



# Animação Facial

por

**Marcos Paulo Moreti**

**Orientador: Roberto Marcondes Cesar Junior**

**Monografia baseada em Iniciação Científica realizada em 2006**

## **Agradecimentos**

Aos meus pais, que sempre me deram todo o apoio necessário durante o curso.

Ao meu orientador Roberto, por fornecer o apoio necessário durante o projeto.

Ao Yossi, por tirar minhas dúvidas.

Aos artistas Maria Hsu e Ricardo Barreto, pelo excelente trabalho realizado.

A todos os professores e colegas do IME, com os quais estive junto durante o curso.

# Índice

|       |                                                   |    |
|-------|---------------------------------------------------|----|
| 1     | Introdução .....                                  | 1  |
| 1.1   | Cinema .....                                      | 1  |
| 1.2   | Jogos Eletrônicos .....                           | 1  |
| 1.3   | Websites interativos e televisão .....            | 2  |
| 1.4   | Auxílio a deficientes .....                       | 2  |
| 1.5   | Arte digital .....                                | 2  |
| 2     | Conceitos e técnicas .....                        | 3  |
| 2.1   | Criação dos modelos tridimensionais .....         | 3  |
| 2.1.1 | Manualmente .....                                 | 4  |
| 2.1.2 | Scanner 3D .....                                  | 4  |
| 2.1.3 | A partir de fotos .....                           | 4  |
| 2.2   | Animação .....                                    | 5  |
| 2.2.1 | Interpolação de poses-chave .....                 | 5  |
| 2.2.2 | Baseada em performance .....                      | 5  |
| 2.2.3 | Parametrização geométrica .....                   | 5  |
| 2.2.4 | Baseada em modelos físicos .....                  | 6  |
| 2.3   | Sincronização entre os visemas e a fala .....     | 6  |
| 2.4   | Expressões faciais e a linguagem não-verbal ..... | 6  |
| 3     | Implementação do sistema .....                    | 8  |
| 3.1   | Modelos tridimensionais .....                     | 8  |
| 3.2   | Script .....                                      | 10 |
| 3.3   | Deformações da face virtual .....                 | 12 |
| 3.4   | Animação da face .....                            | 13 |
| 3.5   | Resultados obtidos .....                          | 14 |
| 4     | Arte digital .....                                | 17 |
| 4.1   | Funcionamento do software.....                    | 17 |
| 4.2   | Arquitetura do software .....                     | 18 |
| 4.2.1 | Cliente .....                                     | 18 |
| 4.2.2 | Script AIML .....                                 | 18 |
| 4.2.3 | Servidor .....                                    | 19 |
| 4.2.4 | Criação das faces virtuais .....                  | 19 |
| 4.2.5 | Animação facial .....                             | 19 |
| 5     | Conclusão .....                                   | 20 |
| 6     | Parte subjetiva .....                             | 21 |
| 6.1   | Desafios e frustrações .....                      | 21 |
| 6.2   | Disciplinas importantes do curso .....            | 21 |
| 6.2.1 | Princípios de Desenvolvimento de Algoritmos ..... | 21 |
| 6.2.2 | Laboratório de Programação II .....               | 21 |
| 6.2.3 | Programação para Redes de Computadores .....      | 22 |
| 6.2.4 | Introdução à Computação Gráfica .....             | 22 |
| 6.3   | Interação com membros da equipe .....             | 22 |
| 6.4   | Futuro .....                                      | 22 |
| 7.    | Referências bibliográficas .....                  | 23 |

# 1 Introdução

Animação de faces humanas é um tópico de pesquisa que vem sendo explorado desde a década de 70. Mesmo assim, ainda há muito a ser pesquisado. Nós estamos acostumados a observar rostos humanos mais do que qualquer outra coisa, por isso fica extremamente difícil criar rostos sintéticos que pareçam realistas em todos os aspectos [7]. Este trabalho pretende mostrar os conceitos e técnicas usados na área, assim como desenvolver um sistema de animação facial simplificado.

Sistemas de animação facial são utilizados no cinema, na indústria de jogos eletrônicos, na criação de websites interativos e até mesmo na educação especial, auxiliando pessoas com deficiências auditivas e com problemas no reconhecimento de expressões faciais, por exemplo.

## 1.1 Cinema

No cinema, animação facial é utilizada tanto para faces realistas, que se assemelham ao rosto humano, como para criaturas ou objetos inanimados, que ganham vida, falam e expressam sentimentos.

No filme “Toy Story”, produzido em 1995 pela Pixar, bonecos criados em computadores falam, riem, ficam tristes, enfim, fazem tudo o que uma face humana pode fazer. Os movimentos da boca são sincronizados com a fala, dublada por artistas. Este filme se tornou um marco na história por ser a primeira animação totalmente produzida por computador.

Atualmente, com a evolução da animação facial, é possível criar filmes inteiros por computador, com faces extremamente realistas. É o caso de “Final Fantasy – The Spirits Within”, filme criado pela produtora japonesa “Square Pictures” em 2001. O realismo é tão grande que, às vezes, torna-se difícil perceber que os personagens não são artistas reais.

Esta tecnologia permite também fazer coisas bizarras e polêmicas, como colocar um ator falecido em um filme. Isso foi feito em “Capitão Sky e o Mundo do Amanhã”. O ator Laurence Olivier, falecido em 1989, “aparece” em uma cena do filme, interpretando o vilão da história.

## 1.2 Jogos eletrônicos

Animação facial é um requisito fundamental para a grande maioria dos jogos produzidos hoje. Graças ao desenvolvimento de hardware específico para renderização de cenas 3D, é possível colocar modelos 3D complexos, como a face humana, por exemplo, dentro dos jogos.

Porém, aqui, o desafio é maior, pois, ao contrário dos filmes, as cenas têm que ser renderizadas em tempo real enquanto, num filme, pode-se levar horas para renderizar uma única cena. Em jogos, devem ser renderizadas em torno de 50 quadros por segundo. Por isso, não é possível usar modelos tridimensionais

muito complexos, ou seja, o desafio é produzir faces virtuais com o maior nível de realismo possível e com a menor complexidade possível (poucos polígonos).

Com o constante aumento do poder de processamento das Unidades de Processamento Gráfico (GPU), a tendência é que, num futuro próximo, os jogos apresentem faces virtuais cada vez mais realistas.

### **1.3 Websites interativos e televisão**

Outro uso interessante é a criação de websites que se interagem com os visitantes através de uma face virtual. É possível colocar uma face virtual para dar informações ao visitante, assim como fazê-la ler automaticamente uma notícia, usando voz sintetizada por computador. A “Guile 3D Studio”, por exemplo, colocou uma face virtual capaz de dar informações ao cliente.

