



Paula Salgado Lucena Rodrigues

**Um Sistema de Geração de Expressões Faciais
Dinâmicas em Animações Faciais 3D com
Processamento de Fala**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Informática
do Departamento de Informática da PUC-Rio como requisito
parcial para obtenção do título de Doutor em Informática

Orientador : Prof. Bruno Feijó
Co-Orientador: Prof. Luiz Carlos P.R. Velho

Rio de Janeiro
Dezembro de 2007



Paula Salgado Lucena Rodrigues

**Um Sistema de Geração de Expressões Faciais
Dinâmicas em Animações Faciais 3D com
Processamento de Fala**

Tese apresentada ao Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio como requisito parcial para obtenção do título de Doutor em Informática. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Bruno Feijó

Orientador

Departamento de Informática — PUC-Rio

Prof. Luiz Carlos P.R. Velho

Co-Orientador

IMPA

Prof. Antonio L. Furtado

Departamento de Informática, PUC-Rio

Prof. Waldemar Celes

Departamento de Informática, PUC-Rio

Prof. Soraia Raupp Musse

Faculdade de Informática, PUCRS

Prof. Mônica Maria Ferreira da Costa

MidiArte

Prof. Lisandro Lovisolo

UERJ

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 20 de Dezembro de 2007

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Paula Salgado Lucena Rodrigues

Graduou-se como Bacharel em Ciência da Computação pela UFPE (Universidade Federal de Pernambuco) em 1999. Obteve em 2002 o Mestrado em Informática pela PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro). Ao longo do Doutorado atuou como pesquisadora dos Laboratórios TeleMídia, VisionLab e TecGraf, do Departamento de Informática da PUC-Rio, e ministra aulas no curso de Graduação em Engenharia dessa mesma instituição desde 2005. Ao longo desses anos atuou como professora em cursos para empresas como a Petrobras (2003-2005) e em cursos de Especialização no CCE PUC-Rio (2002-2003) e na Fejal CESMAC (2004-até o momento).

Ficha Catalográfica

Rodrigues, Paula Salgado Lucena

Um Sistema de Geração de Expressões Faciais Dinâmicas em Animações Faciais 3D com Processamento de Fala / Paula Salgado Lucena Rodrigues; orientador: Bruno Feijó; co-orientador: Luiz Carlos P.R. Velho. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2007.

v., 161 f: il. ; 29,7 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Animação Facial. 3. Expressões Faciais Dinâmicas. 4. Processamento de Fala. 5. Modelo de Emoção. 6. Hipercubo Emocional. 7. Personagens Virtuais. 8. Personagens Virtuais Falantes. 9. MPEG-4. I. Feijó, Bruno. II. Velho, Luiz. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

Aos nossos filhos Milton e Maria Clara que viveram intensamente esta Tese durante as suas gestações. Por eles, eu busquei garra e esperança de terminar este trabalho.

Agradecimentos

A Deus pela vida, saúde, sabedoria, esperança e perseverança.

À minha irmã Cynthia por ser meu exemplo e inspiração para não desistir dos sonhos. Agradeço pela vida, pelo amor, pela amizade, pela cumplicidade, pela doação e dedicação. Mesma na distância física, ela está sempre presente na minha vida.

Ao meu marido Rogério por ser amigo, companheiro, cúmplice, ... por me fazer feliz. Esta Tese é resultado de um sonho nosso, onde ele se sacrificou junto comigo, confiando e apostando nesta vitória.

À Asla Sá pela amizade verdadeira que construímos. Agradeço também por todas as nossas discussões matemáticas, pelas idéias e concretizações. Devo a ela muito do resultado deste trabalho!

Aos meus pais Paulo e Maria de Lourdes por tudo que tenho e que sou. Pelos ensinamentos como ser humano, por acreditarem nos meus sonhos, por serem meus melhores amigos e companheiros.

Ao professor Luiz Velho, exemplo de profissional e de ser humano. Agradeço a ele por acreditar na minha “animação”, desde o Mestrado, por sonhar junto comigo, por me fazer persistir neste sonho. Agradeço também por todo ensinamento, profissional e pessoal, sendo um exemplo a ser seguido.

