

# 1

## Introdução

A animação por computador é uma área da computação gráfica que desperta, tanto no usuário quanto no desenvolvedor, o sonho e o encantamento. Com todas as suas sutilezas e princípios essa é uma área repleta de desafios (JoTh95) (Will02), que pode ser subdividida em várias outras áreas, tais como: animação de partículas, animação comportamental, simulação física, simulação de humanoides etc (Pare01). Esta tese está relacionada à animação facial e ao estudo de expressões faciais dinâmicas associadas à fala e a fatores relacionados à emoção, sub-áreas também da animação por computador.

A animação facial tem como objetivo “simular”, de forma convincente, o comportamento da face humana ou de personagens fictícios (*cartoons*). A pesquisa e as aplicações desenvolvidas nessa área têm se intensificado bastante nos últimos anos, com várias abordagens propostas para gerar animações faciais realistas desde o trabalho pioneiro de Parke (Park74). A animação facial é atualmente um elemento importante em filmes, jogos, televisão, interfaces de comunicação com o usuário etc.; estando presente não apenas na pesquisa, mas também na indústria, principalmente de entretenimento. Os resultados deste trabalho foram aplicados em ambientes de jogos 3D, *storytelling* (Rodr06) e de aplicações de TV digital interativa (Rod04a) (Rod04b), podendo ser estendidos para outros domínios como ensino a distância e comércio eletrônico.

O restante deste capítulo apresenta alguns aspectos introdutórios para o melhor entendimento do trabalho. Inicialmente é apresentada uma síntese sobre o estado da arte na área de animação facial como também alguns dos principais desafios da área, seguida da descrição da parametrização tida como ideal para os sistemas de animação facial. Por fim, é dada uma visão geral das principais contribuições do trabalho proposto nesta tese e da estruturação do restante deste documento.

### 1.1

#### Estado da Arte em Animação Facial

A pesquisa em animação facial tem seu desafio na proposta de novas técnicas para melhor representação computacional dos aspectos faciais

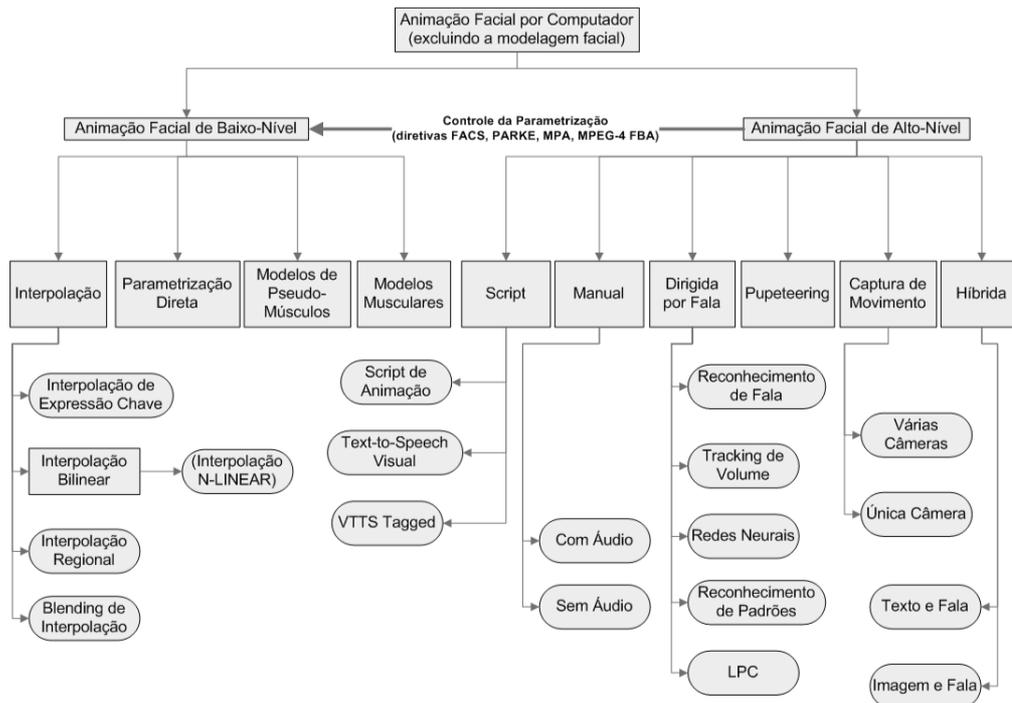


Figura 1.1: Taxonomia da Animação Facial.