Na televisão, temos um exemplo que a maioria das pessoas conhecem: é a Eva Byte, uma apresentadora virtual que aparece no programa Fantástico, da Rede Globo. É tão realista que algumas pessoas se recusam a acreditar que ela foi mesmo criada por computador.

### **1.4 Auxílio a deficientes**

Faces virtuais também podem ser usadas para auxiliar deficientes como, por exemplo, pessoas com dificuldades de audição ou pessoas com graves problemas de comunicação, como autistas.

Uma face virtual chamada “Baldi”, desenvolvida com a ajuda de um psicólogo chamado Dominic Massaro, tem este propósito. Muitas escolas dos Estados Unidos estão utilizando softwares baseados no Baldi para ajudar crianças especiais.

### **1.5 Arte digital**

Faces virtuais também podem ser utilizadas em softwares artísticos. Como exemplo, pode ser citada a criação de Ricardo Barreto e Maria Hsu, cujo desenvolvimento fez parte deste projeto de conclusão de curso. Consiste na criação de personagens virtuais capazes de interagir com o público respondendo perguntas. A “personalidade” é simulada utilizando uma linguagem própria para a criação de chatbots, chamada AIML [14]. O trabalho foi apresentado no FILE 2006 - Festival Internacional de Linguagem Eletrônica [16]. Para este trabalho, foi usado um sistema de animação facial comercial capaz de produzir falas bem realistas.

Este trabalho será explicado com maiores detalhes na seção 4.

## 2 Conceitos e técnicas

Um sistema de animação facial por computador deve ser capaz de exibir faces virtuais que falem e produzam expressões faciais convincentes de acordo com o especificado pelo usuário.

A face humana é capaz de exibir muitas expressões, mas seis delas são tão comuns que podem ser consideradas básicas. As seis expressões faciais básicas do ser humano são: alegria, tristeza, raiva, medo, nojo e surpresa. Todo sistema de animação facial deve apresentar pelo menos estas expressões. Mas, obviamente, elas são insuficientes. Por exemplo, para expressar alegria, podemos fazer um sorriso com a boca aberta ou com a boca fechada; a face humana também possui inúmeras maneiras de comunicar alguma coisa mesmo sem usar a fala como, por exemplo, franzir a testa quando se está com alguma dúvida.

Um conceito importante é o de visema. Visemas podem ser definidos como padrões visuais de movimentação articulatória observáveis na face durante a fala [4]. São os análogos visuais dos fonemas. Porém, um único visema pode ser usado para representar vários fonemas. O agrupamento de visemas para representar os fonemas depende do nível de realismo que se deseja atingir. Uma face extremamente realista usaria medições precisas para representar os visemas. Já uma face menos realista pode ter uma fala animada apenas com a boca abrindo e fechando. Vale notar que alguns fonemas são indistinguíveis pelo ser humano durante a fala apenas observando-se a parte externa da face. Por isso, um visema pode ser usado para representar mais de um fonema.

Por outro lado, temos que tomar cuidado na representação dos visemas. O ser humano, mesmo inconscientemente, consegue extrair informações lingüísticas apenas olhando para a face da pessoa que está falando [3]. Isso pode levar a um fenômeno surpreendente, porém desagradável, chamado efeito McGurk [5], que consiste no entendimento incorreto de uma mensagem falada causada pela incoerência entre a observação visual e a auditiva. Por exemplo, uma pessoa está assistindo ao vídeo no qual um locutor pronuncia a sílaba “ga”, mas o vídeo foi dublado com o áudio “ba”. Esta pessoa, então, percebe o “da”.

Veremos, a seguir, algumas técnicas para criar e animar faces virtuais tridimensionais.

### 2.1 Criação dos modelos tridimensionais

Um modelo tridimensional é uma representação matemática de um objeto em 3D. No nosso caso, estamos lidando com representações tridimensionais de faces humanas. Um modelo 3D pode ser exibido numa tela 2D através de um processo chamado renderização.

Existem várias formas de criar modelos tridimensionais de faces humanas. Vamos ver as mais comuns:

### **2.1.1 Manualmente**

Neste caso, um artista gráfico 3D utiliza algum programa de modelagem 3D para criar uma face tridimensional. A qualidade do objeto modelado depende da habilidade do artista e também das restrições impostas a ele como, por exemplo, um número limitado de polígonos para não tornar o modelo complexo demais.

### **2.1.2 Scanner 3D**

Quando é necessário fazer um modelo 3D extremamente realista de uma face humana, pode-se usar um equipamento capaz de gerar com precisão um modelo tridimensional de um objeto do mundo real, chamado scanner 3D.

Por ser um equipamento muito caro, apenas grandes instituições podem desfrutar dessa maneira de criar representações tridimensionais de faces humanas.

### **2.1.3 A partir de fotos**

Esta é a forma mais barata e mais fácil, permitindo até mesmo que pessoas sem nenhum conhecimento de modelagem 3D crie um modelo tridimensional de uma face humana.

Existem vários softwares que fazem isso. Geralmente, funcionam da seguinte maneira: o usuário fornece ao programa duas imagens ortogonais de um rosto humano, uma de frente e a outra de perfil. Então, o programa gera o modelo 3D correspondente.

Uma maneira de fazer isso é criar um modelo 3D genérico e definir nele alguns vértices-chave. Algoritmos de visão computacional analisam as imagens fornecidas e determinam, a partir delas, a posição para a qual os vértices-chaves devem ser deslocados. Após o deslocamento deste pequeno conjunto de vértices, usa-se um algoritmo que determina a posição de todos os outros vértices do modelo baseado no deslocamento dos vértices-chave. Feito isso, outro algoritmo é usado para mapear a textura para o modelo.

Vários métodos foram propostos para a modelagem automática a partir de fotos, como estes [8, 11, 13], por exemplo.

Apesar de não produzir modelos tão realistas como os criados manualmente por bons modeladores 3D ou gerados por scanners 3D, esta técnica permite fazer softwares de criação de faces tridimensionais que possam ser usados por qualquer pessoa. Por exemplo, podemos criar um jogo no qual o personagem pode ficar com o rosto do jogador, bastando para isso fornecer fotos de frente e perfil. Exemplos de software que fazem isso:

- PeoplePutty [18], usado por artistas para criar as faces virtuais para o software exposto no FILE 2006, e que será explicado nesta monografia.

- FaceGen [15], usado para gerar as faces virtuais para testar o sistema de animação facial desenvolvido durante a Iniciação Científica.

