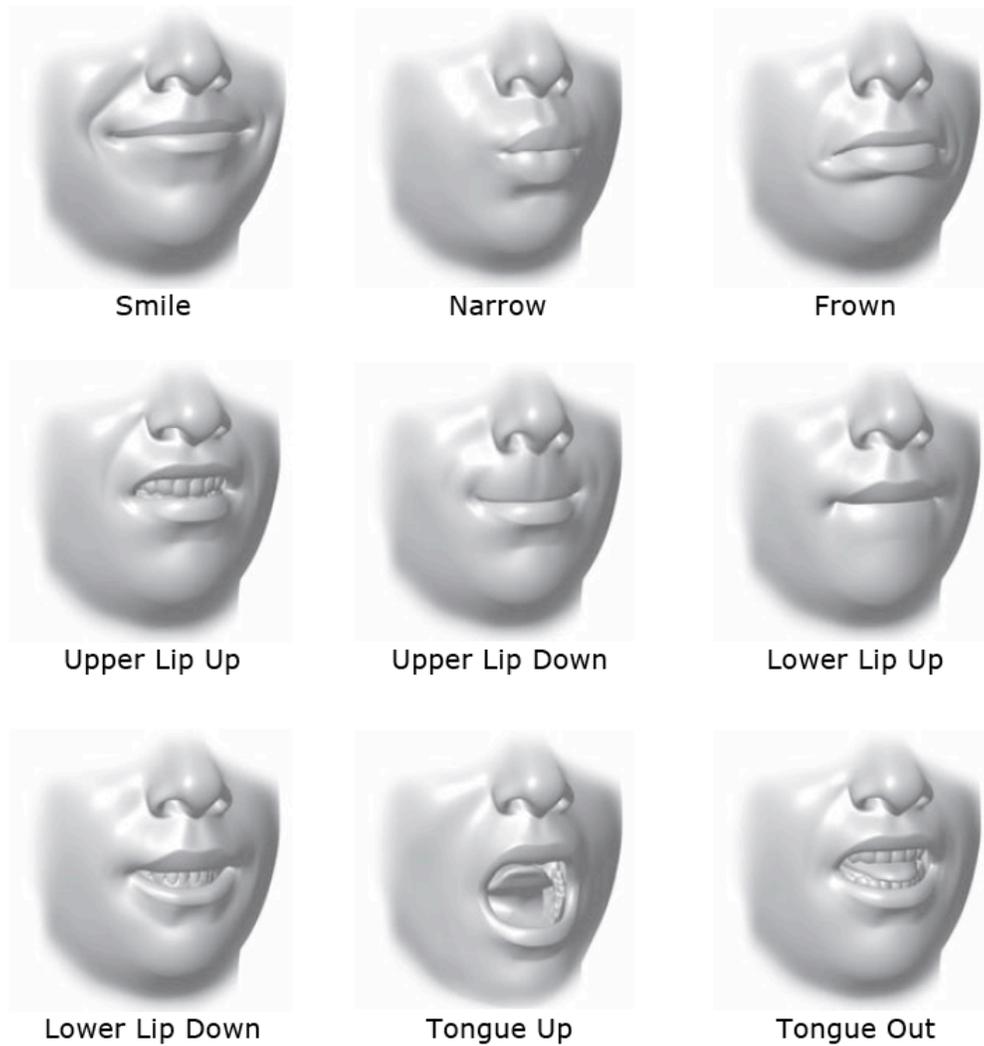


Figura 1.8: Visimes. Los nueve rostros que crean los fonemas visuales (Modelo y *visimes* hechos por J.Osipa [Jas02].)

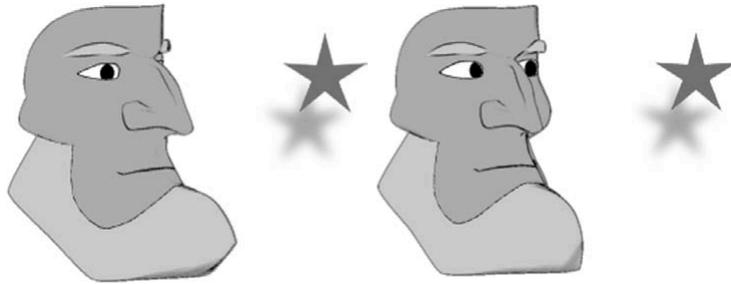


Figura 1.9: Diferencias entre dos líneas visuales; la primera es la correcta, pero la segunda es mucho más interesante y se sabe que es lo que observa el personaje [Jas02].

a ciencia cierta si realmente los ojos están observando lo que se pretende contemplar (la Figura 1.9 lo muestra perfectamente).

La idea, detrás de todo, es primeramente comunicar “algo” a la audiencia; por ejemplo en la figura 1.9 se aprecia que el rostro realmente le expresa algo al público.

### 1.7.3. *Landmarking*

Osipa define como “*landmarking*” a la apreciación del área que se observa o el lugar donde las emociones que podemos observar. Puede ser usado para identificar qué cambios globales se han hecho de forma incorrecta; por ejemplo si al crear una sonrisa se dejan fuera las arrugas, la Figura 1.10 explica de forma visual lo ya comentado.

Este proceso consta de observar alguna de las expresiones y registrar los cambios hechos al modelo inicial (regularmente es el rostro serio del modelo); la sonrisa es un excelente ejemplo porque se analiza el entorno, pero dejando de lado la boca misma, se aprecian cambios en las mejillas e incluso los párpados inferiores, características que son esenciales para poder expresar una verdadera sonrisa. Este mismo concepto puede aplicarse diferentes emociones.

Comprender las características de las expresiones faciales tiene una consecuencia muy útil: el poder intensificar las emociones. Observe la Figura



Figura 1.10: Sonrisa inusual [Jas02].

1.11, la característica del enojo es que las cejas estén lo más abajo posible, sin embargo si se inclina la cabeza hacia delante se crea la ilusión de que la distancia entre las cejas y ojos es más estrecha, por consecuencia se intensifica el enojo y el espectador lo podrá percibir más claramente.

Todos los conceptos anteriores tienen cierta complejidad, la cual radica en qué momento se deben realizar las acciones de los personajes y en cómo se deben de aplicar los conceptos (individualmente o en conjunto), no obstante el potencial que brindan al crear las emociones es enorme.

## 1.8. Cuello de botella

La complejidad de una animación facial radica en el siguiente proceso: la creación de una malla anatómicamente correcta del rostro, un *rig* adecuado y la creación de las expresiones que necesitan expresar los rostros para transmitir una emoción. Pero hay algo que se ha dejado implícito en dicho proceso: es necesario repetirlo para cada personaje que se necesite animar, ésta es la razón por la que es un cuello de botella en cualquier producción de animación.



Figura 1.11: Inclinar el rostro ayuda a intensificar el enojo de una persona. Modelo hecho por Ben Dansie y expresiones faciales hechas por Carlos A. García.

## 1.9. Objetivo

Esta tesis plantea reducir el tiempo de ese cuello de botella al reutilizar los *blend shapes* de algún personaje. El sistema propuesto requiere un modelo facial que cuente con un *rig* basado en *blend shapes*. Posteriormente, el sistema transfiere el *rig* hacia otro modelo facial (incluso si existen diferencias topológicas). Este sistema utiliza una función de deformación no lineal y una elección de puntos característicos sobre las dos mallas, por lo que el resultado es un sistema muy eficiente. Una vez concluido este método, el animador obtendrá la misma calidad en las expresiones faciales en todos los personajes a los cuales se les transfirió el *rig*.

# Capítulo 2

## Estado del arte

*“Cada rostro es único – en su aspecto, forma y proporciones, así como también en las formas específicas de articulación y expresión, tan características de un individuo” [Kah03].* La creación de rostros humanos animados por computadora ha sido un desafío desde principios de los años setentas, y diversas investigaciones han sido abordadas (como las deformaciones musculares y deformaciones de tejido suave), al igual que propuestas de animación facial, pero hoy en día no existe un sistema de propósito general que resuelva de forma satisfactoria las necesidades de los diferentes campos de aplicación, por lo que este capítulo describe los principales problemas relacionados con el *rig* facial y sus posibles soluciones a lo largo de 30 años.

### 2.1. Antecedentes

Diversas investigaciones relacionadas con movimientos faciales y expresiones han sido realizadas desde el milenio pasado (los siglos XVII, XIX y XX); John Bulwer [Bul49] creía que ciertos músculos faciales eran el puente de los deseos de cualquier mente; Charles Darwin [Dar02] planteaba que cualquier animal (incluido el hombre) expresaba el mismo estado emocional.

Frederic I. Parke es el pionero de las animaciones faciales generadas por computadora y se ha mantenido activo en su desarrollo desde 1971; su trabajo [Par72] provee los fundamentos y técnicas que se han convertido en el estándar de la industria. El modelo que propuso fue una malla de puntos en un espacio de tres dimensiones, la malla fue controlada por dos conjuntos de parámetros; el primero modificaba la malla formando nuevos rostros y

el segundo eran acciones faciales (estirar los labios y cerrar los ojos). Desde su primera edición han habido desarrollos significativos y una gran variedad de propuestas por parte de grupos académicos de investigación, la industria del cine y videojuegos; dichos grupos pretenden crear un sistema que: 1) cree animaciones realistas, 2) se ejecute en tiempo real, 3) que se lo más automatizable posible; y 4) que sea adaptable fácilmente a diferentes rostros.

Las investigaciones en el modelado y animación facial se encuentran catalogadas en dos grandes categorías (Figura 2.1):

- Manipulaciones geométricas
- Manipulaciones de imágenes

Ambas categorías contienen una variedad de diferentes propuestas (se recomienda el siguiente trabajo [Pig98] que recopila y explica de una forma más detallada dichas propuestas), nuevos enfoques utilizan una o más propuestas para obtener mejores resultados, debido a esto y la falta de distinción entre propuestas, establecer una clasificación exacta sería muy complicado, los siguientes secciones explican los métodos más importantes de las dos categorías.

### 2.1.1. Interpolación de *keyframe*

En animación un *key* es una asignación de un conjunto de valores a los atributos de un objeto –por ejemplo la rotación, escala, color, etc. – en un tiempo específico (frame), esta definición se le conoce como *keyframe*. Generar una transición fluida entre dos *keyframes* (en este caso especificar las expresiones faciales) se logra si se emplea una función de interpolación, la cual construye en un intervalo de tiempo el conjunto de valores que se necesitan para lograr la transición del *keyframe* inicial (el rostro serio) al final (la sonrisa, vea Figura 2.2), es decir, la interpolación genera los *keyframes* faltantes entre los dos *keyframes*.

Es cierto que la interpolación lineal es una idea muy intuitiva de la animación facial y comúnmente usada por su simplicidad (véase [Pig06] ), sin embargo, al ser el método más antiguo tiene varios puntos en su contra, su limitante principal: no produce una variedad realista de expresiones faciales.