humanos. Muitos pesquisadores têm estudado métodos para modelar e animar faces baseados em abordagens que vão desde o uso de modelos de pseudo-músculos (Karl92) (ThPT88) até a utilização de modelos musculares com vários graus de simulação física de ossos e da dinâmica de tecidos e de articulações (BuHN03) (PaWa96) (Wate87) (PIBa81). Algumas pesquisas baseiam-se em rastrear características faciais em tempo real em vídeos, fazendo uso ou não de marcadores faciais (caracterizando assim uma animação facial produzida a partir de dados reais) (ChXH03) (Josh03) (Zhan03) (EzGP02). Há também trabalhos que estudam a síntese de *talking heads*<sup>1</sup>, abordando questões como coarticulação e simulação de gestos naturais acompanhados da fala (Cao04) (Luce02) (KiPO00). A Figura 1.1<sup>2</sup> apresenta uma taxonomia para animação facial proposta em (PaFo02).

Como ilustrado na Figura 1.1, a animação facial pode ser abordada de duas maneiras complementares: (1) animação facial de baixo-nível e (2) animação facial de alto-nível.

A animação facial de baixo-nível caracteriza-se por trabalhar mais diretamente no modelo poligonal da face, respeitando e conhecendo a sua geometria, podendo ser subdividida em <sup>3</sup>:

<sup>1</sup>A definição de “talking heads” e de outros termos e siglas que serão usados neste documento encontra-se no sumário de notação no final deste documento.

<sup>2</sup>Figura retirada e traduzida de (PaFo02).

<sup>3</sup>Alguns detalhamentos sobre as classes de animação facial de baixo-nível podem ser encontrados em (RoFV05).

- Interpolação: a idéia consiste em coletar dados geométricos que descrevem a face em um certo número de poses diferentes. Depois um coeficiente de interpolação é aplicado a cada pose, gerando a pose resultante (movimento de um determinado vértice sem modificar a topologia da face).
- Parametrização Direta: o seu propósito é criar um modelo encapsulado que gere um intervalo grande de faces e de expressões faciais tendo como base o menor conjunto possível de parâmetros de controle.
- Modelos de Pseudo-Músculos: como a anatomia da face é bastante complexa, essa abordagem objetiva desenvolver modelos faciais que simulem as ações básicas dos músculos da face através do controle de alguns parâmetros.
- Modelos Musculares: apesar de nenhum modelo facial reportar a perfeita anatomia da face humana (PaWa96), foram desenvolvidos alguns modelos simplificados (PIBa81) (Wate87) usando a estrutura facial dos ossos, músculos, tecidos de conectividade e pele. Esses modelos possuem a habilidade de manipular expressões faciais através de uma simulação das características dos músculos.

Diferente da animação de baixo-nível, a animação de alto-nível caracteriza-se por abstrair a estrutura de definição da face. A animação não é desenvolvida diretamente sobre o modelo facial, ela é definida em uma camada de mais fácil manipulação. A partir dessa camada, e fazendo uso dos controles de parametrização, a estrutura poligonal do modelo facial é manipulada. A animação de alto-nível possui a seguinte subdivisão:

- *Script*: especifica um roteiro de ações que uma determinada face deve executar. A animação por *script* pode conter apenas uma descrição da animação, levando em consideração, ou não, o sincronismo labial (*visual text-to-speech*), e pode ser *VTTs Tagged*, onde o sincronismo labial possui marcações para modulação da voz, das expressões faciais e dos gestos.
- Manual: o animador define manualmente cada quadro da animação a ser gerada através de editores *timeline*. Esse tipo de animação de alto-nível pode possuir ou não áudio associado.
- Dirigida por Fala: caracteriza-se pelo sincronismo labial (*lipsync*) com um áudio existente. Essa categoria de animação de alto-nível ainda subdivide-se em: reconhecimento de fala, rastreamento (*tracking*) do volume facial, aprendizagem usando redes neurais, casamento de padrão e LPC (*linear predictive coefficients*).