## **2.2 Animação**

A animação é uma fase extremamente importante do processo de criação de faces virtuais. De nada adianta um excelente modelo tridimensional se a animação é malfeita. Existem várias técnicas para a animação de faces virtuais. Veremos algumas delas:

### **2.2.1 Interpolação de poses-chave**

Esta é uma técnica bem simples e fácil de ser implementada. Primeiramente, devemos ter modelos tridimensionais expressando as poses desejadas (poses-chave). Para gerar poses intermediárias, basta usar um algoritmo para fazer uma interpolação.

Esta técnica foi utilizada no trabalho pioneiro de Parke em 1972 [9]. Foi também a técnica utilizada para o desenvolvimento do sistema de animação facial criado durante a Iniciação Científica.

### **2.2.2 Baseada em performance**

Técnica baseada em obtenção de dados reais para se dirigir a animação. Normalmente isso é feito com equipamentos de “motion tracking”, capazes de obter medições precisas de movimento. É utilizada principalmente por grandes empresas de produção de filmes e jogos eletrônicos.

### **2.2.3 Parametrização geométrica**

A face humana virtual é controlada por alguns parâmetros como, por exemplo, rotação da mandíbula, rotação do globo ocular, translação da língua, etc. Estes parâmetros visam imitar o funcionamento real da face humana, porém sem utilizar nenhum tipo de simulação física.

Parke, em 1982, sugeriu um método de animação facial utilizando esta técnica [10].

Parametrização geométrica é amplamente utilizada por modeladores 3D para a criação de personagens para animações e jogos.



#### **2.2.4 Baseada em modelos físicos**

Visa simular o real funcionamento da face humana através de um modelo físico, que leva em conta as ações dos músculos. Quanto mais realista o modelo, maior é a quantidade de recursos computacionais necessários para realizar a animação.

Um trabalho bem conhecido de animação facial baseada em simulação física de músculos é o de Waters [12].

### **2.3 Sincronização entre os visemas e a fala**

Um vídeo bem sincronizado com o áudio é fundamental para produzir uma boa animação facial.

O mais fácil é lidar com falas sintetizadas por computador, neste caso, a sincronização com o vídeo é trivial. Porém, vozes sintetizadas por computador não são muito realistas, nada pode substituir uma boa dublagem. Uma maneira de colocar vozes dubladas num vídeo de animação facial é ajustar os movimentos da face virtual manualmente para sincronizá-la com a fala. Porém, este é um trabalho extremamente cansativo e demorado. Para resolver este problema, existem algumas técnicas que permitem que a face virtual seja controlada através da fala, ou seja, a boca vai executando os movimentos de acordo com o que está sendo falado. A sincronização nem sempre é perfeita, então pode ser necessário fazer alguns ajustes manuais, porém o tempo de trabalho diminui muito. O People Putty, que já foi mencionado anteriormente, possui um recurso para geração de animação facial controlada pela fala.

### **2.4. Expressões faciais e a linguagem não-verbal**

A comunicação não-verbal pode ser considerada como toda forma de transmissão de informação que não utiliza palavras. Engloba gestos, expressões faciais, postura, etc. A linguagem não-verbal é essencial para as pessoas. Muitos entrevistadores são treinados para analisar o comportamento não-verbal de seus candidatos. Uma boa comunicação não-verbal também é extremamente importante para quem fala em público ou para quem está procurando um namorado ou namorada; saber expressar e interpretar sinais não-verbais pode ajudar muito.

Um estudo realizado por Mehrabian [6] concluiu que, basicamente, existem três elementos na comunicação: palavras, tom de voz e linguagem corporal. Segundo este estudo, para o entendimento de uma mensagem, as palavras utilizadas contam apenas 7%, tom de voz conta 38% e linguagem corporal conta 55%. A partir daí, podemos dizer que 93% do entendimento de uma mensagem provém da linguagem não-verbal.

Diante disso, é extremamente importante que um sistema de animação facial dê suporte à comunicação não-verbal. Por mais bem modelada e animada

que uma face virtual seja, se ela não for capaz de expressar pelo menos os sinais não-verbais mais importantes, ficará pouco realista.

Como já foi dito, existem seis expressões faciais que são consideradas básicas, são elas: alegria, tristeza, raiva, medo, nojo e surpresa. Mas existem também muitos sinais utilizados durante a fala que servem para dar um significado especial ao que está sendo dito como, por exemplo, levantar as sobrancelhas, piscar um olho propositalmente, franzir a testa, etc. Não podemos nos esquecer também dos movimentos involuntários da face humana, como piscar os olhos, por exemplo. Outra coisa que deve ser levado em consideração é que a cabeça não fica estática durante a fala, ela se movimenta, por isso não ficaria muito realista uma face virtual cuja cabeça fique completamente estática. O sistema de animação facial da Haptik leva tudo isso em consideração, sendo bastante realista. Possui diversas poses da face humana, como os visemas, as expressões faciais básicas e inúmeros sinais não-verbais.

### **3. Implementação do sistema**

Nesta seção da monografia será apresentado o principal trabalho desenvolvido durante o período da Iniciação Científica: o desenvolvimento de um sistema de animação facial.

Antes de tudo, é importante deixar claro que este sistema de animação é bem simplificado e sem muitos usos práticos possíveis. Porém, estendê-lo não é muito complicado, requer apenas mais tempo. A idéia inicial era produzir um sistema de animação facial mais completo, o que não foi possível devido ao pouco tempo disponível para trabalhar neste projeto. Mesmo assim, foi possível produzir alguns vídeos interessantes para demonstração.