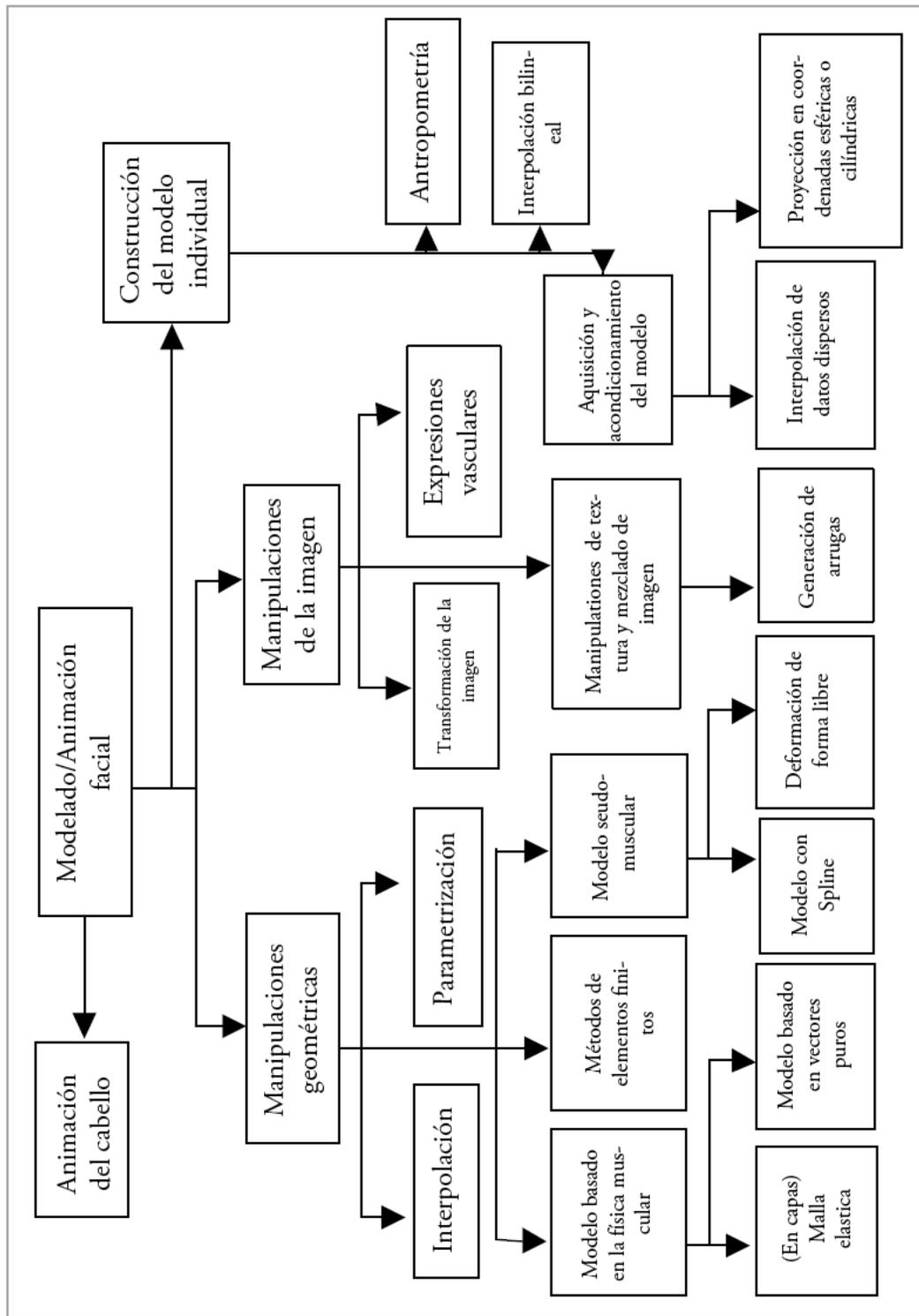


Figura 2.1: Las dos grandes vertientes del modelado y animación.

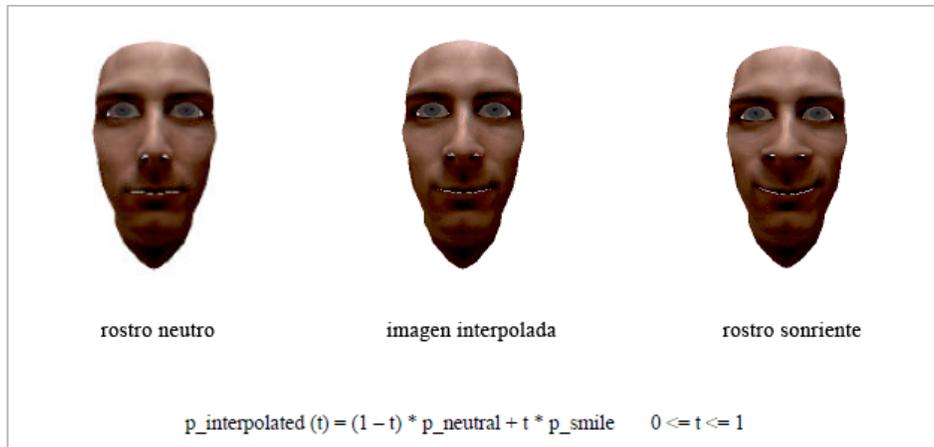


Figura 2.2: La interpolación lineal se realiza sobre valores de contracción muscular [Ter81].

La interpolación es un buen método si sólo se quiere un conjunto limitado de animaciones con base en solo unos cuantos *keyframes*.

### 2.1.2. Deformación geométrica

Proveer interfaces con un alto nivel de control geométrico sobre las deformaciones es el propósito principal de la *Deformación geométrica*, básicamente consisten en usar un objeto que modifica a uno más complejo, mostrando una gran eficiencia al modelar y animar objetos deformables.

En 1986 fué introducido por Parry Sederberg en su trabajo[Sed86] el concepto FFD (*free-form deformation*), una técnica que puede ser aplicada tanto a CSG (Constructive solid geometry) y B-Rep (Boundary representation) que deformaba el espacio en que el objeto se encuentra contenido, en lugar del objeto en sí. Más tarde en 1998 y posteriormente en el 2001 se lograron los mismos resultados, pero utilizando una *Lattice*<sup>1</sup>. Coquillart[Coq90] logró crear un sistema mucho más efectivo y eficiente, al extender FFD; una forma más natural de verlo, es pensar en imitar los procedimientos de esculpir y

<sup>1</sup>Consiste en una rejilla tridimensional de vértices no renderizable. Su principal uso es dar deformaciones adicionales a cualquier objeto que posea. Estos objetos pueden ser mallas, superficies e incluso partículas.

modelar, lamentablemente en el proceso se perdió algo de flexibilidad.

Los métodos de deformación del espacio son independientes de la representación del objeto (mallas, volúmenes en voxel, información de puntos dispersos, entre otros), convirtiéndolos en un gran herramienta de edición de objetos. El trabajo de Lazarus [Laz94] propuso una interactiva técnica llamada A\*Df (deformaciones por ejes y que puede ser combinada con alguna otra técnica de deformación), utiliza los ejes en lugar de una *lattice*, permitiendo a las deformaciones ser controladas por un eje de una forma fácilmente especificada.

Desde entonces se han implementado algunas otras técnicas, brindando un control de fácil manejo para crear animaciones creíbles; sin embargo, automatizar dichas técnicas requiere un esfuerzo enorme.

### 2.1.3. Animación facial con bases físicas

Los sistemas *mass-spring*, los algoritmos de elementos finitos, representaciones vectoriales y las capas de mallas con ciertas propiedad elásticas, tienen algo en común: son modelos que simulan la piel facial y músculos que generan las expresiones faciales. Los dos primeros al día de hoy son las tecnologías dominantes, que dependiendo de la simulación, pueden ser combinadas.

Hace casi tres décadas Platt y Badler [Pla81] mostraron una representación de bajo nivel de un rostro humano. El modelo fue posible al interconectar un conjunto de redes de puntos tridimensionales que simulaba la piel, músculos y huesos; destacándose aún más por utilizar FACS y *mass-springs* que representaron las fibras musculares. En [Wat87] fue construido un modelo muscular que fuera lo más “natural” posible; dos tipos de movimiento muscular fueron definidos: lineales/paralelos (que jalan) y esfínter (que se contraen). Los parámetros (que representaban movimientos musculares) fueron medidos con base a varios rostros, con el fin de saber la máxima y mínima zona de influencia, así como el desplazamiento total que sufren los puntos de la superficie; como consecuencia de este análisis, los parámetros pueden ser utilizados en alguna otra topología, mostrando ciertas expresiones fáciles, pero puede apreciarse la falta de sutileza de algunos movimientos. El trabajo posterior de Terzopoulos y Waters [Ter81] combinó un modelo con bases físicas del tejido facial y otro basado en el proceso muscular que se observa en la anatomía facial. Dicho modelo permitió una mejora significativa del realismo en la animación. Al día de hoy los principios básicos del modelo de Waters siguen vigentes.



Figura 2.3: Un “algoritmo” que hace uso de un modelo muscular con bases físicas [Sif06].

Lee et al[Lee95] hicieron uso del escáner cyberware<sup>2</sup> que constituyó un modelo virtual, eficiente y totalmente funcional. Su principal característica fue contraer músculos sintéticos y animar un modelo anatómicamente correcto de la piel (compuesta por cinco capas y similares al de una persona común). La simulación física propagaba las fuerzas musculares por toda la piel sintética, pero su propuesta hizo difícil definir de forma precisa los parámetros musculares, debido a la complejidad de los músculos humanos.

En lugar de depender totalmente de modelos matemáticos para crear expresiones faciales, Choe y Ko [Cho05] permitieron que los artistas esculpieran manualmente un conjunto de expresiones llamadas actuaciones musculares primarias, considerando en todo momento la anatomía facial, dejando que al final el programa sintetizara las expresiones por medio de combinaciones lineales de elementos básicos (cada uno es en esencia el modelo afectado por una sola expresión muscular). Esto condujo a una falta de precisión anatómica con resultados poco naturales, requiriendo esculpir nuevamente los modelos primarios.

Hace tan solo tres años se presentó un increíble modelo en [Sif06], que utilizó elementos finitos no lineales con el fin de determinar cuáles son los músculos que realizan alguna acción en particular, como se muestra en la Figura 2.3. Lo más interesante fue que las fuerzas musculares pueden ser combinadas con fuerzas externas, como el impacto de un objeto que colisiona, modificando la apariencia del rostro.

---

<sup>2</sup>[www.cyberware.com](http://www.cyberware.com)

#### 2.1.4. Captura de movimiento

El procedimiento originario de Scotland registra los movimientos más sutiles de las expresiones faciales, posteriormente mapea el movimiento a un modelo digital, y por ultimo son reproducidas en el modelo, animando a los personajes sintéticos. Concepto que puede ser similar al de Rotoscopía (se basa en dibujar directamente sobre la referencia, que pueden ser los cuadros de la filmación de una persona real<sup>3</sup>).

