6
Trabalhos Relacionados

Este capítulo apresenta alguns sistemas para geração de expressões faciais. As quatro primeiras seções descrevem em maiores detalhes trabalhos relacionados diretamente ao modelo *VeeM* (Capítulo 3) e à ferramenta *DynaFeX* (Capítulo 5): a Seção 6.1 aborda um trabalho na linha de linguagens de marcação para especificação das animações; a Seção 6.2 descreve um trabalho em que a especificação das animações é direcionada pelos fonemas que compõem a fala da personagem; a Seção 6.3 descreve um modelo de emoção para animações faciais no padrão MPEG-4 encapsulando os parâmetros de baixo-nível do padrão (FAPs) em parâmetros de alto-nível; e a Seção 6.4 aborda um modelo de emoção para geração de expressões dinâmicas segundo a arquitetura DER (Seção 2.5).

Outros trabalhos relacionados são apresentados em menores detalhes na Seção 6.5. Trabalhos que abordam aspectos como linguagens de marcação para emoções, geração automática de emoções e expressões faciais, movimentos de componentes faciais não-verbais e integração de personagens virtuais a outros sistemas são comentados nessa seção.

As principais vantagens e desvantagens de cada trabalho relacionado são comentadas na própria descrição dos trabalhos. Ao final, a Seção 6.6 apresenta uma breve análise comparativa entre a proposta desta tese e as demais contribuições já realizadas para a área de animação facial.

6.1
Especificação de Animação Facial através de um Sistema Multicamadas

A referência (Karl91) descreve um sistema, chamado SMILE, para especificação de animação facial baseada em uma abordagem multicamada. O trabalho especifica camadas com diferentes níveis de complexidade e propõe uma linguagem para sincronização da fala, das emoções e do movimento dos olhos para fornecer um modo natural de especificar sequências de animações. A Figura 6.1 ilustra as camadas, do nível mais baixo de abstração (camada 0) para o nível mais alto (camada 4). Cada camada é independente, possuindo suas próprias entradas e saídas. O objetivo é tornar o sistema flexível e permitir

PUC-Rio - Certificação Digital Nº 0321125/CA

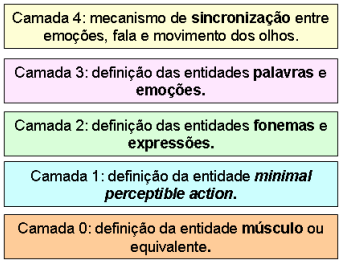


Figura 6.1: As cinco camadas do sistema SMILE.

que o comportamento de um elemento do sistema possa ser modificado sem impactar os demais.

A camada 0 (músculos abstratos) corresponde a um sistema de animação básico. No SMILE (Karl91), a implementação é baseada nos procedimentos AMA (*Abstract Muscle Action*), definidos por outro trabalho do grupo (ThPT88). As ações são muito específicas para os vários músculos e dão a ilusão da presença de uma estrutura óssea. Um problema com essa abordagem é que as deformações são baseadas em modelos empíricos e não em leis físicas.

A camada 1 (*minimal perceptible actions*) corresponde a parâmetros básicos de movimento facial; ações da face que são minimamente percebidas. O intervalo desse movimento é normalizado entre 0% e 100% (por exemplo, elevar a sobrancelha direita de 70%).

A camada 2 (*facial snapshot*) é obtida pela especificação do valor de cada ação mínima. Uma vez definido, um *snapshot* tem a forma $< framenumber >< snapshot >< intensity >$. Vários *snapshots* podem ser ativados ao mesmo tempo, permitindo, por exemplo, especificar um fonema e uma expressão facial simultaneamente. Baseado no trabalho das expressões faciais de Ekman (Ekma71), as seis expressões universais podem ser utilizadas.

A camada 3 (seqüência de *snapshots*) define as entidades "palavras" e "emoções". Uma palavra é especificada como uma seqüência de fonemas. Já uma emoção é definida como uma evolução da face humana ao longo do tempo.

A camada 4 (mecanismo de sincronização) é responsável por introduzir mecanismos para especificação do tempo inicial e da duração de cada ação. Isso implica que cada ação pode ser executada independentemente do estado atual do ambiente porque a sincronização é apenas dependente do tempo. Para especificação do sincronismo é definida uma linguagem de manipulação chamada HLSS (*High Level Script Scheduler*).