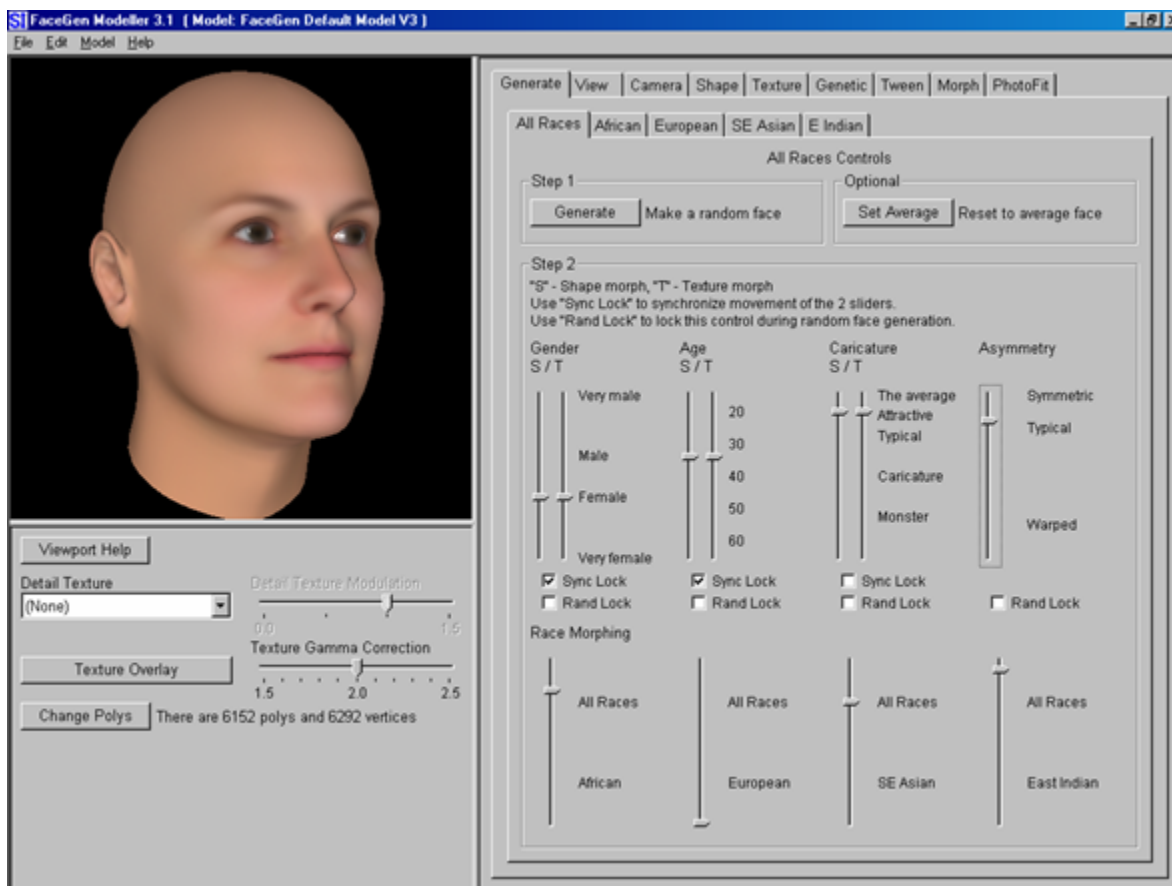
O sistema é capaz de exibir faces virtuais mostrando as seis posições básicas, que são, conforme já mencionadas anteriormente, alegria, tristeza, raiva, medo, nojo e surpresa. Também é possível fazer combinações entre elas, por exemplo, uma face virtual triste e raivosa ao mesmo tempo ou especificar poses intermediárias, como 50% triste, por exemplo.

Basicamente, o programa funciona da seguinte maneira: primeiramente, deve ser fornecido ao programa sete modelos tridimensionais de uma face, um modelo de face neutra (não exibindo nenhuma expressão) e outros seis, cada um representando uma expressão facial; feito isso, o usuário escreve um script que define uma sequência de poses que a face virtual deverá realizar ao longo do tempo; após carregar este script, pode-se configurar alguns parâmetros, como cor, intensidade e posição da fonte de luz, zoom, campo de visão da câmera (FOV), etc.; feito isso, o usuário pode salvar um vídeo no formato MPEG que correspondente ao script carregado e também pode salvar todos os quadros (também conhecidos como “frames”) deste vídeo.

A seguir, será visto com maiores detalhes o funcionamento do programa.

#### **3.1 Modelos tridimensionais**

Conforme mencionado acima, é preciso fornecer sete modelos tridimensionais. Para testar o programa, foram utilizados modelos tridimensionais de alta qualidade e bem realistas, gerados por um programa profissional de criação de faces humanas virtuais chamado FaceGen. Além de criar modelos 3D a partir de fotos ortogonais de faces humanas, este software permite alterar vários detalhes do rosto humano através de parâmetros, como tamanho do nariz, boca, etc., e permite que o modelo seja distorcido para exibir determinada expressão facial ou visema. O modelo tridimensional criado no FaceGen pode ser exportado para vários formatos conhecidos.

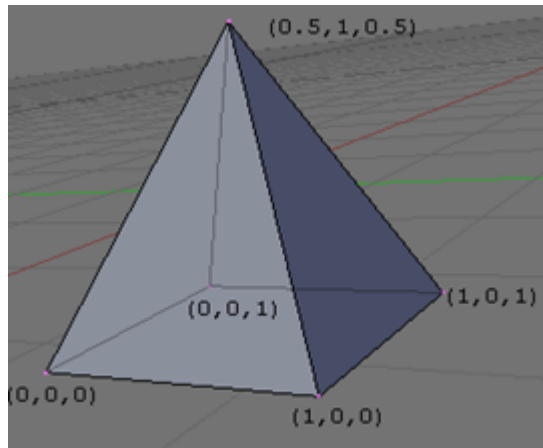


screenshot do FaceGen

Foi utilizado o FaceGen para criar os modelos tridimensionais e exportá-los para o formato de dados 3D OBJ, criado pela Wavefront.

O sistema é capaz de ler arquivos OBJ, mas não há garantias de que funcione com qualquer arquivo deste tipo. Devido a restrições de tempo, a preocupação maior foi apenas em poder ler alguns modelos 3D para realizar testes, por isso a única garantia é de que o sistema seja capaz de ler apenas modelos 3D no formato OBJ exportados pelo FaceGen. A textura deve estar no formato JPEG; não há compatibilidade com outros formatos de imagem.

Para entender como funcionam as deformações, é importante saber como a geometria de um modelo tridimensional é armazenado num arquivo. No formato OBJ, temos listas de vértices e listas de faces. Os vértices estão situados num espaço tridimensional, assim, cada vértice é representado por três coordenadas (x, y, z). Cada vértice é identificado pela posição na qual aparece. Uma face é dada por referências a três ou mais vértices. Nos modelos gerados pelo FaceGen, tem-se apenas faces de três e quatro vértices. A seguir, tem-se o exemplo de uma pirâmide:



```
v 0.0 0.0 0.0
v 1.0 0.0 0.0
v 1.0 0.0 1.0
v 0.0 0.0 1.0
v 0.5 1.0 0.5
```

```
f 1 2 3 4
f 1 2 5
f 2 3 5
f 3 4 5
f 4 1 5
```

O vértice (0.0, 0.0, 0.0) é o primeiro na ordem, portanto, é identificado como 1. (1.0, 0.0, 0.0) é o 2, e assim por diante. A face representada por “f 1 2 3 4” está fazendo referência aos quatro vértices de altura 0, portanto, é uma face de 4 vértices, quadrada, que representa a base da pirâmide. As outras faces representam o resto da pirâmide. Como pode-se notar, um vértice pode ser utilizado por muitas faces.