El método hace uso de técnicas manipulación geométrica y de imagen. Los primeros intentos pueden observarse en [Wat87, Lee95]; en todos se utilizaron sensores de escaneo, importando la imagen digitalizada de la geometría facial, creando así una malla lista para animarse; como consecuencia se inicio una mayor investigación basada en video. Un primer intento se dio en [Wil90] el trabajo fue dividido en la construcción de un modelo que fuera creíble y la demostración del sistema que animaba dicho modelo. La solución fue registrar las expresiones de un flujo de video, el motivo de la elección fue simple: el video no restringe la actuación del actor. Como paso final la información fue mapeada a un modelo en 3D. Las limitantes de estos trabajos fueron que el actor y el modelo digital tuvieran una fuerte semejanza, si existían pobres animaciones, entonces se optaba por otra técnica llamada Reasignación (sección 2.1.5). [Fid00, Cha03] hizo uso de una secuencia de video, analizando la posición de la cabeza y de pequeños grupos de expresiones, separando la información en 2 clases de control de parámetros: expresiones faciales y posición del cráneo. El trabajo sobresalía sobre algunos otros porque adaptaban los movimientos sintetizados a cualquier tipo de geometría facial, por lo que cualquier modelo podía producir expresiones sumamente creíbles. Desde el momento de su concepción los sistemas basados en marcas sobre el rostros (observe Figura 2.4) han soportado una pequeña cantidad de marcas (30-160) limitando el registro de sutiles expresiones, esta limitante condujo al desarrollo del sistema de captura de movimiento sin el uso de marcas [Bla03].

Han pasado algunos años desde el nacimiento de esta técnica, que ha ido evolucionado y conduciendo la investigación de aplicaciones especializadas, incluyendo: flujo óptico y análisis de imagen. Es indiscutible los cambios que ha sufrido y sufrirá esta tecnología [Zha04][Bor06], y que en conjunto con algunas técnicas la calidad de los resultados finales aumentaría considerablemente, por ejemplo: *Blend shapes*; dentro de la animación comercial, la

---

<sup>3</sup><http://es.wikipedia.org/wiki/Animación#Rotoscop.C3.ADa>



Figura 2.4: Puntos característicos [Cha03].

película Los Increíbles (2004) combinó la captura de movimiento con la animación tradicional de *keyframes*, beneficiando a los artistas que controlaron totalmente los medios que manipulan a los personajes.

Los resultados que pueden lograrse con esta tecnología son muy realistas y bastante notorios, sin embargo, los costos son muy elevados e inclusive pueden ser innecesarios si se produce algo con personajes anormales/fantásticos.

### 2.1.5. Reasignación

Este concepto, consiste en adaptar la información existente del movimiento animado desde un personaje a otro, quizá distintos entre sí, pero siempre aprovechando los movimientos creados para un personaje. Tal concepto fue introducido en [Gle98], pero no era apto para las animaciones faciales, debido a que no podía captar las sutilezas de las expresiones, es por esto que su aplicación solo se enfocaba a estructuras del cuerpo humano. Sin embargo el método como tal fue presentado por [Lee95]; método que consistía en mapear la información grabada de la actuación de cierto individuo, hacia otro, el resultado fue un rostro virtual que expresaba detalladamente las emociones. La reasignación mejora aun mas los métodos de captura de movimiento, al mapear el movimiento de la animación entre personajes disímiles en forma,

proporción, resolución y apariencia. Existen hoy en día diferentes implementaciones de este procedimiento basadas en el análisis de la información del movimiento o por medio de la parametrización facial, de este último caso se se derivan las parametrizaciones explícitas e implícitas.

La parametrización explícita necesita la directa correspondencia entre los parámetros de los modelos, la información del modelo original es entonces transmitida hacia el otro modelo. Regularmente el inicio del procedimiento es establecer las características clave del rostro y definir la relación entre los modelos. Para transferir la información hacia un modelo facial, los vectores desplazados (la diferencia entre las marcas correspondientes de una expresión y del rostro serio, Figura 2.4) son normalizados en orientación y magnitud, correspondiendo así el tamaño y proporción del modelo. Por último, los vértices restantes del modelo base son calculados usando un método de interpolación de información dispersa. Se propuso en [Noh01] *Expression Cloning*, método que selecciona manualmente o automáticamente de 10 a 35 vértices en ambos modelos (dependiendo de la complejidad del modelo), hecho esto se realiza una correspondencia uno a uno entre ambos modelos; el segundo paso será transferir los vectores de movimiento desde los vértices del primer modelo al segundo, además se realizaba una modificación usando funciones de base radial (o en inglés RBF). Por último la animación del modelo objetivo fue posible debido al ajuste de la magnitud y dirección de los vectores de movimiento que fueron transferidos. En 2003 [Pan03] el movimiento del rostro en 3D fue copiado hacia otro, pero a diferencia de Noh, él utilizó MPEG-4 FBA (Animación del cuerpo y facial, por sus siglas en inglés) y lo mencionado en [Pan02]; calcular los movimientos del rostro original, fue bastante simple, solo fue una diferencia entre la posición de los vértices del rostro animado y los del rostro estático. Finalmente se obtuvo la animación al agregar los movimientos faciales a los vértices del objetivo. Trabajos similares a estas propuestas fueron presentados posteriormente, sin embargo la mayoría de ellas solo se enfocan en la transferencia de la información de movimiento, sin abarcar y analizar cada expresión. Para superar esta limitación, se definen parámetros asociados con la apariencia física en la malla y en la estructura subyacente del rostro, estableciendo una capa mas de control, justo como FACS [Ekm78] relacionando la superficie del músculo a los movimientos de superficie del rostro.

Ahora bien la Parametrización implícita se enfoca en analizar la relación entre los diferentes *shapes* (expresiones). Para mapear la animación desde el modelo base hacia el modelo objetivo, es necesario que ambos tengan los

mismos *shapes* correspondientes, además de transferir los parámetros que definen los *shapes* para su posterior interpolación; los parámetros son los pesos o cantidades necesarias para lograr la deformación hacia *shapes* específicos. En [Chu02] fue desarrollado un método para crear animaciones faciales, combinando la captura de movimiento y *blend shapes*. La propuesta mantiene la flexibilidad del modelado de los *shapes* hechos por los artistas, mientras que gana la eficiencia de la animación al usar la captura de movimiento, evitando así que el artista haga a mano la animación. El procedimiento es sencillo: se capturan las expresiones del actor usando técnicas de visión computacional, al terminar se descomponen las expresiones faciales en combinaciones de pesos de *shapes*, eligiendo automáticamente un conjunto base de los *shapes*; por último el artista modela *shapes* semejantes a los descompuestos y los valores o pesos que se obtuvieron en la descomposición son usados para crear la interpolación en el modelo. El punto débil de este método es el tiempo que tarda el artista en modelar cada *shape* a mano.

La elección regular es la parametrización explícita, pero escoger la más apropiada siempre es con base a que tal útil es cada una. La parametrización implícita es la típica elección de los artistas porque se deriva de la animación tradicional.

## 2.2. Parametrización facial

Dada una representación del rostro humano, lo más lógico sería saber como debe ser controlada, es por eso que la parametrización define un conjunto óptimo de parámetros, los cuales son utilizados para controlar los movimientos faciales. Como lo menciono Parke [Par74], la animación puede ser desarrollada más fácilmente por los artistas, si en lugar de trabajar directamente sobre la geometría, se usa una parametrización. Otra ventaja sería que una expresión facial puede ser determinada por un pequeño número de parámetros; los artistas pueden trabajar con un mayor nivel de control, en lugar de manipular la geometría directamente. Por último al transferir sólo estos parámetros, una animación facial puede ser transferidas entre personajes porque comparten la misma configuración.

### 2.2.1. ¿Cuál es una buena parametrización?

Las investigaciones han arrojado que la parametrización ideal no existe; es difícil satisfacer lo que cada usuario requiere debido al gran gama de aplicaciones faciales. Para crear una buena parametrización para alguna animación facial, se deben considerar:

- **Completez:** ¿Los parámetros permiten la especificación de todas la expresiones que son posibles en cualquier personaje?
- **Restricción:** ¿Pueden ser sólo construidas configuraciones significativas con la parametrización?
- **Interpretación:** ¿Es intuitivo y claro para el usuario los efectos de los parámetros y sus controladores?
- **Tamaño:** ¿Cuál es el mínimo número de parámetros en un conjunto?
- **Entorno:** ¿Cuál el propósito de la parametrización? (animaciones, video juegos, etc.).

Una vez que se elige el propósito de la parametrización, es importante que el conjunto de parámetros permita mostrar una gran variedad de expresiones, aunque su utilidad puede ser limitada a la gran gama de expresiones faciales. De acuerdo con Ekman y Friesen, existen 5000 diferentes combinaciones de AU's, las cuales cubren aproximadamente 30 diferentes significados. Ekman y Friesen [Ekm78] definieron FACS, para describir y medir el comportamiento del rostro y sin ningún propósito para la animación. El valor FACS radica en la validación de las expresiones faciales.

A continuación se presentarán dos corrientes principales dentro de la parametrización (la elección de alguna corriente depende enormemente hacia que área será aplicada).

### 2.2.2. Parametrización basada en los rasgos faciales

Este método hace uso del conocimiento del rostro y las expresiones que puede realizar, si una nueva expresión se añade un nuevo parámetro controlará la expresión, así como lo expone [Par82]. En 1982 Parke uso un conjunto diferente de parámetros de conformación es usado para darle forma al rostro, es decir, para controlar las animaciones del rostro (estos incluían el cuello,

mejillas, etc.) . Más adelante C. Cohen y Massaro usaron los parámetros para controlar el largo de la lengua, ángulo, ancho y espesor para una síntesis de un discurso en [Coh93]. En ambos sistemas, la geometría facial esta directamente distorsionada por los parámetros y la creación del movimiento sólo fue cuestión de utilizar la interpolación de los vértices entre dos posiciones extremas del modelo.

El trabajo de [Gue98] concluyó que el movimiento de las expresiones puede ser definido por el desplazamiento de los puntos característicos del rostro. Este sistema hizo además uso de la captura de movimiento. El punto característico se relacionó con el punto de control de la geometría facial, es así que al cambiar la posición del primero se provocaba una deformación en el modelo. El estándar MPEG-4 [Koe02], también emplea puntos característicos, haciendo solo una distinción entre los parámetros de definición facial(FDP) y los parámetros de animación facial (FAP), donde FDP definen el *shape* del rostro y FAP controlan la animación. Aunque El MPEG- 4 no define el movimiento de los puntos característicos y la deformación de la piel; esto deben proporcionar las aplicaciones.