O SMILE é um sistema que oferece uma complexidade relativamente baixa para o animador. As entidades definidas correspondem a conceitos intuitivos tais como fonemas, expressões, palavras, emoções, sentenças e movimento

PUC-Rio - Certificação Digital Nº 0321125/CA

dos olhos, o que torna o sistema natural de manipular. Um aspecto negativo é o modelo de camadas não ter sido incorporado em nenhum sistema de animação facial, apesar dos autores afirmarem que o sistema em camadas é independente do sistema de animação e, conseqüentemente, de simples incorporação.

6.2 Síntese e Edição de Animação Facial Expressiva com Controle através de Isomapas de Fonemas

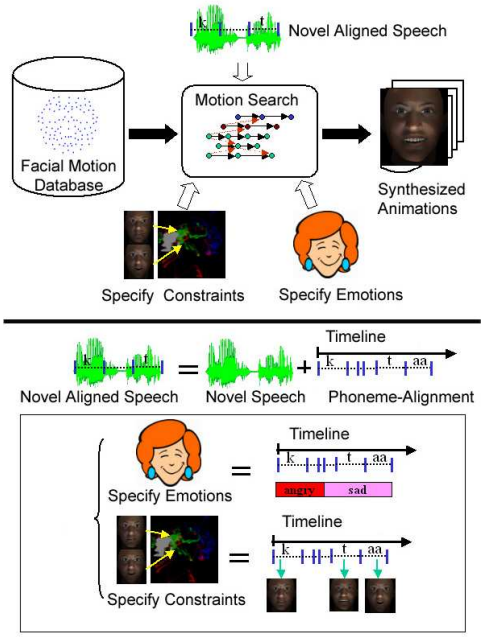


Figura 6.2: Visão geral do sistema *eFASE*.

O sistema *eFASE* (*expressive Facial Animation Synthesis and Editing system*) (DeNe06) é um sistema para síntese e edição de animações faciais dirigidas por fala. A Figura 6.2 apresenta uma visão do sistema e de seus componentes principais. Basicamente, o *eFASE* gera animações faciais expressivas através da concatenação de movimentos oriundos de dados capturados enquanto que animadores estabelecem restrições e objetivos, interativamente.

Para geração da animação facial, o *eFASE* passa por uma fase de pré-processamento formada pela captura dos dados (formando uma base de movimento facial, na Figura 6.2 referenciada como *Facial Motion Database*) e pelo alinhamento dos fonemas da fala a ser utilizada na animação facial.

Para construção da base de dados de movimentos faciais, uma atriz usando 102 marcadores em sua face foi orientada para falar quatro vezes um conjunto de sentenças com cerca de 225 fonemas balanceados. Cada repetição foi falada com uma expressão facial diferente (natural, feliz, com raiva e triste). O total de dados inclui mais de 105.000 quadros de movimento capturados. Das

PUC-Rio - Certificação Digital Nº 0321125/CA

102 marcas capturadas apenas 90 foram realmente utilizadas pelo trabalho. Cada quadro de movimento foi rotulado com o nome da expressão facial, sendo essa a única marcação necessária para o algoritmo de busca que é usado.

Após a captura de dados, a base de movimento facial foi normalizada. Com a base de dados normalizada, o trabalho optou por remover o movimento de cabeça da base. Os movimentos dos 90 marcadores para cada quadro de movimento foram empacotados em um vetor de movimento com dimensão igual a 270. Uma redução da dimensionalidade desse vetor foi feita aplicando análise de componentes principais (PCA - *Principal Components Analysis*) (Shle05), conseguindo reduzí-lo a uma dimensionalidade de 25, cobrindo cerca de 98.53% da variação de movimento. No *eFase*, o termo *Motion Frames* faz referência aos vetores de coeficientes PCA ou, correspondentemente, às configurações de marcações faciais.

Para o reconhecimento e o alinhamento dos fonemas da fala da personagem contida no áudio capturado, o *eFASE* faz uso da ferramenta *Festival Speech System* (BlTa97).

Após a fase de pré-processamento, o usuário pode navegar e selecionar restrições de expressões para os fonemas utilizando uma ferramenta de edição e visualização (DeNe06). O usuário pode também, opcionalmente, especificar modificadores de emoção sobre intervalos arbitrários de tempo. Essas interações do usuário são alinhadas por fonemas com o intuito de fornecer um controle relativo à fala. As restrições e os modificadores são aplicados sobre as emoções capturadas para formação da base de dados.