### 3.2 Script

Apesar do sistema não possuir uma maneira interativa de criar um script de animação, ele é extremamente simples e fácil de ser entendido. O script define uma animação. O formato geral é:

```
[instante1]:[pose1];
[instante2]:[pose2];
[instante3]:[pose3];
.....
.....
.....
.....
[instanten]:[posen];
```

Os instantes devem ser especificados na ordem crescente, assim:  
[instante<sub>1</sub>] < [instante<sub>2</sub>] < ... < [instante<sub>n</sub>]

Cada [instante<sub>i</sub>], em segundos, está relacionado com uma [pose<sub>i</sub>] da face virtual, ou seja, no instante [instante<sub>i</sub>], será exibida a pose [pose<sub>i</sub>]. Dessa forma, várias poses são especificadas. Para criar a animação, é necessário achar poses intermediárias entre dois instantes especificados. Para isso, interpolação linear é utilizada, sendo possível fazer uma animação suave.

Uma pose  $[pose_i]$  é representada por uma palavra-chave, representando uma expressão facial com uma determinada intensidade ou uma combinação delas.

O script trabalha com sete poses-chaves, que correspondem às expressões faciais básicas do ser humano mais a face neutra. A seguir, temos uma tabela mostrando como especificar estas expressões faciais no script:

| Palavra-chave utilizada no script | Expressão facial correspondente |
|-----------------------------------|---------------------------------|
| std                               | face neutra                     |
| hpy                               | face alegre                     |
| sad                               | face triste                     |
| ang                               | face raivosa                    |
| fea                               | face assustada                  |
| dgu                               | face enojada                    |
| spr                               | face surpresa                   |

Para especificar a intensidade de uma pose, usamos o símbolo “\*” (asterisco), desta forma:

$[intensidade] * [pose-chave]$

Onde intensidade é um valor numérico maior que 0 e pose-chave é uma das palavras-chave expostas na tabela anterior. Multiplicar por 0 não faz sentido, pois não haverá deformação alguma, multiplicar por 1 significa especificar o modelo tridimensional correspondente à palavra-chave. Assim:

$1 * ang$  é equivalente à  $ang$ .

Multiplicar por valores maiores que 1 causa uma deformação exagerada e pode trazer resultados indesejáveis, como uma face com uma distorção monstruosa ou completamente sem sentido.

Resultados entre 0 e 1 significam uma deformação parcial. Por exemplo:

$0.5 * sad$

Acima, temos a especificação de uma face 50% triste.

Também é possível combinar mais de uma expressão facial básica na mesma pose, utilizando o símbolo “+”, da seguinte forma:

$[intensidade_1] * [pose_1] + \dots + [intensidade_n] * [pose_n]$

Assim, teríamos uma acumulação de deformações. Exemplo:

$0.5 * Ang + sad$

Desta maneira, especificamos uma pose 50% nervosa e 100% triste.

Acumular mais de duas deformações também pode trazer resultados indesejáveis.

A palavra-chave `std` não corresponde à nenhuma deformação, apenas representa a face neutra. Por isso, não faz sentido especificar a intensidade para ela.

A seguir, temos um exemplo de um script completo:

```
0: std;  
2: 0.1*ang;  
4: 0.2*ang + sad;  
9: fea;
```

No instante 0, temos a face neutra; no instante 2, temos uma face 10% raivosa; no instante 4, a face está 20% raivosa e 100% triste (as duas deformações são acumuladas); já no instante 9, a face está 100% assustada. Desta forma, definimos uma animação de 9 segundos na qual a face faz transições suaves entre as 4 poses especificadas no script.

Nas seções a seguir, será visto em detalhes como são feitas as deformações e o que elas representam, e também o cálculo para determinar as posições intermediárias do script.

### 3.3 Deformações da face virtual

Nesta seção será explicado em detalhes como o sistema calcula as poses especificadas no script. Conforme já foi dito, o sistema trabalha com sete modelos 3D fornecidos pelo usuário. Cada um é representado por uma palavra-chave no script. Portanto, uma face `ang`, por exemplo, corresponde ao modelo 3D da face raivosa, fornecido pelo usuário. Agora será explicado como fazer para achar posições mais complexas, como `0.2*ang`, `ang + sad`, por exemplo.

Primeiramente, para o sistema funcionar, as faces de todos modelos 3D devem ser as mesmas, ou seja, devem fazer referências aos mesmos vértices. O que muda de um modelo para outro são os vértices. Assim, todos os modelos devem ter o mesmo número de vértices e de faces.

Cada um dos seis modelos 3D correspondentes às expressões faciais básicas, representados por `hpy`, `sad`, `ang`, `fea`, `dgu` e `spr`, são comparados em relação ao modelo 3D da face neutra. A comparação é feita vértice a vértice e a diferença entre eles é armazenada. Ou seja, suponha que, para dois determinados modelos, os vértices em um determinada posição sejam diferentes um do outro. Sejam  $U$  e  $V$  estes vértices, tais que  $V = (x_v, y_v, z_v)$  e  $U = (x_u, y_u, z_u)$ . Suponha também que  $V$  seja o vértice de um modelo da face neutra e  $U$  de uma deformação. Assim, a diferença é representada por  $U - V = (x_u - x_v, y_u - y_v, z_u - z_v)$ .

Dessa forma, a geometria de cada expressão facial é representada por um conjunto de diferenças entre os vértices da face neutra e os da expressão facial em questão.

Assim, criar uma deformação do tipo [intensidade]\*[pose] consiste em multiplicar todas as coordenadas (x, y, z), que representam a diferença entre a pose em questão e a face neutra, pela intensidade e depois somar estas coordenadas resultantes às coordenadas correspondentes no modelo da face neutra. Ou seja: suponha que V seja um vértice do modelo da face neutra e U o vértice correspondente no modelo deformado. O vértice correspondente no modelo resultante será:

$$V + [\text{intensidade}] * (U - V)$$

Para acumular mais de uma deformação, a idéia é a mesma: suponha que V seja um vértice do modelo da face neutra,  $U_1, \dots, U_n$  os vértices correspondentes nas n deformações que serão aplicadas. Assim, o vértice resultante será:

$$V + [\text{intensidade}_1] * (U_1 - V) + \dots + [\text{intensidade}_n] * (U_n - V)$$

Como pode ser visto, as deformações que resultam nas expressões faciais especificadas no script são criadas através de um processo muito simples. Mesmo assim, os resultados são convincentes, só não se pode exagerar na intensidade das deformações.