### 2.2.3. Parametrización muscular

Este método llegó a ser muy popular para manipular las deformaciones; su popularidad se debió a que la geometría facial era una simulación de las propiedades de los músculos. El uso del conocimiento anatómico permitió definir la movilidad y la forma del rostro. Aunque no es perfecto, debido a que la animación se vuelve más compleja (el rostro virtual tiene demasiados músculos que necesitan ser controlados para obtener una simulación realista del movimiento del rostro). Es así que es necesario tener una capa superior relacionada con los músculos que simplifique la interfaz de control de la animación. FACS es la piedra angular para el desarrollo de la investigación de las parametrizaciones musculares. En [Wat87, Wat89] fue desarrollado un modelo del tejido facial y su articulación fue representada al sintetizar los músculos, permitiendo su observación, análisis y previendo la movilidad del tejido suave sobre el rostro. Esto conduce a diferentes aplicaciones de animación facial, que no requieren predefinir la ejecución de las acciones. Más adelante se presentó en [Ter81] un método que incorporaba una aproximación entre el tejido facial y un conjunto anatómicamente correcto de músculos faciales. En dicho trabajo fueron estimadas las contracciones musculares realizadas por actores y registradas en secuencias de video. En 1999 la M. en C. Lour-

des Guerrero presentó en [Gue99] un modelo más general y flexible para la parametrización, dicho modelo podía representar los 18 músculos requeridos por FACS, ignorando los atributos físicos, pero imitando las características primarias del rostro.

#### 2.2.4. El MPEG-4

Este método extiende diferentes aproximaciones y fue el primer control parametrizado de un rostro en ser un estándar: MPEG-4 FAB [Pan02]. Usado para investigaciones académicas y aplicaciones comerciales. Mark Sagar [Sag06] inició el uso de FACS en las animaciones fáciles dentro de la industria de entretenimiento. El resultado de su investigación fue un sistema que capturaba el movimiento de sutiles expresiones y posteriormente las trasladaba al rostro de un modelo sintético; dicha técnica fue utilizada en varios largometrajes. El instituto Fraunhofer HHI<sup>4</sup> desarrollo un avanzado sistema de video conferencia usando MPEG-4, el sistema es capaz de codificar las secuencias de imagen de la cabeza de 1 Kbits/s. Los modelos en 3D fueron usados para representar a la persona en la escena y MPEG-4 específico los parámetros que definieron los cambios temporales de las expresiones faciales.

La gama de aplicaciones que usan alguna forma de parametrización está en constante crecimiento y sobre todo nuevas técnicas continúan emergiendo; por ejemplo Xface es un conjunto de herramientas de uso libre para crear y editar MPEG-4 [Bal04].

### 2.3. Industria del entretenimiento: herramientas internas

Como se mencionó al inicio de este capítulo actualmente no existe una solución automática para el *rig* facial de los personajes, es por eso que la mayoría de las compañías desarrollan sus propias herramientas para acelerar proceso de *rigging*, aun así, al menos el 50% del trabajo tiene que ser hecho por personas; usualmente tal trabajo se realiza para los personajes principales de las películas animadas, tales como Shrek (2001) o King Kong

---

<sup>4</sup><http://www.hhi.fraunhofer.de>

(2005). Los personajes necesitaban mostrar una gran cantidad de deformaciones detalladas, las cuales se reflejaban en sutiles y creíbles emociones; el *rig* es completamente personalizado, además de que técnicas específicas son implementadas para cumplir los requerimientos del director técnico. Para personajes secundarios, el proceso de *rigging* combina scripting y ajustes hechos por los modeladores; todo se realiza para acelerar algunas tareas como la transferencia de animaciones, pero si los personajes son muy disimiles los *scripts* no pueden ser reusados.

## 2.4. Aplicaciones libres y patentadas

Existen una gran variedad de componentes en el mercado, libres o con patentes, algunas han sido desarrolladas por académicos, algunas otras por la industria, pero ambas proveen algunas herramientas para crear un flujo de trabajo requerido para desarrollar animaciones faciales de personajes virtuales. La mayoría de las aplicaciones que realizan un análisis facial automático y animación son extremadamente pobres en calidad y sólo son usados en videojuegos o aplicaciones web. Las soluciones basadas en la captura de movimiento muestran resultados excelentes en películas, pero aún queda mucho para que puedan ser automatizadas y reducir el cuello de botella en las producciones de animación.

## 2.5. Futuro de la animación facial

La investigación de la animación facial está orientada hacia conseguir mejores modelos y herramientas que controlen el rostro, ofreciendo resultados sutiles y realistas. Una interesante dirección es agregar una capa con la que se pueda abstraer el lenguaje natural y que se encuentre encima de los actuales controles de animación, permitiendo quizá que el propio director opere y decida que expresión utilizar en cada escena.

Aparte de las sofisticadas aplicaciones que aparecen continuamente en la industria del entretenimiento y en la medicina, existe un interés para nuevas aplicaciones, por ejemplo: la interacción humano-computadora, computación afectiva y agentes de conversación. La meta es desarrollar un sistema capaz de detectar, procesar y responder apropiadamente a las emociones del usuario.

Consecuentemente, para las películas y videojuegos, la emoción parece

ser la siguiente gran frontera. Con la contribución de métodos de inteligencia artificial (IA), los personajes serán capaces de simular complejos comportamientos, enfatizando las cualidades esenciales de los humanos.

### 2.5.1. El porvenir del *rig* facial

Existen cuestiones relacionadas con el *rig* facial, que no han sido abordadas y son de suma importancia, como mapear el comportamiento físico del rostro humano hacia cualquier objeto (como un libro) y mostrar una mayor credibilidad a cualquier personaje que no sea un humano [Pig06].

Regularmente el método elegido de cualquier artista es *blend shape*. La técnica por sí misma no provee suficientes detalles para producir los sutiles movimientos de los rostros reales. Así como también existen otros problemas con los demás métodos, es necesario la investigación de nuevos métodos, pero sobre todo el objetivo principal es disminuir el tiempo que se necesita en todo momento de la producción, manteniendo la libertad que el artista pueda requerir, así como resultados de alta calidad.

# Capítulo 3

## Automatización del *rig* facial

*“Dios les ha dado un rostro y ustedes crearán el suyo”*

*William Shakespeare*

Durante la construcción de rostros virtuales, diversas técnicas y herramientas son empleadas (comerciales o hechas por los animadores), además del gran tiempo que se le dedica, es por eso que ese trabajo debería ser reducido. Debido a que los modelos virtuales comparten características básicas de geometría y estructuras anatómicas, existe la posibilidad de transferir el *rig* facial.

### 3.1. El problema a considerar

El tiempo para poder realizar el *rig* facial depende de las características o las aplicaciones que se requieran, aunque regularmente es un proceso de una a cuatro semanas (incluso para personas experimentadas). Los anteriores capítulos han mostrado los retos que se enfrentan al crear un solo rostro virtual y los métodos que se utilizan para animar dicho rostro. El método que se presentará utiliza y conserva la calidad de *rig* facial basado en *blend shapes*, pero antes de continuar es necesario tener en cuenta los siguientes problemas:

#### 3.1.1. La singularidad de los rostros

Alrededor del globo existen diversas caras y cada una tiene diversas características que la hacen ser única (ver sección 1.2), por lo que es probable

que al transferir el *rig* el animador podría obtener una animación que el público no acepte. Esto se debe a que las personas son capaces de detectar un comportamiento anormal en cualquier rostro. Debido a la imaginación del hombre se ha ampliado aun más la diversidad debido al diseño de personajes caricaturescos e incluso fantásticos.

La gran variedad de rostros que pueden crearse es una de las razones por la cual el *rig* facial debe ser tan personalizado.

### 3.1.2. Dificultades técnicas del *rig* facial

Contar con un *rig* facial permite a la animación en 3D alejarse de la animación tradicional (en la cual los movimientos del personaje deben realizarse frame por frame), sin embargo, está lejos de ser un método perfecto, por lo que como cualquier tecnología cuenta con diversas dificultades:

- No hay un estándar; dependiendo de las necesidades de los personajes se crea su *rig*.
- Los cambios a la geometría inicial obligan a recrear el *rig*.
- Agregar un nuevo *shape* puede alterar el desempeño del *rig*.
- Si los *blend shapes* están controlados por alguna interfaz de usuario(IU), esta se debe probar y reajustar con cada cambio en la interfaz y con cada *shape* que se agrega.
- La IU puede ser confusa e incluso podría tener otra apariencia para cada personaje.
- Debido a la variedad de los rostros no existe un estándar para el área del rostro que afectan los controladores.

## 3.2. Una solución: deformación de modelos faciales

Algunos métodos como la captura de movimiento, brindan la posibilidad de transferir las animaciones que hayan sido registradas a diversos personajes, sin embargo para algunas producciones esto es inaccesible.

Además de ser afable para la producción y para los animadores, la solución que se presenta reutiliza el *rig* hecho a un personaje, adaptándolo perfectamente a la fisonomía del rostro al cual se le transfiere los *blend shapes*. Por otro lado se debe tener en cuenta cuidar bien la calidad del *rig* del modelo que se tomará como base, para asegurar la misma calidad del *rig* en los demás personajes.

Una característica que integra al sistema es que emplea puntos característicos (o *landmarks*) en ambas mallas, con el fin de obtener una función de deformación (o función de interpolación), la cual será utilizada para ajustar los parámetros que se utilizan en el *rig* de la malla inicial y así formar el nuevo *rig*. Sin embargo su principal desventaja reside en que la precisión de la deformación en una área local depende de cierta cantidad de puntos característicos.

En las siguientes secciones se expondrá el sistema y las características o funciones que lo integran.

### 3.3. Descripción del sistema

Bookstein [Boo97] sabía que las computadoras tendrían un gran papel en el desarrollo de las animaciones, por eso sus trabajos muestran un análisis acerca de cómo es que las modificaciones en las formas biológicas pueden ser presentadas e interpretadas. Pero el tema de mayor relevancia para el desarrollo de este trabajo (*Thin-plate spline*) se encuentra en [Boo89]. Esta sección brinda los detalles en la construcción de la función de deformación para un conjunto de puntos característicos del modelo original y al que se le desea transferir el *rig*.

#### 3.3.1. *Thin-plate spline*

A la información que es recolectada de forma irregular, se le conoce como dispersa o al azar o no-estructurada. La información dispersa (*scattered data*) puede encontrarse en diferentes problemas de diferentes campos científicos (geología, meteorología, cartografía y en la digitalización de modelos de superficie) la cual necesita ser interpolada o aproximada. *Thin plate spline* (TPS) es el análogo en dos dimensiones de la función de interpolación cubic spline en una dimensión y su nombre proviene de la analogía que involucra doblar una delgada hoja de metal.