Tendo como base os tempos limitantes de cada fonema (provenientes da fase de alinhamento dos fonemas), a seqüência de movimento capturado é dividida em pequenas subseqüências onde cada uma corresponde à duração de um fonema específico. Todos os nós de movimento de cada fonema representam configurações faciais para o fonema específico e esse conjunto é referenciado como "*cluster* do fonema". Cada quadro de movimento no *cluster* do fonema possui um rótulo de emoção e uma propriedade relativa ao tempo (na verdade, referente à duração do nó de movimento que o fonema pertence). Além do *cluster* de fonema, também é construída uma base de dados para movimentos faciais relacionados. Essa base forma conceitualmente um espaço 3D representado pelos "eixos" sentença, emoção e ordem do nó de movimento (DeNe06).

Todos esses passos são efetuados sobre os dados da base de movimento facial para que se tenha uma maior eficiência no algoritmo de busca ("*Motion Search*" na Figura 6.2). O algoritmo sintetiza uma seqüência de movimento facial buscando os melhores casamentos dos quadros de movimento capturado, tendo como base a nova seqüência de fonemas da fala, as restrições especificadas

PUC-Rio - Certificação Digital Nº 0321125/CA

pelo usuário das expressões dos fonemas e os modificadores de emoção. A saída gerada automaticamente pelo *eFASE* consiste de uma animação facial expressiva sintética obtida através da concatenação dos dados de movimento capturado ("*Synthesized Animations*" na Figura 6.2).

Segundo os autores, o *eFASE* possui duas contribuições principais. A primeira consiste em fornecer um controle intuitivo para síntese e edição de animações faciais dirigidas por dados. A segunda contribuição consiste do novo controle baseado em fonemas (*phoneme-Isomap* em (DeNe06)) fornecendo ferramenta para edição e gerenciamento de uma grande base de dados de movimento. No entanto, o sistema requer um processo custoso para formação da base de dados de treinamento, além de operar apenas com um subconjunto das emoções universais.

6.3 Modificadores Paramétricos Qualitativos em um Sistema de Animação Facial

O trabalho descrito em (ByBa02) propõe um método para síntese de animação facial utilizando um conjunto de elementos paramétricos de alto-nível que encapsulam os FAPs de baixo-nível definidos no padrão MPEG-4. O objetivo do trabalho é simplificar o processo de síntese de animação facial permitindo que parâmetros de alto-nível controlem a face. O desafio desse trabalho está em identificar um conjunto razoável de tais parâmetros de alto-nível (PaWa96).

O sistema criado é chamado de FacEmote, sendo uma adaptação para face do sistema EMOTE (*Expressive MOTion Engine*) (Chi00), originalmente desenvolvido para gestos e posições dos braços e do tronco. Ambos os sistemas são uma interpretação da "análise do movimento de Laban" (LMA - *Laban Movement Analysis*[1]).

O LMA é um método para descrição (linguagem textual e simbólica) de movimentos, sendo composto por cinco componentes principais: *Body* (corpo), *Space* (espaço), *Shape* (forma), *Effort* (esforço) e *Relationship* (relacionamento). O EMOTE (Chi00) utiliza apenas as componentes *Shape* e *Effort* para descrever seus movimentos, enquanto que o FacEMOTE faz uso somente da componente *Effort* para modelar os movimentos faciais.

A componente *Effort* descreve como o corpo concentra seu esforço quando está executando movimentos. O *Effort* é geralmente comparado a termos

[1]LMA é promovido pelo *the Laban/Bartenieff Institute of Movement Studies (LIMS) - URL: http://www.limsonline.org*(acesso em 24/Out/2007).

PUC-Rio - Certificação Digital Nº 0321125/CA

musicais dinâmicos como forte, doce, contínuo etc. O *Effort* é composto por quatro parâmetros, como descreve a Tabela 6.1.

Tabela 6.1: Parâmetros e elementos do *Effort*.