- *Puppeteering*: é a animação dos *animatronics* (robôs animados)<sup>4</sup> estando os seus controles no hardware.
- Captura de Movimento (*Motion Capture*): baseia-se em capturar os movimentos reais de um ator, podendo essa captura ser feita com uma ou múltiplas câmeras. No caso de uma única câmera, a face pode possuir marcadores para ajudar na recuperação da informação capturada.
- Híbrida: une mais de um tipo de entrada de dados para a produção da animação. É possível ter uma animação facial híbrida usando texto e fala ou usando imagem e fala.

O trabalho proposto nesta tese caracteriza-se por utilizar uma animação facial de baixo-nível do tipo “parametrização direta”. Como será discutido ao longo deste documento, o modelo facial usado é baseado em uma face tridimensional que segue a especificação do padrão MPEG-4 (Capítulo 4). Sobre essa face são definidos pontos de controle que, a partir do estado de neutralidade da face, deslocam a posição dos vértices gerando novas poses.

Do ponto de vista de animação facial de alto-nível, este trabalho pode ser classificado como manual, porém também fazendo uso de *scripts*. Na etapa inicial de execução, o usuário pode fornecer *scripts* que definem o curso da animação. Adicionalmente, através de uma ferramenta de edição, a animação pode ser refinada de forma manual e interativa pelo usuário. A partir da definição na abordagem de alto-nível, os parâmetros de baixo-nível são determinados.

## 1.2

### Desafios da Área

Analisando o comportamento humano é possível identificar várias nuances. A forma de pensar e agir de cada ser humano é algo muito particular e dinâmico, não seguindo uma regra específica. Vários fatores estão envolvidos quando uma pessoa demonstra alguma emoção através de uma expressão facial, como, por exemplo, humor, personalidade, se a pessoa está sozinha, se a pessoa está falando ou ouvindo (papel na conversação) etc.

Estudos mostram que se duas ou mais pessoas forem submetidas à mesma situação elas podem reagir de forma bastante diferente<sup>5</sup>. Desse modo, é bastante desafiador representar o comportamento facial humano através de um modelo computacional. Infelizmente, não há ainda um sistema que

<sup>4</sup>Maiores informações em <http://www.animatronics.org/> (acesso em 28/março/2007).

<sup>5</sup>Definições e testes podem ser encontrados em <http://library.thinkquest.org> (acesso em 29/março/2007).

consiga reproduzir o comportamento facial humano com exatidão. Os sistemas perdem na naturalidade da animação gerada e, conseqüentemente, no grau de realismo (convencimento) da expressão facial desejada. Quando a fala precisa ser incorporada para geração da animação, surge o desafio de como combinar as posições dos lábios (definir o visema apropriado, onde “visema” vem a ser a representação visual de um fonema) com a expressão facial a ser gerada. Para que haja um sincronismo entre os elementos que compõem a fala (fonemas) e sua representação visual (visemas), o sistema proposto nesta tese executa um processamento de fala para extração destas informações, como será descrito no Capítulo 5 deste documento.

Este trabalho aborda também um outro problema em aberto na área de animação facial: os movimentos não-verbais e não-emocionais do ser humano. O ser humano dificilmente fica totalmente parado. Seja enquanto pensa ou escuta em silêncio, seja enquanto fala, movimentos corporais e faciais se fazem presentes. Por exemplo, em cada uma das situações anteriores citadas há um movimento dos olhos e da cabeça, ambos podendo ser aleatórios ou direcionados (com um foco). Com isso, pode-se concluir que a geração automática da dinâmica das expressões faciais ainda é um grande desafio para área.

Outros aspectos são também importantes quando se deseja produzir a animação facial natural e convincente de uma personagem sintética. Como exemplos, é possível citar a personalidade da personagem, o humor da personagem para um determinado instante de tempo e a simulação do ambiente físico que a personagem se encontra (por exemplo, em uma reunião de trabalho e em uma festa com amigos a personagem exprime suas emoções de forma diferente). A junção desses fatores com o tratamento das expressões faciais dinâmicas em sincronismo com a fala também se constitui em uma questão desafiadora para a pesquisa na área de animação facial.

### 1.3

#### **Parametrização Ideal para Sistemas de Animação Facial**

Satisfazer todos os requisitos para um sistema de animação facial não é algo fácil. Um ponto inicial é propor requisitos para uma parametrização ideal. Uma primeira lista de requisitos foi proposta em (PaWa96) e posteriormente ampliada em (PaFo02), consistindo nos seguintes requisitos desejáveis:

- *Intervalo de possíveis faces e expressões*: idealmente, deve ser possível expressar qualquer face com qualquer expressão.
- *Facilidade de uso*: este requisito é determinado pelo número de parâmetros, a complexidade deles e o quão intuitivos esses parâmetros

são. Este requisito é normalmente inversamente proporcional ao anterior: quanto mais flexibilidade se deseja nas expressões faciais, os parâmetros se tornam mais complexos.