Aunque TPS es una aproximación de información en dos dimensiones, fácilmente puede adaptarse a un conjunto de tres dimensiones [Boo89]. La adaptación al problema de interpolación puede ser planteada de la siguiente forma: Sea  $\vec{x} = T_\lambda^\mu(x)$  de  $\mathbf{R}^3$  en  $\mathbf{R}^3$ . Dado un conjunto de  $n$  *landmarks*  $\lambda_i$  en  $\mathcal{S}_S$  y el correspondiente conjunto  $\mu_i$  en  $\mathcal{F}_S$ , la función de TPS satisface que  $T(\lambda_i) = \mu_i$  y se define como la función general:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n w_{i,x} U(\mathbf{x}, \lambda_i) + a_{1,x} + a_{x,x}x + a_{y,x}y + a_{z,x}z \\ \sum_{i=1}^n w_{i,y} U(\mathbf{x}, \lambda_i) + a_{1,y} + a_{x,y}x + a_{y,y}y + a_{z,y}z \\ \sum_{i=1}^n w_{i,z} U(\mathbf{x}, \lambda_i) + a_{1,z} + a_{x,z}x + a_{y,z}y + a_{z,z}z \end{bmatrix} \quad (3.1)$$

De acuerdo a Bookstein [Boo89] la solución fundamental a la ecuación bioarmónica  $\Delta^2 U = 0$  (la cual es  $|r^3|$  en una dimensión y  $r^2 \log r^2$  en dos dimensiones) es para  $\mathbf{R}^3$ :  $U(\mathbf{x}, \lambda_i) = |\mathbf{x} - \lambda_i|$ . Dicha solución minimiza la energía de doblado en la deformación. TPS trabaja muy bien con información dispersa e incluso se obtiene una descomposición en características ortogonales de escalas progresivas geoméricamente pequeñas, que pueden ser usadas para tratar con aristas curvas entre los *landmarks*.

Para comprobar el mapeo de la interpolación definida en (3.1) se tiene que computar los coeficientes para el conjunto de *landmarks*. Esto se puede lograr al imponer una correspondencia entre el conjunto de *landmarks*, obteniendo un sistema lineal de ecuaciones de  $(n+4) \times (n+4)$ :

$$\begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{Q} \\ \mathbf{0} \end{bmatrix} \quad (3.2)$$

$\mathbf{K}$  es una matriz de  $n \times n$  definida por elementos  $\mathbf{K}_{i,j} = U(\lambda_i, \lambda_j)$ ,  $\mathbf{0}$  es una matriz de  $4 \times 4$  y  $\mathbf{P}$  es una matriz de  $n \times 4$ , la cual está definida por columnas de *landmark* del modelo que se le realizó el *rig* ( $\mathbf{P}_i = (1, x_i, y_i, z_i)$ ) y se muestra a continuación:

$$P = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & z_n \end{bmatrix} \quad (3.3)$$

Las matrices desconocidas son descompuestas en una matriz  $\mathbf{W}$  de  $n \times 3$  y una matriz  $\mathbf{A}$  de  $4 \times 3$  definidas como:

$$\mathbf{W} = \begin{bmatrix} w_{1,x} & w_{1,y} & w_{1,z} \\ \vdots & \vdots & \vdots \\ w_{n,x} & w_{n,y} & w_{n,z} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} a_{1,x} & a_{1,y} & a_{1,z} \\ \vdots & \vdots & \vdots \\ a_{z,x} & a_{z,y} & a_{z,z} \end{bmatrix} \quad (3.4)$$

Finalmente, los términos independientes de lado derecho de la ecuación (3.2) son una matriz  $\mathbf{Q}$  de  $n \times 3$  (formada por columnas de los *landmarks* del modelo objetivo  $\mathbf{Q}_i = (\vec{x}_i, \vec{y}_i, \vec{z}_i)$ ) y una matriz  $\mathbf{0}$  de  $4 \times 3$ :

$$\mathbf{Q} = \begin{bmatrix} \vec{x}_1 & \vec{y}_1 & \vec{z}_1 \\ \vdots & \vdots & \vdots \\ \vec{x}_n & \vec{y}_n & \vec{z}_n \end{bmatrix} \quad (3.5)$$

El término  $\mathbf{KW} + \mathbf{PA} = \mathbf{Q}$  en (3.2) asegura la exacta correspondencia entre los puntos del modelo original y el objetivo. El término  $\mathbf{P}^T \mathbf{W} = \mathbf{0}$  representa una condición de frontera que regulariza la curvatura no-lineal con el fin de desaparecer su energía al infinito.

Este sistema lineal puede resolverse usando una descomposición LU estándar con pivote y por consiguiente un punto puede ser transformado por medio de

$$\vec{x} = T_\lambda^\mu(x) \quad (3.6)$$

Cabe hacer notar que debido a la fuerte interconexión de los *landmarks* y a las propiedades de TPS se obtendrán buenos resultados al deformar la malla e incluso usando sólo algunos *landmarks*.

### 3.3.2. Deformación de los rostros

Dada la función (3.6) definida por los conjuntos de *landmarks*  $\lambda$  y  $\mu$  que fueron colocados en los rostros, la *malla* es entonces deformada por la directa aplicación de esta función a sus vértices (Figura 3.1).

Posteriormente se hace uso de una *correspondencia densa* para garantizar un mejor posicionamiento de la estructura. Una correspondencia densa entre superficies se define como lo siguiente: un punto es mapeado desde  $\mathcal{S}_S$  hacia

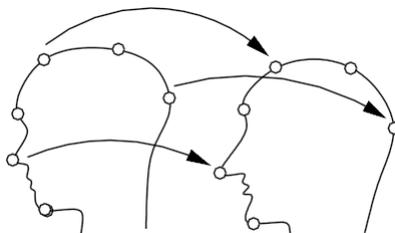


Figura 3.1: Los *landmarks* del rostro inicial son mapeados directamente a los *landmarks* de la malla objetivo (Figura hecha por [Tei07]).

el punto más cercano de  $\mathcal{F}_S$ . Como resultado, se obtiene el punto correspondiente en la malla objetivo para cada punto de la malla original (Figura 3.2).

Una vez definido lo anterior se puede realizar la transferencia del *rig* con los pasos que a continuación se muestran:

1. Se calcula el desplazamiento entre los vértices correspondientes de un *blend shape* y la malla en su posición base (el rostro en posición relajada, coincide con una expresión neutra, “serio”).
2. Se encuentra la correspondencia densa entre los vértices de los modelos.
3. Se crean los *blend shapes* (los cuales son adaptados a las proporciones del modelo objetivo usando TPS) de la malla objetivo al cambiar la posición de sus vértices.

La Figura 3.3 muestra el proceso de transferencia en 2D, mientras que la figura 3.4 muestra el proceso con modelos en 3D.

### 3.3.3. Blender y Python

La aplicación que se eligió para modelar las expresiones e implementar el sistema es Blender<sup>1</sup>, que es una aplicación de gráficos en 3D. Entre algunas de sus características es que tiene integrado Python *scripting* (esto no

---

<sup>1</sup>[www.blender.org](http://www.blender.org)

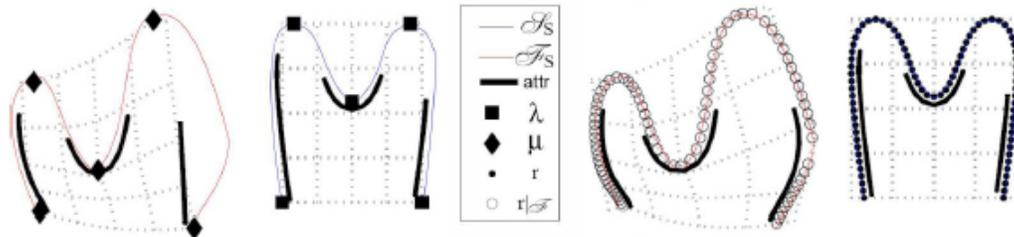


Figura 3.2: A la izquierda se encuentra una curva uniforme que será transformada en otra curva usando TPS y un conjunto pequeño de *landmarks*. Sólo las posiciones de los *landmarks* tienen una exacta deformación, mientras que el resto de los puntos quedan fuera de la curva. A la derecha los resultados de aplicar la correspondencia densa, la cual asegura una mejor posición de los vértices que integran a la curva y de la estructura (Gráficas hechas por [Tei07]).

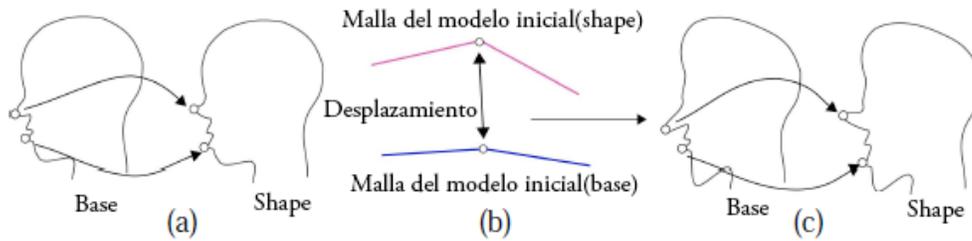


Figura 3.3: Representación en 2d de la transferencia del *rig*: (a) a la izquierda se muestra el modelo inicial sin ningún *shape* o deformación; (b) se obtienen los vectores de desplazamiento entre los vértices correspondientes de la base y el *shape*; (c) se deforma la malla base de acuerdo a los vectores de desplazamiento (Figura hecha por ) [Tei07].

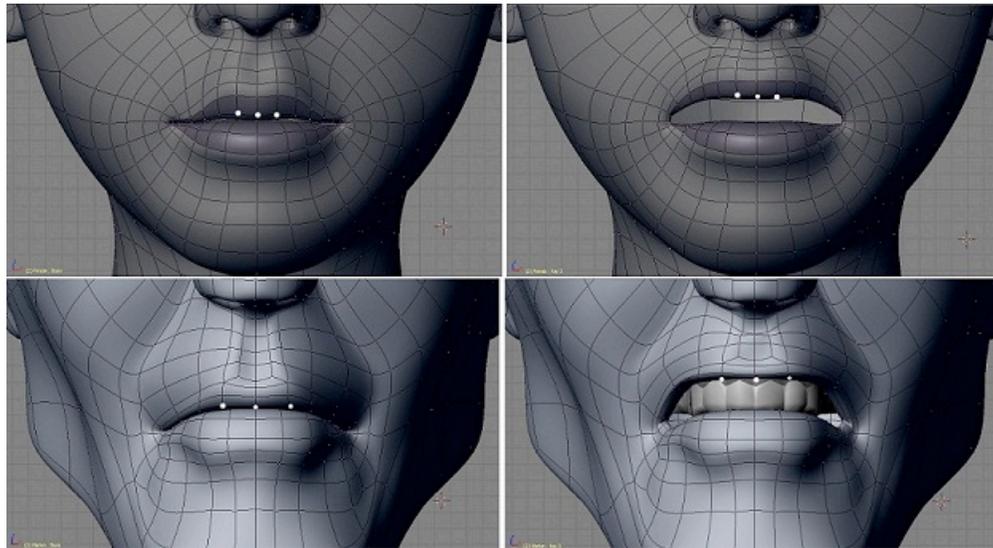


Figura 3.4: En la parte superior se muestra el modelo inicial en su estado neutro (izquierda) y su *shape* (derecha) ; de la misma forma en la parte baja para el modelo objetivo. Los puntos representan a los vértices en las mallas base y los correspondientes en los *shapes*. El *shape* del modelo objetivo se calcula con respecto a la información del modelo inicial. Debido a una mayor definición del *wireframe* la apariencia del modelo mejora y los resultados son más claros ( el *shape* o *visime* es adaptada correctamente a la topología del objetivo). Modelos hechos por Ben Dansie.