### 3.4 Animação da face

Através do processo descrito acima, conseguimos encontrar o modelo 3D correspondente a cada pose especificada no script. Resta agora saber como proceder para achar as poses intermediárias (que estão num instante de tempo não especificado no script).

A idéia é semelhante à da seção anterior: usar interpolação linear entre duas poses adjacentes especificadas no script. Seja  $t_i$  um instante de tempo no qual o modelo com um vértice V está especificado e seja  $t_j$  um instante de tempo no qual o modelo com o vértice U, correspondente a V (na mesma posição), está especificado. O objetivo é achar o vértice correspondente ao modelo do instante  $t_k$ , sendo que  $t_i < t_k < t_j$ . Assim, o vértice a ser encontrado é:

$$V + [(t_k - t_i) / (t_j - t_i)] * (U - V)$$

Realizando este procedimento com todos os vértices, conseguimos encontrar o modelo 3D no instante  $t_k$ .



### 3.5 Resultados obtidos

Apesar de não ter havido tempo para fazer um sistema de animação facial mais completo, os resultados obtidos foram interessantes. Para testar o programa, conforme já foi mencionado, foram utilizados modelos tridimensionais gerados pelo FaceGen.

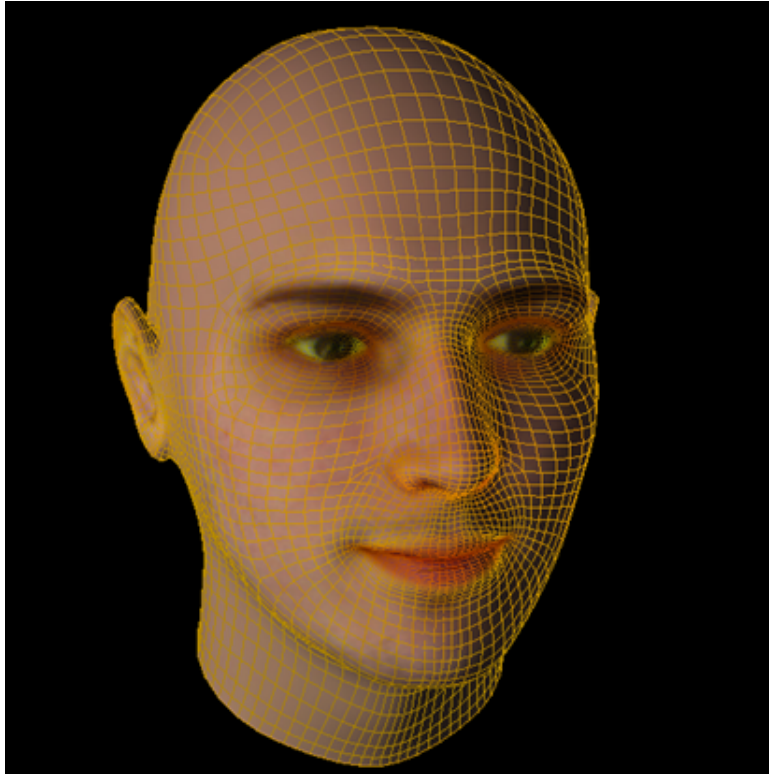
Um exemplo de um vídeo gerado pelo sistema pode ser baixado na página da monografia. Segue abaixo o script usado para gerar o vídeo:

```
0: std;  
0.999: ang;  
1: std;  
1.999: sad;  
2: std;  
2.999: hpy;  
3: std;  
3.999: spr;  
4: std;  
4.999: fea;  
5: std;  
5.999: dgu;
```

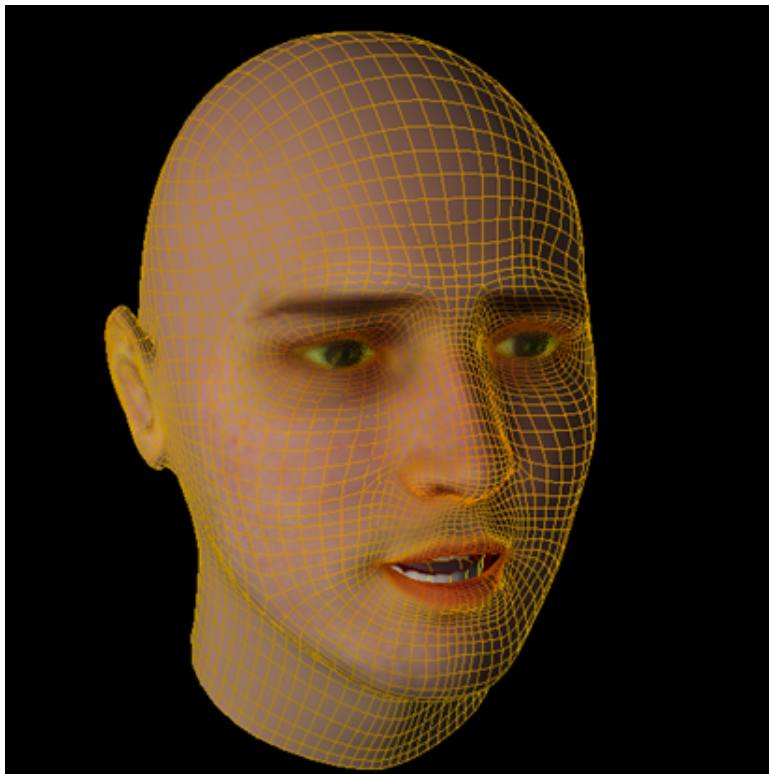
Este modelo foi gerado utilizando o recurso de transformação de duas fotos ortogonais em um modelo 3D do FaceGen.

A geração de vídeo é feita através de uma biblioteca extremamente simples, mas que funciona tanto no Linux como no Windows, chamada ezMPEG. Além de gerar o vídeo, o programa pode gerar todas os quadros (frames) deste vídeo no formato de imagem PPM, que não possui nenhuma compressão.

A seguir temos algumas screenshots, mostrando inclusive a wireframe, que é uma espécie de conjunto de arames que compõem o modelo 3D:



face alegre



face triste

Mais uma screenshot, tirada com o campo de visão (FOV) da câmera em torno de 140°:



face raivosa, FOV de aproximadamente 140°

## 4 Arte digital

Conforme já especificado na seção introdutória, foi desenvolvido um software para fins artísticos que possui personagens virtuais capazes de se comunicar com o público de uma maneira semelhante a um chatbot. Chatbot é um programa que consiste em simular uma conversa inteligente com um ou mais seres humanos. A maioria dos chatbots disponíveis na Internet fornecem comunicação com usuários apenas por texto. O trabalho descrito nesta seção vai além, pois os personagens possuem uma face virtual e se comunicam através de um vídeo, no qual falam de maneira realista. Outra diferença dos chatbots normais é que este trata de assuntos bem específicos, respondendo através de citações do filósofo Nietzsche.