- *Sutileza*: humanos são sensíveis a mudanças tênues de uma expressão facial. Os parâmetros devem ser capazes de expressar e controlar de forma precisa os movimentos suaves.
- *Ortogonalidade*: um parâmetro não deve ser afetado por nenhum outro parâmetro.
- *Capacidade de ser base para uma abstração de alto-nível*: é desejável que se possa controlar a face em um alto-nível de abstração, por exemplo, através do uso de textos com marcações que são automaticamente convertidos em fala com expressões faciais. Com esse propósito, a parametrização deve permitir agregação de parâmetros em controles de alto-nível como visemas e emoções.
- *Previsibilidade*: o resultado da combinação de qualquer parâmetro deve ser previsível.
- *Portabilidade*: é desejável que os mesmos parâmetros de controle dêem o mesmo resultado em termos de expressões de alto-nível quando aplicados a diferentes modelos. Isso possibilita a troca fácil de modelos faciais em aplicações e o reuso de um conjunto ou de uma seqüência de parâmetros.
- *Possibilidade de medir os parâmetros*: em várias aplicações é interessante derivar os parâmetros faciais através da medida de uma face real e de seus movimentos, e esse processo de medida deve ser preferencialmente automático. Conseqüentemente, os parâmetros devem ser passíveis de medição. Isso significa que eles devem ser descritos em unidades físicas que possam ser visualmente medidas na face.
- *Eficiência (largura de banda)*: para aplicações que envolvem comunicação de animação facial através da rede, a representação de parâmetros deve ser eficiente e possibilitar a codificação e o *streaming*.

Conhecer os requisitos para construir uma parametrização em sistemas de animação facial favorece o reuso de faces existentes assim como a interoperabilidade entre diferentes ferramentas de animação facial. Não há uma especificação de quão fiel um sistema de animação deve ser a essa parametrização, mas é importante buscar atender ao máximo os requisitos mencionados vislumbrando definir faces padronizadas. O reuso e a interoperabilidade de faces evita uma redefinição de funcionalidades nos sistemas e pode proporcionar, idealmente, um enriquecimento com extensões sobre aplicações já existentes.

## 1.4

### Contribuições Esperadas

Esta tese apresenta um estudo dos principais modelos de emoção e suas aplicações em sistemas computacionais e tem como principal contribuição a proposta de uma releitura do modelo clássico de emoções básicas definido por Plutchik (Plut80). O círculo emocional de Plutchik é generalizado permitindo novas formas de combinação das oito emoções puras propostas nesse modelo (gerando novas emoções, chamadas de emoções derivadas) e também possibilitando uma maior flexibilidade para criar novos modelos de emoção a partir de uma facilidade de reconfiguração das emoções puras para o modelo a ser criado.

O modelo de emoção proposto nesta tese chama-se *VeeM* (*Virtual emotion-to-expression Model*). O *VeeM* é um modelo onde representações de emoções são conectadas a representações de expressões faciais com a finalidade de criar personagens virtuais plausíveis e convincentes. Não é objetivo do *VeeM* ser um modelo cognitivo de emoções.

Uma outra contribuição deste trabalho para o avanço do estado da arte na área de animação é levar em consideração, na construção de um sistema desse tipo, algumas características que fazem com que cada indivíduo expresse suas emoções de forma particular. Como já mencionado nesta introdução, o humor e a personalidade do indivíduo fazem parte desse conjunto de modificadores. Além das emoções e do contexto emocional da personagem, esta tese também aborda o aspecto da fala na animação a partir de áudios previamente capturados (gravados), assim como o comportamento dos movimentos não-verbais ao longo de uma animação facial.

Esta tese adota o modelo de face definido pelo padrão MPEG-4, sendo a integração do modelo de emoção proposto com esse padrão também uma importante contribuição deste trabalho. Adotar o MPEG-4 tem como principal vantagem satisfazer os requisitos de parametrização enumerados nesta introdução. Além disso, a adoção do padrão favorece a interoperabilidade com outros sistemas e faces. Assim, uma mesma descrição de animação facial pode ser mapeada para diferentes malhas faciais automaticamente. Descrições de animações podem também ser intercambiadas independente da implementação da face.