Parâmetros	Elementos do Parâmetro
espaço (*space*)	indireto (*indirect*) e direto (*direto*)
peso (*weight*)	leve (*light*) e pesado (*strong*)
tempo (*time*)	ininterrupto (*sustained*) e rápido (*quick*)
fluxo, fluência (*fluxo*)	livre (*free*) e limitado(*bound*)

Com os quatro parâmetros que compõem o *Effort*, o FacEmote descreve a expressividade dos movimentos faciais através de um mapeamento para os parâmetros de baixo-nível do padrão MPEG-4 (ver Capítulo 4 e Apêndice B deste documento para descrição completa dos FAPs). O objetivo é tornar o trabalho de síntese de animação facial mais simples através de dois fatores principais: a redução do número de parâmetros a um número pequeno de parâmetros intuitivos e a aplicação desses parâmetros sobre uma seqüência de quadros ao invés de um único quadro, produzindo assim uma seqüência de animação para apresentação em tempo real.

Dos 66 FAPs de baixo-nível do padrão MPEG-4 (excluindo visemas e expressões que são os dois FAPs de alto-nível do padrão), o FacEMOTE utiliza apenas 47 para serem afetados pelos quatro parâmetros do *Effort*, não levando em consideração os FAPs para as regiões da língua, nariz, orelhas e pupilas. Esses 47 parâmetros são reunidos em dois grupos: parâmetros primários, composto por 8 FAPs independentes entre si; e parâmetros secundários, composto por 39 FAPs cujos valores são obtidos como interpolações dos parâmetros primários.

O FacEMOTE gera novos dados de animação usando uma etapa preparatória *offline* que modifica o fluxo de FAPs de entrada em três passos. A etapa preparatória é responsável por gerar uma tabela de intervalos de movimentos dos FAPs para cada modelo facial, para evitar extrapolações nos valores de cada FAP.

No primeiro passo, a trajetória de movimento de cada parâmetro é analisada como preparação para os dois próximos passos. Um mínimo local e um máximo local são encontrados definindo, assim, o perfil.

O segundo passo modula os FAPs primários. O FacEMOTE trabalha como um filtro, modificando o valor de cada FAP para cada quadro da animação. Um conjunto de funções $F_1, F_2, ..., F_8$ é computado para os valores

PUC-Rio - Certificação Digital Nº 0321125/CA

dos parâmetros do EMOTE atuando como filtros que ajustam os oito FAPs primários.

Por fim, o terceiro passo modula o restante dos FAPs (parâmetros secundários) com as funções obtidas pela combinação das funções dos parâmetros primários. O fluxo de FAPs gerado após essas três etapas pode ser aplicado em um modelo facial para renderização.

Os autores do FacEMOTE (ByBa02) afirmam que o FacEMOTE pode ser incorporado em qualquer sistema de animação facial. Um ponto a ser destacado desse trabalho é o uso da teoria de Laban (LMA) para gerar animações faciais, sendo um trabalho inovador em termos do uso de teorias clássicas para definição das expressões faciais. A principal desvantagem está no descarte de alguns FAPs, restringindo um pouco as expressões faciais obtidas.

6.4 Um Modelo de Representação de Emoções Dinâmicas usando a Arquitetura CogAff

EE-FAS - Emotionally Expressive Facial Animation System (Tang06) (TaWB05) é um sistema de animação facial cujo principal objetivo é permitir que animadores não profissionais, ou outros programas, criem animações faciais usando expressões faciais significativas. O EE-FAS utiliza a arquitetura DER (Seção 2.3 e Seção 2.5) para ampliar o número de expressões faciais visualizadas.

O EE-FAS distingue e produz dois tipos de expressões faciais: expressões emocionais e ações de funções comunicativas. As expressões emocionais são representadas através das seis expressões universais definidas por Ekman (Seção 2.2). No EE-FAS, expressões emocionais são mostradas como reações a impulsos emocionais, não estando relacionadas com a fala. Na verdade, elas podem interromper uma fala do ator virtual.

As expressões emocionais são usadas para comunicar uma mensagem emocional única e clara, com uma rápida reação para uma mudança no ambiente. Por exemplo, o ator virtual pode ser interrompido durante a fala e mostrar uma expressão facial de surpresa.

Já as ações de funções comunicativas são expressões faciais relacionadas e sincronizadas com a fala. No EE-FAS, a descrição das ações comunicativas foram baseadas no trabalho (BaCh00). Basicamente, ações comunicativas têm duas componentes: uma função comunicativa, como "enfatizar palavras", e um sinal, que é um movimento visível da face, como levantar as sobrancelhas.

A Figura 6.3 ilustra a arquitetura do sistema EE-FAS. Os módulos do sistema se comunicam por troca de mensagens, através de um *middleware*

PUC-Rio - Certificação Digital Nº 0321125/CA

chamado *Psyclone* (Tang06).