Ao professor Bruno Feijó pela confiança e pelas oportunidades propiciadas.

À minha família e a família de Rogério por respeitarem o “momento Tese”, por toda torcida, paciência, orações e pensamento positivo.

As minhas “irmã cariocas” Christiane Caula, Liliane Thompson, Luciana Lima, Lucimar Martins, Maíra Greco, Taciana Melcop e Viviane Braconi pela amizade, companheirismo, torcida, cumplicidade... por serem minha família aqui no Rio.

Ao professor Luiz Fernando Gomes Soares pela oportunidade de trabalhar no TeleMídia e no projeto de TV Digital. Agradeço também aos amigos do Laboratório Telemídia, incluindo Lorenza Moreno e Suzana Maranhão.

Aos professores Antonio Furtado e Soraia Raup pelo trabalho no projeto Virtus.

Aos professores Marcelo Gattass e Waldemar Celes pela oportunidade de retornar ao TecGraf que foi de grande incentivo para a conclusão desta Tese.

Aos amigos do TecGraf, principalmente, ao Grupo de Visualização pela torcida, incentivo e inspirações diários.

Aos amigos de Recife, Sérgio Soares, Isabella Aragão, Maria Luisa Ribeiro, Conceição e Breno Sayão, Patricia Ramos, por serem minha fortaleza e me passarem a certeza que estamos sempre juntos, mesmo na distância física.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus professores do Departamento de Informática e do IMPA. Com o conhecimento dado por eles consegui crescer profissionalmente e realizar meus sonhos acadêmicos.

Ao pessoal do Departamento de Informática pela ajuda de todos os dias, em particular à Fatima, Emmanuelle e ao Alex.

Resumo

Rodrigues, Paula Salgado Lucena; Feijó, Bruno; Velho, Luiz. **Um Sistema de Geração de Expressões Faciais Dinâmicas em Animações Faciais 3D com Processamento de Fala.** Rio de Janeiro, 2007. 161p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese apresenta um sistema para geração de expressões faciais dinâmicas sincronizadas com a fala em uma face realista tridimensional. Entende-se por expressões faciais dinâmicas aquelas que variam ao longo do tempo e que semanticamente estão relacionadas às emoções, à fala e a fenômenos afetivos que podem modificar o comportamento de uma face em uma animação. A tese define um modelo de emoção para personagens virtuais falantes, denominado *VeeM* (*Virtual emotion-to-expression Model*), proposto a partir de uma releitura e uma reestruturação do modelo do círculo emocional de Plutchik. O *VeeM* introduz o conceito de um hipercubo emocional no espaço canônico do R^4 para combinar emoções básicas, dando origem a emoções derivadas. Para validação do *VeeM* é desenvolvida uma ferramenta de autoria e apresentação de animações faciais denominada *DynaFeX* (*Dynamic Facial eXpression*), onde um processamento de fala é realizado para permitir o sincronismo entre fonemas e visemas. A ferramenta permite a definição e o refinamento de emoções para cada quadro ou grupo de quadros de uma animação facial. O subsistema de autoria permite também, alternativamente, uma manipulação em alto-nível, através de *scripts* de animação. O subsistema de apresentação controla de modo sincronizado a fala da personagem e os aspectos emocionais editados. A *DynaFeX* faz uso de uma malha poligonal tridimensional baseada no padrão MPEG-4 de animação facial, favorecendo a interoperabilidade da ferramenta com outros sistemas de animação facial.

Palavras-chave

Animação Facial. Expressões Faciais Dinâmicas. Processamento de Fala. Modelo de Emoção. Hipercubo Emocional. Personagens Virtuais. Personagens Virtuais Falantes. MPEG-4.