A fim de validar a proposta, esta tese implementa uma ferramenta que integra o modelo de emoção especificado e uma face definida segundo o padrão MPEG-4. A ferramenta é denominada *DynaFeX* (*Dynamic Facial eXpression*) e é implementada como um ambiente de edição e apresentação de animações faciais, que permite que um usuário crie uma descrição em alto-nível das

emoções sincronizadas a uma fala através de *scripts*. A ferramenta também permite uma edição refinada tanto em alto-nível como em baixo-nível dos pontos de controle para a animação facial. A ferramenta foi desenvolvida com base no projeto *Xface* (Balc04). Um módulo de sincronização recebe as informações editadas e gera a animação facial sincronizada à fala da personagem, previamente gravada.

A tese também descreve dois trabalhos de integração de uma personagem virtual com expressões e fala a outras aplicações. Uma integração deu origem ao uso da face expressiva como narradora de um sistema de *storytelling* (Rodr06). O outro trabalho incorporou a face como um exibidor de mídia (*player*) de um sistema multimídia interativo (Rod04a) (Rod04b), podendo esse exibidor estar integrado a sistemas *Web*, bem como aplicações de TV digital interativa.

## 1.5

### Estrutura do Documento

Este trabalho está estruturado do seguinte modo.

O Capítulo 2 apresenta os principais modelos para representação de emoção. A emoção que um ser humano pode sentir é algo muito subjetivo e particular de cada indivíduo, conseqüentemente, “discretizar” uma emoção é algo difícil e perigoso. Na maioria das vezes, cada ser humano vivencia uma emoção diferente quando submetido à mesma situação. Devido a esses aspectos, o Capítulo 2 aborda as diferentes correntes (visões) para modelos emocionais, suas formas de representação e conceituação.

O Capítulo 3 propõe uma releitura e, conseqüente, generalização de um modelo clássico de representação das emoções. A nova abordagem é denominada *VeeM* (*Virtual emotion-to-expression Model*) e é capaz de dada uma combinação de emoções puras gerar novas emoções (emoções derivadas). Como comentado na Seção 1.4, ao *VeeM* é associada uma função de mapeamento para composição de expressões faciais a serem posteriormente visualizadas. Fenômenos afetivos, como humor e personalidade, podem compor a função de mapeamento para geração da expressão, como também a dinâmica da personagem durante um discurso ou uma conversação (podendo estar no papel de ouvinte ou falante).

O Capítulo 4 apresenta, de forma resumida, o padrão MPEG-4 para animação facial. É importante frisar que esse é o primeiro padrão existente para modelagem (definição dos pontos característicos e, conseqüente, definição da malha poligonal) e para animação. O capítulo apresenta como o padrão define e aplica sua unidade básica, chamada FAP (*Facial Animation Parameter*). A partir da definição de FAPs é possível representar expressões faciais e visemas.

Por fim, o capítulo apresenta o mecanismo necessário para gerar uma animação facial conforme com o padrão MPEG-4.

O Capítulo 5 descreve a ferramenta implementada para incorporação e validação do *VeeM*, denominada *DynaFeX* (*Dynamic Facial eXpression*). A ferramenta tem o propósito de permitir a edição e a visualização de animações faciais em uma personagem tridimensional realista especificada segundo o padrão MPEG-4. O dinamismo da *DynaFeX* é possível a partir da definição de características faciais da personagem para cada instante de tempo, podendo essa especificação emocional estar sincronizada com a fala da personagem. O capítulo também descreve os estudos de caso desenvolvidos durante o doutorado, exemplificando aplicações onde pode ser interessante a incorporação de uma personagem tridimensional realista com fala e expressão facial sincronizadas.

O Capítulo 6 resume alguns dos principais trabalhos relacionados ao trabalho proposto nesta tese. Além de oferecer uma visão geral dos trabalhos relacionados, o capítulo faz também uma análise comparativa com o modelo de emoção proposto (*VeeM*) e a ferramenta implementada (*DynaFeX*).

Por fim, o Capítulo 7 destaca as principais contribuições alcançadas com esta tese e aponta alguns trabalhos futuros que podem ser explorados.