### 4.1 Funcionamento do software

O software foi criado para funcionar em dois computadores ligados em rede. Ambos os computadores exibem faces virtuais nos seus respectivos monitores. O usuário digita uma frase para se comunicar com o personagem virtual, uma face virtual repete a pergunta e, após o término da mesma, a outra face virtual responde com a frase selecionada pelo componente responsável pela inteligência artificial.

O software foi apresentado no Festival Internacional de Linguagem Eletrônica realizado em 2006. Esta exposição de arte é específica para trabalhos de arte digital.

Durante a exposição do trabalho, o público podia ver dois monitores: o da esquerda exibindo a face virtual que repetia a pergunta e o da direita exibindo a face virtual que dava a resposta. Apenas o teclado do computador da esquerda estava disponível ao público. Através dele as frases eram digitadas.



a face da esquerda repete a frase digitada e a face da direita responde

## 4.2 Arquitetura do software

Este software possui três componentes principais: um cliente, que recebe a frase digitada do usuário; um script escrito em AIML, que é responsável por analisar a frase e escolher a resposta; e um servidor, que fica aguardando um sinal avisando que a frase digitada já foi pronunciada e, ao receber este sinal, responde ao usuário, ou seja, pronuncia a frase escolhida pelo script AIML.

### 4.2.1 Cliente

O programa cliente é o que coleta as frases digitadas pelo usuário e as envia para serem analisadas pelo script AIML, responsável pela inteligência artificial do personagem virtual.

Pelo fato de o usuário possuir acesso ao teclado, algumas medidas de segurança foram tomadas para evitar que o mesmo cometesse algum erro ou ato proposital que causasse a finalização do programa. As teclas responsáveis pelo suporte à multitarefa do sistema operacional foram desabilitadas via software, impedindo que o usuário minimizasse o programa ou alternasse para outra janela ativa; os comandos importantes que permitiam alterar certas características, como o plano de fundo, por exemplo, tinham nomes complexos e impossíveis de serem digitados ao acaso por um usuário.

### 4.2.2 Script AIML

AIML é uma linguagem criada para fazer chatbots. É extremamente simples e fácil de ser aprendida e pode ser usada até mesmo por pessoas que não têm nenhuma noção de programação. Isso pôde ser verificado na prática, já que as pessoas responsáveis pela criação do script, cujo objetivo era imitar Nietzsche, são apenas artistas e não possuíam nenhuma formação na área de Computação.

A tarefa de criar um chatbot através da linguagem AIML, embora requeira apenas conhecimentos básicos de Computação, é extremamente longa e cansativa. Não pela dificuldade em usar a linguagem, mas pela dificuldade intrínseca de criar uma entidade virtual capaz de responder a frases que serão digitadas pelos usuários sem qualquer critério, ou seja, podem ser sobre qualquer assunto.

O script foi criado pelos artistas Maria Hsu e Ricardo Barreto. Por causa da falta de tempo, não foi possível criar um chatbot muito extenso capaz de lidar com muitos assuntos mas, mesmo assim, os resultados foram convincentes e algumas pessoas ficaram intrigadas ao ver o software em funcionamento:

*“Dá a impressão de que é uma pessoa falando com a gente mesmo, de carne e osso.”*, comentário de um vendedor ambulante feito durante uma reportagem sobre a exposição [1].

*“Por exemplo: uma personagem fica só observando quem passa. Ela, por assim, dizer, pensa.”*, comentário do repórter Fabio Turci em outra reportagem sobre a exposição [2].

#### **4.2.3 Servidor**

O programa servidor é responsável por dar a resposta, recebida do programa cliente. Ele fica aguardando um sinal, indicando o término da pronúncia da frase pelo programa cliente. Ao terminar, o programa cliente estabelece uma conexão com o programa servidor e envia a resposta.

Durante a exposição, os usuários não tiveram acesso ao teclado do computador executando o programa servidor, já que não era necessário, por isso, nenhuma medida de segurança foi tomada para proteger o funcionamento do programa.

#### **4.2.4 Criação das faces virtuais**

As faces virtuais utilizadas neste software foram criadas por Maria Hsu e Ricardo Barreto. Para isso, eles utilizaram o PeoplePutty, conforme já foi mencionado.

Assim como o FaceGen, o PeoplePutty também possui a funcionalidade de criar modelos 3D a partir de fotos e possui alguns parâmetros a serem configurados para alterar diversas partes do rosto humano.

Ele exporta os modelos para o formato proprietário da Haptik, que pode ser utilizado pelo Haptik Player [17] para produzir as animações desejadas.

#### **4.2.5 Animação facial**

O sistema de animação facial utilizado foi o da Haptik, que é capaz de produzir animações bem realistas de rostos humanos. A face virtual carregada não fica estática na tela, ela se move, pisca os olhos, aumentando bastante o realismo.

É extremamente fácil de ser usado. Basta fornecer ao player alguns comandos bem simples para que ele faça o que você deseja. O player vem na forma de um controle OCX, que pode ser incorporado a aplicações que funcionem sobre Windows.

O objetivo inicial, conforme descrito na proposta, era fazer o software funcionar com o sistema de animação facial desenvolvido durante a Iniciação Científica, porém isso não foi possível, pois o mesmo não ficou pronto, não sendo capaz de produzir animações com fala.

## 5. Conclusão

Animação facial é um tópico de pesquisa ainda bastante ativo. Técnicas cada vez mais complexas vem sendo descobertas e o aumento na capacidade de processamento do hardware dá suporte a essas técnicas.

A indústria milionária do cinema, por exemplo, utiliza técnicas avançadas para produzir resultados cada vez mais impressionantes, tanto para imitar faces humanas realistas como para criar faces caricaturadas ou seres fantasiosos.

Porém, através dos trabalhos desenvolvidos durante a Iniciação Científica, foi possível perceber que utilizar um sistema de animação facial é algo muito simples e que pode ser feito mesmo por pessoas que não são especialistas na área. Softwares como FaceGen ou PeoplePutty permitem a criação de complexos modelos tridimensionais de faces humanas por pessoas que não tenham nenhum conhecimento de modelagem 3D. Sistemas de animação facial, como o Haptik Player, podem ser utilizados por pessoas que conheçam apenas um pouco de programação.