Figura 3.5: Los 44 *landmarks*. Copyright 2005 Dygrafilms.

quiere decir que no pueda desarrollarse en otro lenguaje o en otra aplicación de gráficos en 3D). Blender está escrito en C, C++ y en Python, algo muy importante, porque resulta muy sencillo crear *scripts* que se integren perfectamente con Blender.

Las diferentes expresiones que fueron modeladas y utilizadas para poder comprobar la eficacia del sistema fueron elegidas debido a que son muy diferentes entre sí y cada una de ellas afecta a una zona en particular del modelo.

La cantidad de *landmarks* que utiliza el sistema (44 en total) no es una cantidad arbitraria; no solo son los suficientes, sino son los más destacados y distintivos en el rostro [Far87] (vea Figura 3.5). Aunque su elección depende de cada modelo, es un proceso que debe realizarse para todos los modelos y *landmark* por *landmark*.

Una vez que se concluyó todo lo anterior, se inició la construcción del sistema que obtiene la función (3.6) y que implementa cada uno de los pasos mencionados en la sección 3.3.2. El sistema es bastante simple en su estructura, debido a que cada función es un paso de la transferencia del *rig*.

### 3.4. Flujo de trabajo

En las anteriores secciones fue mostrada en diversos aspectos la aplicación, pero en esta sección se verá el panorama general del flujo de trabajo. La entrada en el sistema es el modelo S. La salida sería el modelo F con su apropiado *rig* y totalmente listo para asignarle los controladores para poder

animarlo. Los pasos son los siguientes:

**Seleccionar los *landmarks*** para mantener la correspondencia entre ambos modelos se definen sus *landmarks*.

**La correspondencia entre superficies** asegura que los *landmarks* estén bien apareados así como la correcta interpolación entre otros puntos.

**Correspondencia densa** para evitar el uso de mas puntos, debido a que se asegura la exacta deformación para cada punto de la superficie.

**Transferencia de *shapes*** Se transfiere el *rig* facial al modelo objetivo.

La Figura 3.6 muestra el panorama general de lo anterior.

Animar personajes virtuales es para muchos un reto, sobre todo el rostro, que puede hacer más de 5000 expresiones<sup>2</sup>. Diversos personajes en diversas películas pueden tener un *rig* sumamente complicado; la presente aplicación permite obtener los mismos resultados que un artista logra en un lapso máximo de tres semanas, en unos cuantos minutos (depende de lo rápido que se coloquen los landmarks y la velocidad del procesador para realizar las operaciones).

---

<sup>2</sup>[http://www.ehow.com/about\\_5553754\\_human-face.html](http://www.ehow.com/about_5553754_human-face.html)

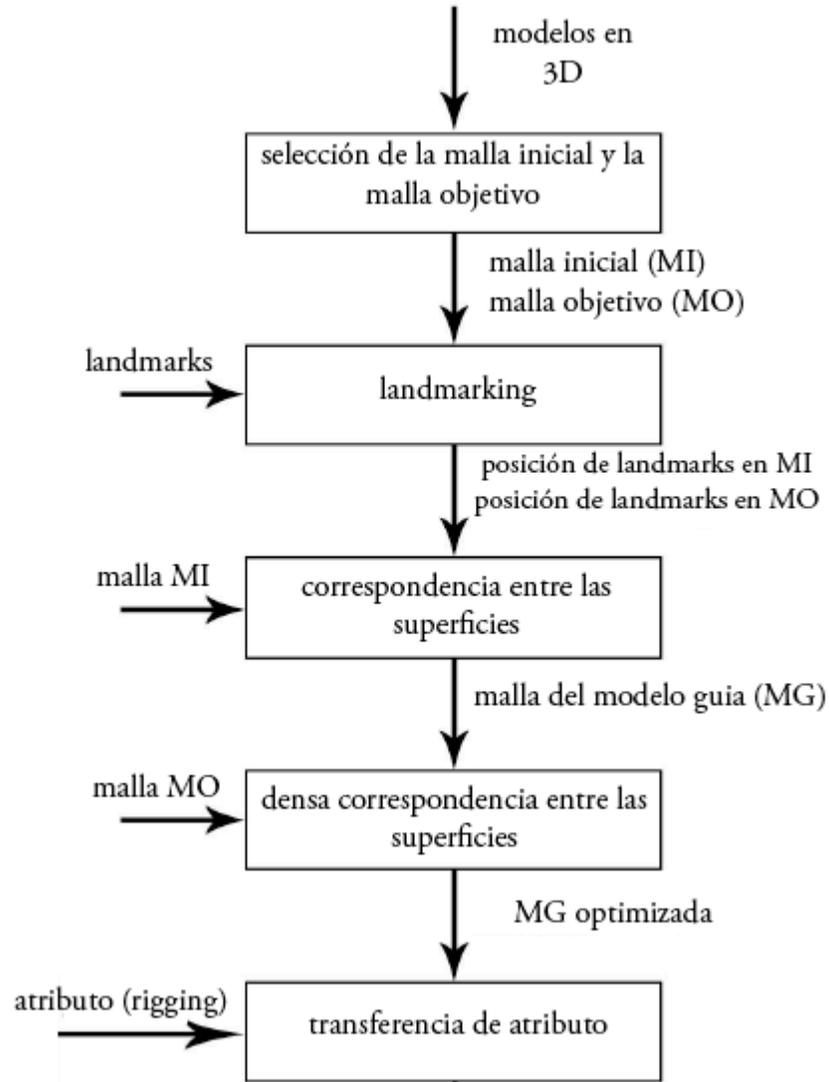


Figura 3.6: Diagrama de flujo del sistema. El diagrama muestra las acciones que se realizan para obtener la transferencia del *rig*, el segundo paso requiere de algún tiempo, esto es porque se deben colocar uno por uno los *landmarks* en ambos modelos, pero una vez finalizada esta tarea el sistema es automático.

# Capítulo 4

## Resultados

*“Muestra mi cabeza a las personas, vale la pena verla”*

*Georges Jaques Danton, a su verdugo*

Aunque los resultados al trasladar el *rig* facial parecen ser satisfactorios fue necesario comprobarlos de alguna manera; por lo que el *rig* debe ser hecho a mano para los otros dos personajes que se utilizaron en el sistema (un personaje masculino joven y uno maduro) y comparar entonces ambos resultados.

### 4.1. Aplicación del sistema a un rostro joven y a un rostro maduro

La meta de este trabajo (como se planteo en sección 1.9 ) es reutilizar el *rig* creado en un personaje, en otros personajes. El *rig* usado corresponde al del modelo llamado Rowena<sup>1</sup>, un rostro realista de una mujer joven, por lo que el reto del sistema es obtener expresiones similares y adecuadas en los modelo masculinos, que tienen como característica un diseño más caricaturesco. En total el sistema hace uso de tres modelos en 3D, proporcionados por Ben Dansie y las expresiones faciales hechas por el autor del presente trabajo. Se utilizó una computadora personal con las siguientes características: Intel Core 2 Duo CPU E8400 con 2 GB de RAM y Windows 7.

---

<sup>1</sup>Los modelos llamados Rowena, Marlan y Sandrak fueron hechos para el cortometraje animado de la historia de Arthur C. Clarke, “El despertar”

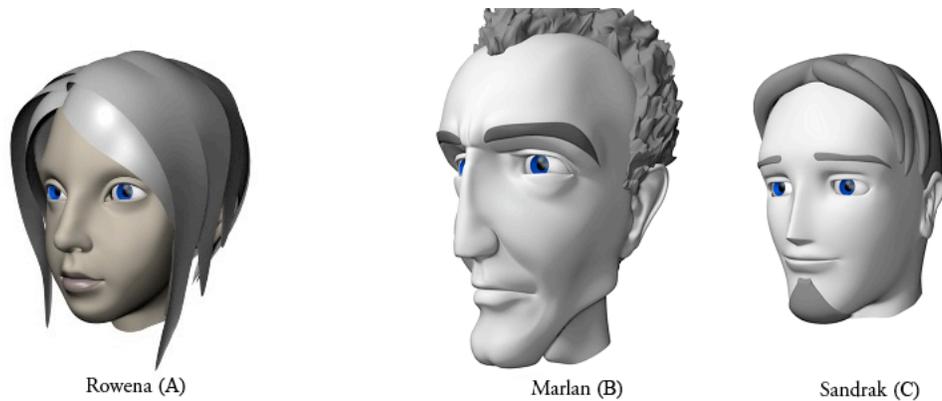


Figura 4.1: Los modelos que fueron utilizados en el sistema. Puede apreciarse las grandes diferencias en su geometría. (A) Rowena, (B) Marlan y (C) Sandrak.

Los modelos caricaturescos (como Marlan) tienen diferente composición y diferente representación visual en su comportamiento, por lo que no siguen las proporciones humanas, la figura 4.1 muestra el modelo inicial y los dos modelos objetivos. Lo que se pretende revisar y comprobar son los *blend shapes* generados en cada uno de los modelos.

#### 4.1.1. Verificación del sistema

Para verificar la salida del sistema, se verifica visualmente que el *rig* que se le transfirió a cada personaje sea correcto y correspondiente al *rig* de Rowena, es decir, que se comprueba que los *shapes* de las expresiones faciales se reproducen correctamente en todos los modelos y que todas esas expresiones corresponden a las que tiene Rowena (observe la figura 4.1 donde Rowena es más realista y Sandrak tiene un rostro más redondo) .

#### 4.1.2. Comprobación del sistema

La prueba que se realizó para asegurar la eficacia del sistema fue realizar el *rig* facial a mano, es decir, hacer las mismas expresiones faciales de Rowena a cada uno de los personajes que se usaron. La figura 4.2 muestra las

	Rowena	Sandrak
Triángulos	2976	2116
Deformación del rostro(TPS)	–	< 1 seg
Correspondencia Densa	–	< 1 seg
Creación de <i>shape</i> (cada uno)	4/6 Horas	< 1 seg

Cuadro 4.1: Comparación de tiempos en cada paso del sistema.

expresiones generadas entre las dos técnicas.

Como segundo paso se comparó el *rig* generado por el sistema y el creado manualmente por el autor de esta tesis. Se puede observar por medio de la figura 4.2 que son similares las expresiones entre ambas técnicas. Es por esto que ahora se debía asegurar que el sistema genera mucho más rápido las expresiones que el *rig* hecho a mano. La tabla 4.1 muestra una comparativa en tiempos entre las dos técnicas por cada expresión facial, mostrando así la eficacia del método ante el *rig* hecho a mano para los demás personajes.

### 4.1.3. Transferencia y validación del *rig* al modelo masculino maduro

El rostro femenino es más realista y muy diferente en proporciones al de Marlan por lo que en esta sección se muestran los resultados de trasladar el *rig* entre dos mallas con tan diferente topología. Lo importante de esta prueba era verificar que el *rig* fuera transferido correctamente y que las expresiones fueran las deseadas en el modelo masculino (con una geometría un poco fantástica).