Abstract

Rodrigues, Paula Salgado Lucena; Feijó, Bruno; Velho, Luiz. **A System for Generating Dynamic Facial Expressions in 3D Facial Animation with Speech Processing**. Rio de Janeiro, 2007. 161p. PhD Thesis — Department of Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis presents a system for generating dynamic facial expressions synchronized with speech, rendered using a tridimensional realistic face. Dynamic facial expressions are those temporal-based facial expressions semantically related with emotions, speech and affective inputs that can modify a facial animation behavior. The thesis defines an emotion model for speech virtual actors, named *VeeM* (*Virtual emotion-to-expression Model*), which is based on a revision of the emotional wheel of Plutchik model. The *VeeM* introduces the emotional hypercube concept in the R^4 canonical space to combine pure emotions and create new derived emotions. In order to validate *VeeM*, it has been developed an authoring and player facial animation tool, named *DynaFeX* (*Dynamic Facial eXpression*), where a speech processing is realized to allow the phoneme and viseme synchronization. The tool allows either the definition and refinement of emotions for each frame, or group of frames, as the facial animation edition using a high-level approach based on animation *scripts*. The tool player controls the animation presentation synchronizing the speech and emotional features with the virtual character performance. *DynaFeX* is built over a tridimensional polygonal mesh, compliant with MPEG-4 facial animation standard, what favors tool interoperability with other facial animation systems.

Keywords

Facial Animation. Dynamic Facial Expressions. Speech Processing. Emotion Model. Emotional Hypercube. Virtual Characters. Virtual Talking Characters. MPEG-4.

Sumário

1	Introdução	15
1.1	Estado da Arte em Animação Facial	15
1.2	Desafios da Área	18
1.3	Parametrização Ideal para Sistemas de Animação Facial	19
1.4	Contribuições Esperadas	21
1.5	Estrutura do Documento	22
2	Espaços de Emoção	24
2.1	Estado da Arte para os Modelos de Emoção	24
2.2	Emoções Básicas	26
2.3	Emoções na Teoria da Avaliação	28
2.4	Modelo de Humor	30
2.5	Modelos de Emoção em Sistemas Computacionais	32
3	VeeM: Um Modelo de Emoção para Personagens Virtuais	35
3.1	Esquema de Classificação de Emoção em Eixos Bipolares	36
3.2	Combinação de Emoções Puras para Geração de Emoções Derivadas	37
3.3	O Hipercubo Emocional para o <i>VeeM</i>	39
3.4	Taxonomia das Emoções Derivadas a partir da Combinação Binária de Emoções Puras	41
3.5	Incorporação de Fenômenos Afetivos e Expressões Não-Verbais	50
3.6	Visualização da Emoção: Expressão Facial	52
3.7	Análise do <i>VeeM</i> e seu Hipercubo Emocional segundo a Parametrização Ideal para Animação Facial	53
4	Espaços de Expressões Faciais no Padrão MPEG-4	55
4.1	Espaço de Expressões Faciais	56
4.2	Apresentação do Padrão MPEG-4	57
4.3	Especificação da Face no Padrão MPEG-4	58
4.4	Expressões Faciais no Padrão MPEG-4	62
4.5	Visema e <i>Lip-Sync</i> no MPEG-4	64
4.6	Especificação de um Modelo de Face no Padrão MPEG-4	66
4.7	Análise do Padrão MPEG-4 segundo a Parametrização Ideal para Animação Facial	67
5	DynaFeX: Uma Ferramenta de Edição e Exibição de Expressões Faciais Dinâmicas	70
5.1	Visão Geral da <i>DynaFeX</i>	71
5.2	Fornecimento de Arquivos de Entrada	74
5.3	Geração da Estrutura Fonética da Fala	77

5.4	Definição e Refinamento de Emoções para Geração de Expressões Faciais	82
5.5	Combinação dos FAPs	85
5.6	Execução da Animação Facial com Sincronismo da Fala e dos Aspectos Emocionais	87
5.7	Implementação dos Fenômenos Afetivos e das Expressões Não-Verbais	89
5.8	Estudos de Casos	92
6	Trabalhos Relacionados	97
6.1	Especificação de Animação Facial através de um Sistema Multicamadas	97
6.2	Síntese e Edição de Animação Facial Expressiva com Controle através de Isomapas de Fonemas	99
6.3	Modificadores Paramétricos Qualitativos em um Sistema de Animação Facial	101
6.4	Um Modelo de Representação de Emoções Dinâmicas usando a Arquitetura CogAff	103
6.5	Síntese de Outros Trabalhos Relevantes	105
6.6	Análise Comparativa	108
7	Conclusão	110
7.1	Contribuições da Tese	111
7.2	Trabalhos Futuros	113
	Referências Bibliográficas	115
	Sumário das notações	126
A	Especificação dos Pontos Característicos do Modelo Facial no Padrão MPEG-4	127
B	Especificação dos Parâmetros da Animação Facial no Padrão MPEG-4	138
C	Especificação dos FAPs na DynaFeX para as Expressões Faciais dos Visemas e das Emoções Puras a partir do Hipercubo Emocional	151