Desenvolver um sistema de animação facial também não é uma tarefa muito complexa. Através de técnicas bem simples, como a interpolação de poses-chave, utilizada no sistema desenvolvido durante o ano, é possível criar animações bem realistas.

Enfim, a animação facial por computador, hoje, é algo que está ao alcance de todos que possuem conhecimentos básicos de Computação.

## **6. Parte subjetiva**

Esta seção da monografia é pessoal, onde descreverei minha experiência ao participar de um projeto de Iniciação Científica.

### **6.1 Desafios e frustrações**

Antes do início do trabalho não havia muita noção de quais softwares seriam utilizados, por isso, tive que gastar algum tempo pesquisando para descobrir os softwares mais adequados. Pesquisar e avaliar ferramentas é algo um pouco complicado, pois a escolha de uma ferramenta incorreta pode atrasar muito o desenvolvimento de um projeto.

Aprender novos conceitos e tecnologias também é algo desafiador. Durante o ano, tive que pesquisar alguns conceitos de computação gráfica, animação facial e aprender a utilizar a biblioteca gráfica OpenGL.

Porém, a parte mais difícil foi conciliar os estudos das outras matérias com o desenvolvimento do projeto. Por eu ter trabalhado sozinho no desenvolvimento do sistema de animação facial, não foi possível fazer um sistema mais completo. A minha maior frustração foi a de não conseguir desenvolver tudo o que eu havia planejado no início.

### **6.2 Disciplinas importantes do curso**

Aqui serão relacionadas as disciplinas mais importantes do curso para o desenvolvimento do meu projeto de Iniciação Científica.

#### **6.2.1 MAC0122 - Princípios de Desenvolvimento de Algoritmos**

Nesta disciplina eu aprendi os conceitos básicos de algoritmos, como recursão, por exemplo, e estruturas de dados básicas como listas ligadas, árvores, assim como a importância de um algoritmo eficiente.

Sem estes conceitos básicos não seria possível desenvolver praticamente nada interessante.

#### **6.2.2 MAC0242 - Laboratório de Programação II**

Foi nesta disciplina na qual eu tive meus primeiros contatos com o desenvolvimento de compiladores. Durante o semestre tivemos um projeto que consistiu no desenvolvimento de um compilador simplificado.



### **6.2.3 MAC0448 - Programação para Redes de Computadores**

Nesta disciplina eu aprendi como criar softwares que funcionem em rede, como softwares com arquitetura cliente-servidor, como o que foi desenvolvido junto com os artistas e apresentado no FILE 2006.

### **6.2.4 MAC0420 - Introdução à Computação Gráfica**

Graças a esta disciplina, aprendi os conceitos básicos da Computação Gráfica, que foram úteis no desenvolvimento do projeto.

Além dos conceitos, a disciplina também teve exercícios práticos que me permitiram aprender a criar programas que lidem com imagens 2D e cenas 3D.

## **6.3 Interação com membros da equipe**

Apesar de ter trabalhado sozinho no desenvolvimento do sistema de animação facial, haviam reuniões semanais na qual o orientador ficava sabendo de tudo o que estava sendo feito. Sempre que precisei, tive a ajuda necessária de todos.

Quanto ao desenvolvimento do softwares de fins artísticos apresentado na exposição, trabalhei junto com os dois artistas, principalmente com Maria Hsu que, mesmo sem conhecimentos de programação, conseguiu, com a minha ajuda, fazer o script AIML para simular Nietzsche.

## **6.4 Futuro**

Gostei muito de trabalhar com animação facial e talvez eu resolva continuar o desenvolvimento do software, mas isso não está decidido.

Gosto muito de Computação Gráfica e pretendo trabalhar nesta área, não necessariamente com animação facial.

## 7 Referências bibliográficas

- [1] ARTE ELETRÔNICA. **Bom Dia Brasil**. São Paulo, Rede Globo de Televisão, 16 de agosto de 2006. Programa de TV. Disponível em: <http://bomdiabrasil.globo.com/Jornalismo/BDBR/0,,AA1249674-3682,00.html>
- [2] EXPOSIÇÕES DIVERTEM PAULISTANO NESTE FERIADO. **Jornal Hoje**. São Paulo, Rede Globo de Televisão, 8 de setembro de 2006. Programa de TV. Disponível em: <http://jornalhoje.globo.com/JHoje/0,19125,VJS0-3076-20060908-241605,00.html>
- [3] JEFFERS, J.; BARLEY, M. **Speechreading (Lipreading)**. Editora Charles C. Thomas, 1971.
- [4] MARTINO, J. M. **Animação facial sincronizada com a fala: visemas dependentes do contexto fonético para o português do Brasil**. UNICAMP, 2005. Tese (Doutorado).
- [5] MCGURK, H.; MACDONALD, J. **Hearing lips and seeing voices**. Nature 264, pág. 746-748, 1976.
- [6] MEHRABIAN, A. **Silent Messages**. Editora Wadsworth Publishing Company, 1971.
- [7] OSIPA, J. **Stop staring: facial modeling and animation done right**. Editora Sybex, 2003.
- [8] PARK, I. K.; ZHANG, H.; VEZHNEVETS, V.; CHOH, H. K. **Image-based photorealistic 3-D face modeling**. Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.
- [9] PARKE, F. I. **Computer generated animation of faces**. University of Utah, 1972. Dissertação (Mestrado).
- [10] PARKE, F. I. **Parameterized models for facial animation**. IEEE Computer Graphics and Applications, 1982.
- [11] PIGHIN, F.; AUSLANDER, J.; LISCHINSKI, D.; SALESIN, D.; SZELISKI, R. **Realistic facial animation using image-based 3D morphing**. Microsoft Research, 1997. Relatório técnico.
- [12] WATERS, K. **A muscle model for animation three-dimensional facial expression**. ACM Press, 1987.
- [13] ZHANG, K.; HUANG, Z.; CHUA, T. S. **A method for animating a customized face**. Computer Graphics International, 2003.

[14] AIML - Artificial Intelligence Markup Language. Linguagem utilizada para criação de chatbots. Página na Internet: <http://www.alicebot.org/aiml.html>

[15] FaceGen. Software de criação de faces humanas tridimensionais. Página na Internet: <http://www.facegen.com>

[16] FILE - Festival Internacional de Linguagem Eletrônica, 2006, São Paulo.

[17] Haptek Player. Software de animação facial. Página na Internet: <http://www.haptek.com/products/player>

[18] PeoplePutty. Software de criação de faces humanas tridimensionais. Página na Internet: <http://www.haptek.com/peopleputty>