De igual forma que en la sub-sección 4.1.2 se presenta en la figura 4.1.3 el resultado de aplicar el sistema a la malla y los resultados de hacerlo a mano. Y de nueva cuenta se presenta la tabla 4.2 que muestra los tiempos que se necesitan para lograr en ambos casos las expresiones.

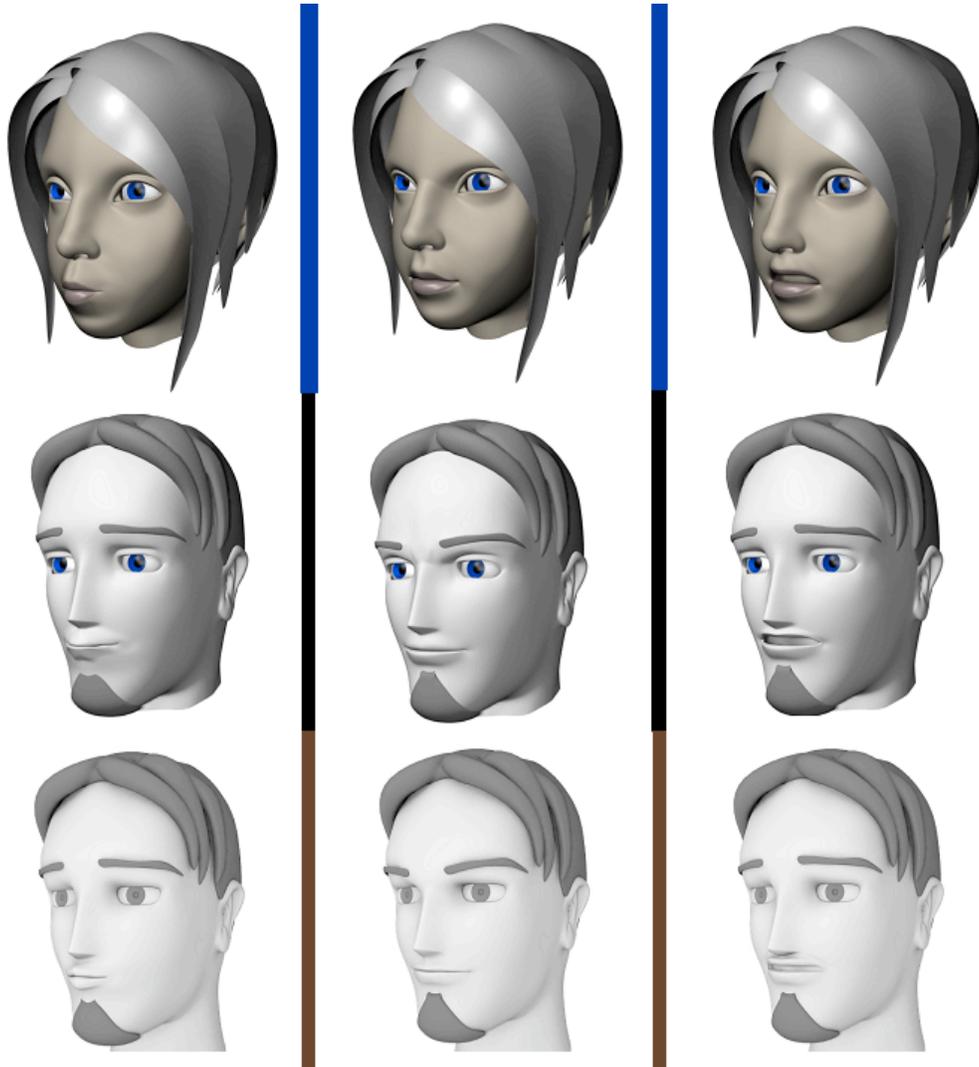


Figura 4.2: Primera comparación. En la parte superior de la imagen se aprecian las expresiones que se realizaron a Rowena; la segunda fila (la parte de en medio) se muestra la salida del sistema; al final los *shapes* que se modelaron para Sandrak.

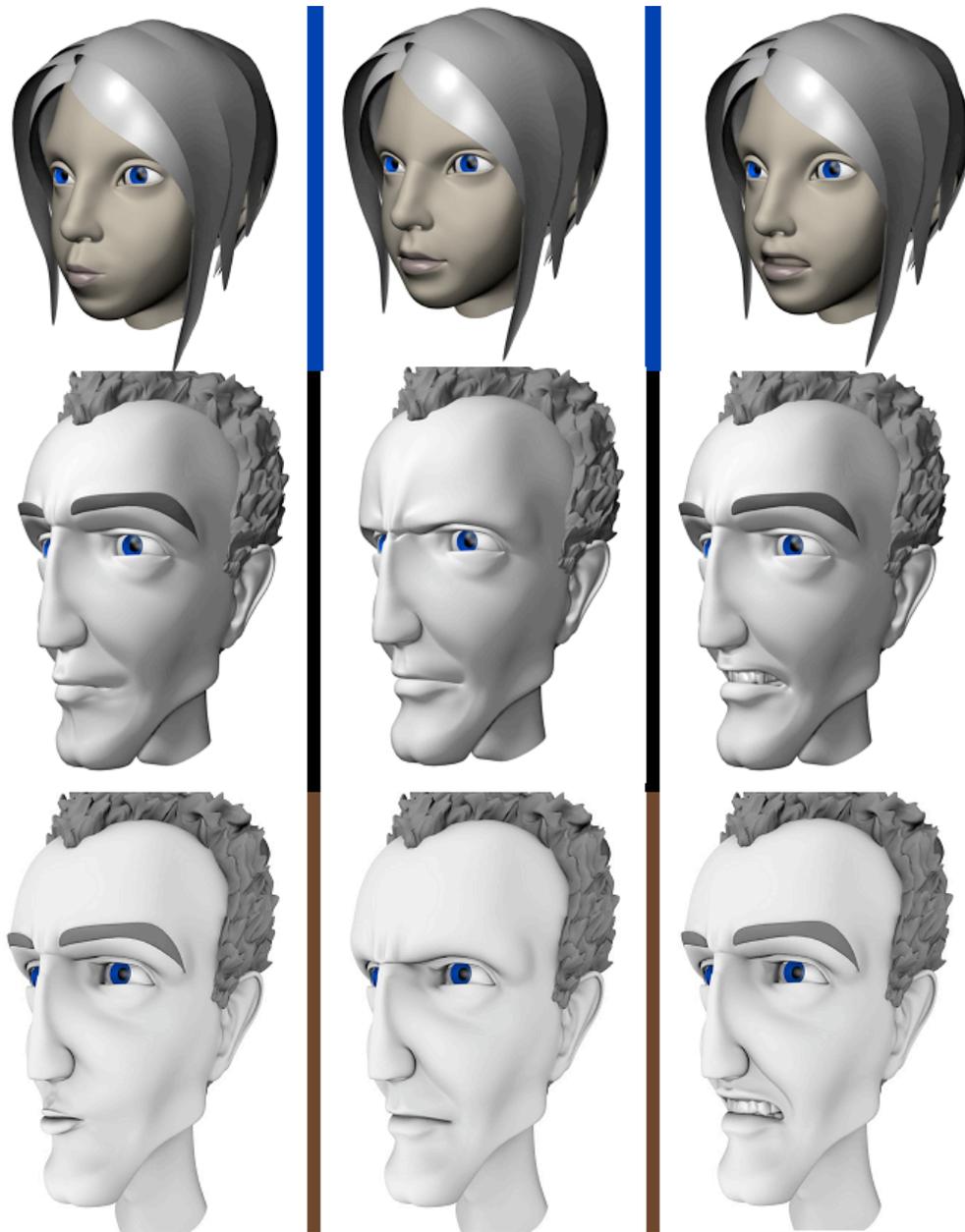


Figura 4.3: Comparación de las expresiones faciales. De la misma manera que en la figura 4.2 se muestra en la primer fila los *shapes* de Rowena (son el *visime Narrow*, con el ceño fruncido y el *visime Upper Lip Up*); en la segunda fila se observa los *shapes* que genero el sistema; en la última fila se encuentran las expresiones que se modelaron a mano para Marlan.

	Marlan
Triángulos	3116
Deformación del rostro(TPS)	< 1 seg
Correspondencia Densa	< 1 seg
Creación de <i>shape</i> (cada uno)	< 1 seg

Cuadro 4.2: La tabla muestra los tiempos en los que el el sistema completa la transferencia del *rig* de Rowena.



Figura 4.4: Usando curvas y *locators* se asignan los *landmarks* a la malla original y objetivo.(Copyright 2005 Dygrafilms)

## 4.2. Observaciones

A lo largo del desarrollo del sistema se descubrió lo tardado de colocar los *landmarks*, ya que solo se pueden hacer manualmente; para facilitar la tarea podría implementarse una interfaz que facilite al usuario dibujar curvas (figura 4.2) en los modelos que definen regiones faciales(como las cejas, los bordes de la nariz y labios), en lugar de marcar cada *landmark* del rostro y también es posible predefinir *landmarks* en locaciones específicas. Como siguiente paso el usuario escoge el numero de *landmarks* que cada curva debería tener. Finalmente esta aplicación generaría los *landmarks* a lo largo de las curvas que describen las características importantes del rostro.

Si se incrementaran los *landmarks* mayor sería el número de cálculos computacionales, aunque es cierto que se mejora la correspondencia (por lo

tanto los resultados del traslado serán mejores), el tiempo en que se ejecuta es muy importante, así que es preferible que el número de *landmarks* sólo sea de 44. Finalmente la calidad del *rig* de los modelos objetivo depende de que también se haya realizado el *rig* inicial, en otras palabras el modelo fuente se convertiría en el molde para las expresiones faciales de los demás.

# Capítulo 5

## Conclusión

*“Tu rostro, mi barón, es como un libro, donde los hombres  
pueden leer cosas extrañas”*

*William Shakespeare.*

En este trabajo se describió el proceso de hacer las expresiones faciales de un rostro virtual, así como el método que automatiza el *rig* facial en base a un modelo con previo *rig*, un método de deformación TPS y seleccionar los *landmarks*. Con los resultados mostrados en el capítulo cuatro es clara la ventaja de este sistema, pues reduce tiempo de producción, aunque este sistema aun es el principio de una mejor y más eficiente solución.

### 5.1. El sistema

Las seis horas en promedio que requirió el autor de esta tesis para realizar un *shape*, quizá no se aproximen al tiempo que un modelador experimentado pudiera requerir, pero lo que sí es seguro es que el tiempo requerido para crear todo el *rig* es suficiente para tratar de optimizar su elaboración.

A lo largo de los capítulos tres y cuatro se presentó un sistema que reutiliza los *blend shapes*, la técnica los aplica y ajusta en diferentes modelos, creando así el *rig* de los personajes en un menor tiempo que si se hicieran a mano los *blend shapes*.

Este pequeño sistema puede ser usado en cualquier producción que necesite personajes animados, debido a que los nuevos *blend shapes* de los personajes son de la misma calidad y consistencia artística del modelador, por

lo que los animadores pueden hacer su trabajo en menor tiempo, con menos recursos y que se ajusten a sus consideraciones.