Lista de figuras

1.1	Taxonomia da Animação Facial.	16
2.1	Emoções universais definidas por Ekman.	27
2.2	Emoções básicas definidas por Plutchik.	27
2.3	Caricaturas para emoções puras definidas por Plutchik.	28
2.4	As três camadas da arquitetura CogAff.	29
2.5	Interface gráfica do sistema <i>EmoSpeak</i> .	33
3.1	Visualização completa do círculo de Plutchik.	37
3.2	Configuração para emoções opostas.	38
3.3	Combinações de emoções.	38
3.4	Emoções básicas nos eixos do <i>VeeM</i> .	39
3.5	Emoções do plano xy.	43
3.6	Emoções do plano xz.	44
3.7	Emoções do plano xw.	45
3.8	Emoções do plano yz.	46
3.9	Emoções do plano yw.	48
3.10	Emoções do plano zw.	49
3.11	Esquematização do <i>VeeM</i> .	50
4.1	Músculos das expressões faciais.	57
4.2	A face neutra e definição das FAPUs.	59
4.3	Pontos característicos do padrão MPEG-4.	60
4.4	Exemplo da animação de um FAP.	62
4.5	Expressões faciais no MPEG-4.	63
4.6	Visemas para expressão facial feliz.	66
5.1	Visão geral da ferramenta <i>DynaFeX</i> .	71
5.2	Padrão de projeto aplicado na <i>DynaFeX</i> .	73
5.3	Exemplo de descrição textual e de emoção.	76
5.4	Exemplo de arquivo <i>fap</i> .	76
5.5	Módulo de estrutura fonética da fala.	77
5.6	Arquivos de saída do reconhecedor de fala.	78
5.7	Expressão facial dos 15 visemas da <i>DynaFeX</i> .	81
5.8	Expressões puras do <i>VeeM</i> na <i>DynaFex</i> .	82
5.9	Edição alto-nível.	83
5.10	Edição baixo-nível.	85
5.11	Módulo de combinação de FAPs.	86
5.12	Interface gráfica para movimento dos olhos.	90
5.13	Configuração dos olhos para algumas emoções.	91
5.14	Análise comparativa do movimento dos olhos.	92
5.15	Integração das ferramentas ao formatador <i>HyperProp</i> .	94
5.16	Exemplo de execução.	94
5.17	Arquitetura da junção do <i>ETHs</i> com o <i>LOGTELL</i> .	95
5.18	Cena renderizada no ambiente de narração.	96

6.1	As cinco camadas do sistema SMILE.	98
6.2	Visão geral do sistema <i>eFASE</i> .	99
6.3	Arquitetura do sistema EE-FAS.	104
6.4	Definição das categorias de emoções.	106
6.5	Disco emocional para combinação de emoções.	107
A.1	Pontos característicos no padrão MPEG-4.	127
C.1	Expressão facial dos 15 visemas na <i>DynaFeX</i> .	151
C.2	Emoções puras na <i>DynaFeX</i> .	157

Lista de tabelas

2.1	Emoções Básicas	34
3.1	Semântica das emoções puras.	40
4.1	Grupo de FAPs	61
4.2	Expressões faciais primárias	64
4.3	Visemas e os fonemas relacionados.	65
5.1	Mapeamento dos fonemas para visemas.	80
6.1	Parâmetros e elementos do <i>Effort</i> .	102
C.1	Intervalo sugerido dos FAPs nos eixos x e y .	160
C.2	Intervalo sugerido dos FAPs nos eixos z e w .	161

“Animation is not the art of drawings that move, but rather the art of movements that are drawn. What happens between each frame is more important than what happens in each frame. Animation is the art of manipulating the invisible interstices that lie between frames.”

Norman McLaren, *Chuck Jones: Life of Animated Cartoonist*.