Al final se mostró que el sistema funciona con modelos caricaturescos. El *rig* que el sistema generó siempre será la adaptación del *rig* original a la topología del los modelos objetivo, brindando así, expresiones faciales que realmente correspondan a su imagen. Por último para comprobar los resultados o la salida del sistema en ambos tipos de modelos, se compararon las mismas expresiones hechas por el modelador.

Observando los resultados se aprecia que la calidad depende de la correcta colocación de los *landmarks* y sobre todo de qué tan bien se hicieron los *blend shapes* del modelo inicial; es así que con el sistema el modelador puede emplear las horas que utilizaría en el *rig* de los demás personajes para mejorar el *rig* del modelo inicial. Como última observación si la calidad del *rig* no fue lo suficientemente buena, es posible realizar pequeñas correcciones al *rig* que regresó el sistema, aunque esto debería ser opcional.

## 5.2. Mejoras del *rig*

Existen otro tipo de atributos que pueden ser parte del *rig* y pueden ser asignados a los modelos para poder no sólo controlarlos sino que pueden ayudar a las animación o modelar cierta expresión facial:

**El esqueleto** es un grupo de “huesos” que están posicionados debajo de la piel. Estos huesos definen la pose de la cabeza y controlan a bajo nivel la deformación de la superficie.

**Objetos de influencia** estos objetos afectan la forma de la malla y ayudan a los animadores a controlar los modelos en 3D. Este tipo incluye: superficies NURBS, curvas NURBS, deformadores *lattice*, deformadores *cluster*, mallas de polígonos, entre otros.

**Scripts de animación** consisten en una lista de curvas de animación que determinan el movimiento del objeto. Cada curva de animación representa algún cambio en el valor de algún atributo ( *blend shapes* o huesos).

Todos estos atributos tienen algo en común: deben ser construidos de nueva cuenta para cada uno de los personajes que estén dentro de la animación, por lo que aún el tiempo de producción es considerable.

Otro punto en común y de suma importancia es que cualquiera de los anteriores atributos está sumamente ligado con los vértices del modelo, incluso estos atributos pueden estar hechos a base de vértices del espacio virtual.

### **5.2.1. Trabajo futuro**

El método y los resultados expuestos a lo largo de los capítulos, muestran la gran ayuda que ofrece el sistema a los animadores, pero como ya se observó aun quedan otros atributos que ayudan a definir un mejor *rig*; el sistema podría considerarse que es apenas un punto de partida.

Como se mencionó los demás atributos tienen una relación con los vértices del mallado de los rostros de los personajes; dicho eso es muy posible crear un mejor sistema que traslade todos estos atributos a cualquier modelo objetivo; ése es el siguiente paso para desarrollar.

# Bibliografía

- [Bal04] Balci,K. Xface: Mpeg-4 based open source toolkit for 3d facial animation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 399–402. ACM Press, New York, NY, USA, 2004.
- [Bir70] Birdwhistell, R. *Kinesics And Context*. University Of Pennsylvania Press, 1970.
- [Bla03] Blanz,V.,Basso,C.,Poggio,T. y Vetter,T. Reanimating faces in images and video. In *In EUROGRAPHICS 2003 the European Association for Computer Graphics, 24th Annual Conference*, pages 641–650. Blackwell, Granada, Spain, P. Brunet and D. W. Fellner, Eds., 2003.
- [Boo89] Bookstein,F.L. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 567–585, 1989.
- [Boo97] Bookstein,F.L. *Morphometric Tools for Landmark Data*. Cambridge University Press, 1997.
- [Bor06] Borshukov,G., Montgomery,J. y Werner,W. Playable universal capture: compression and real-time sequencing of image-based facial animation. In *In SIGGRAPH '06: ACM SIGGRAPH Courses* , page 8. ACM Press, New York, NY, USA, 2006.
- [Bul49] Bulwer,J. *Pathomyotomia*. Printed by W.W. for H. Moseley, 1649.
- [Cha96] Chatters, J.C. *Ancient Encounters: Kennewick Man and the First Americans*. NEW YORK SIMON & SCHUSTER 2001, 1996.

- [Cha03] Chai, J., Xiao, J. y Hodgins, J. Vision-based control of 3d facial animation. *Eurographics Association*, 2003.
- [Cho05] Choe, B. y Ko, H. Analysis and synthesis of facial expressions with hand-generated muscle actuation basis. *ACM Press, New York, NY, USA*, 2005.
- [Chu02] Chuang, E. y Blegler, C. Performance driven facial animation using blend shape interpolation. Technical report, Stanford University, 2002.
- [Coh93] Cohen, M.M. y Massaro, D.W. Modeling coarticulation in synthetic visual speech. *In Models and Techniques in Computer Animation*, pages 139–156, 1993.
- [Col08] Coleman, Loren. The Ainu and The Kennewick Man, December 2008. <http://www.cryptomundo.com/cryptozoo-news/ainu-kennewick/>.
- [Coq90] Coquillart, S. Extended free-form deformation: a sculpturing tool for 3d geometric modeling. *In SIGGRAPH 90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 187–196. ACM Press, New York, NY, USA, 1990.
- [Dar02] Darwin, C. *The Expression of the Emotions in Men and Animals*. Oxford University Press, 3 edition, 2002.
- [Ekm72] Ekman, F. *Universals and cultural differences in facial expressions of emotion*. University of Nebraska Press., 1972.
- [Ekm73] Ekman, F. *Darwin and facial expression; a century of research in review*. New York: Academic Press., 1973.
- [Ekm78] Ekman, P. y Friesen, W. *Facial Action Coding System*. Consulting Psychologist Press, Palo Alto, CA, USA, 1978.
- [Far87] Farkas, L. y Munro, I. *Anthropometric facial proportions in medicine*. Charles Thomas Ltd., 1987.
- [Fid00] Fidaleo, D., Noh, J-Y., Kim, T., Enciso, R. y Neumann, U. Classification and volume morphing for performance-driven facial animation.

In *In International Conference on Digital and Computational Video*, 2000.

- [Gle98] Gleicher, M. Retargetting motion to new characters. In *In SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 33–42. ACM Press, New York, NY, USA, 1998.
- [Gue98] Guenter, B., Grimm, C., Wood, D., Malvar, H. y Pighin, F. Making faces. In *In SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 55–66. ACM Press, New York, NY, USA, 1998.
- [Gue99] Guerrero Zarco, María de Lourdes. Síntesis y modelo de expresiones faciales basado en un modelo muscular. Master's thesis, UNAM, 1999.
- [Jas02] Jason Osipa. *Stop Staring: Facial Modeling and Animation Done Right*. Sybex, 2002.
- [Kah03] Kahler, K. *A Head Model with Anatomical Structure for Facial Modeling and Animation*. PhD thesis, Universita et des Saarlandes, 2003.
- [Kel] Kelsey, Sandra. Face Up Exercise. <http://www.face-up-exercise.com/facialmuscles.htm/>.
- [Kla06] Klaran, R., Oleg, A. y Karim, B. *The Art of Rigging*. CG Toolkit, 2006.
- [Koe02] Koenen, R. Overview of the mpeg-4 standard, 2002. <http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm>.
- [Laz94] Lazarus, F., Coquillart, S. and Jancene, P. Axial deformations: an intuitive deformation technique. *Computer-Aided Design*, pages 607–613, 1994.
- [Lee95] Lee, Y., Terzopoulos, D. y Walters, K. Realistic modeling for facial animation. In *In SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 55–62. ACM Press, New York, NY, USA, 1995.

- [Noh01] Noh, M. y Neumann, U. Expression cloning. In *In SIGGRAPH '01: Proc. of Conf. on Computer graphics and interactive techniques*, pages 277–288. ACM Press, New York, NY, USA, 2001.
- [Pan02] Pandzic, I.S. y Forchheimer, R. *MPEG-4 Facial Animation*. John Wiley & Sons Ltd, 2002.
- [Pan03] Pandzic, I.S. Facial motion cloning. *Graph Models*, pages 385–404, 2003.
- [Par72] Parke, F.S. Computer generated animation of faces. In *In ACM 072: Proc. of the ACM annual Conf.*, pages 451–457. ACM Press, New York, NY, USA, 1972.
- [Par74] Parke, F.,. *A Parametric Model for Human Faces*. PhD thesis, A Parametric Model for Human Faces, 1974.
- [Par82] Parke, F. Parametrized models for facial animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, 1982.
- [Pig98] Pighin, F., Hacker, J., Lischinski, D., Szeliski, R. y Salesin, D.H. Synthesizing realistic facial expressions from photographs. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 75–84. ACM Press, New York, NY, USA, 1998.
- [Pig06] Pighin, F. y Lewis, J.P. Facial motion retargeting. In *In SIGGRAPH 06: ACM SIGGRAPH Courses*. ACM Press, New York, NY, USA, 2006.
- [Pla81] Platt, S.M. y Badler, N.I. Animating facial expressions. *ACM Press, New York, NY, USA*, pages 245–252, 1981.
- [Sag06] Sagar, Mark. Facial performance capture and expressive translation for king kong. In *In SIGGRAPH 06: ACM SIGGRAPH Courses*. ACM Press, New York, NY, USA, 2006.
- [Sed86] Sederberg, T. y Parry, S. Free-form deformation of solid geometric models. In *In SIGGRAPH 86: Proc. of Conf. on Computer graphics and interactive techniques*, pages 151–160. ACM Press, New York, NY, USA, 1986.

- [Sif05] Sifakis, E. , Neverov, I. y Fedkiw, R. Automatic determination of facial muscle activations from sparse motion capture marker data. *In Models and Techniques in Computer Animation*, pages 417–425, 2005.
- [Sif06] Sifakis, E., Selle, A., Robinson-Mosher, A. y Fedkiw, R. Simulating speech with a physics-based facial muscle model. In M. Cani, editor, *In ACM SIGGRAPH/Eurographics Symp. on Computer Animation*, pages 261–270. Eurographics Association, 2006.
- [Tei07] Teixeira Orvalho, Verónica Costa. *Reusable Facial Rigging and Animation: Create Once, Use Many*. PhD thesis, Universitat Politècnica de Catalunya, 2007.
- [Ter81] Terzopoulos, D. y Waters, K. Physically-based facial modelling, analysis and animation. *The Journal of Visualization and Computer Animation*, 1-2:73–80, 1981.
- [Wat87] Water, K. A muscle model for animation three-dimensional facial expression. *ACM Press, New York, NY, USA*, pages 17–24, 1987.
- [Wat89] Water, K. Modeling 3d facial expressions: Tutorial notes. In *SIGGRAPH '89 Course Notes 22: State of the Art in Facial Animation*, pages 127–152, 1989.
- [Wil90] Williams, L. Performance-driven facial animation. In *In SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 235–242. ACM Press, New York, NY, USA, 1990.
- [Zha04] Zhang, L., Snavely, N., Curless, B. y Seitz, S.M. Spacetime faces: high resolution capture for modeling and animation. In *In SIGGRAPH '04: ACM SIGGRAPH 2004 Papers* , pages 548–558. ACM Press, New York, NY, USA, 2004.