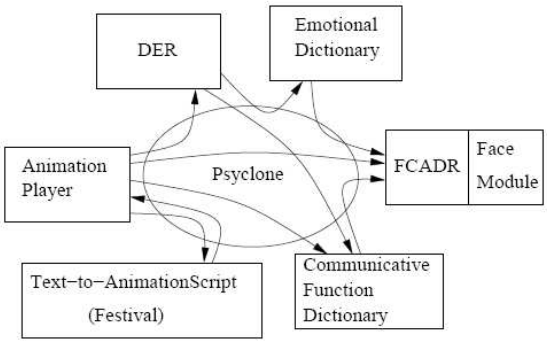


Figura 6.3: Arquitetura do sistema EE-FAS.

O módulo de entrada, chamado *Animation Player*, recebe um texto com marcações descritas em XML que descrevem funções comunicativas e impulsos emocionais. Esse texto com marcações é enviado para o módulo *Text-to-AnimationScript* que o transforma em um *script* da animação. O *script* é composto por uma lista de marcações, incluindo marcações de visemas, marcações das funções comunicativas e dos impulsos emocionais. Todas essas marcações possuem a informação de tempo provenientes da ferramenta *Festival Speech Synthesis* (BlTa97).

O *script* de animação é enviado para o módulo *Animation Player*, que é responsável por produzir a animação. Cada tipo de marcação é enviado para um módulo diferente: as marcações de visemas são enviadas para o módulo *Facial Component Action Dynamic Representation (FCADR)*, as marcações das funções comunicativas e impulsos emocionais são enviados, respectivamente, para os módulos *DER* e *Communicative Function Dictionary*.

As expressões emocionais oriundas dos impulsos emocionais não são sincronizadas com a fala no EE-FAS. No momento da apresentação de uma expressão emocional, a sincronização labial pára, a expressão emocional é exibida no seu tempo estabelecido, e depois a animação facial com fala prossegue segundo especificado pela entrada.

Por sua vez, as funções comunicativas possuem relação com a fala e, por essa razão, elas devem ser sincronizadas. Na verdade o que ocorre no EE-FAS não é exatamente a sincronização das funções comunicativas; a partir das funções comunicativas são extraídos os sinais faciais que são sincronizados em paralelo com os visemas. Um sinal facial é definido como uma lista de ações dos componentes faciais, excluindo a boca. No EE-FAS há um mapeamento de um-para-um entre fonemas e visemas. No instante da apresentação do áudio, são aplicados na face os sinais faciais e o visema correspondentes àquele instante de tempo, havendo assim um casamento com a fala do personagem.

PUC-Rio - Certificação Digital Nº 0321125/CA

Para animar uma malha facial 3D, é utilizado um modelo de músculos abtratos (PaWa96) (TaWB03) (BuHN04). O sistema foi desenvolvido para controlar uma face sintética independente da técnica utilizada para exibir a face na tela (Tang06). Para obter tal independência, o sistema manipula ações de componentes faciais, que estão relacionadas com as unidades de ações (AUs - *Action Units*) definidas pelo FACS (*Facial Action Coding System*) (EkFr75).

O EE-FAS propõe uma interessante dinâmica nas expressões faciais geradas a partir dos estímulos emocionais, respeitando as três camadas de diferentes níveis de emoção da arquitetura DER (Seção 2.5). O aspecto negativo está na distinção entre expressões emocionais que não estão sincronizadas com a fala e os sinais faciais sincronizados com a fala. Não há um casamento entre visemas, movimentos faciais e expressões emocionais, sendo cada um deles tratados e exibidos de forma individual.

6.5 Síntese de Outros Trabalhos Relevantes

Assim como a proposta do sistema SMILE (Karl91), o trabalho *Avatar Markup Language - AML* (KiGK02) define uma linguagem de alto-nível baseada em XML. A linguagem AML encapsula elementos para síntese de texto (*Text-To-Speech*), animação da face e animação do corpo. Os elementos para animação facial e corporal definem duas linguagens derivadas de AML que especificam marcações usando como base os parâmetros baixo-nível (FAPs) do padrão MPEG-4. Essas linguagens receberam os nomes *Avatar Face Markup Language* (AFML) e *Avatar Body Markup Language* (ABML), respectivamente.

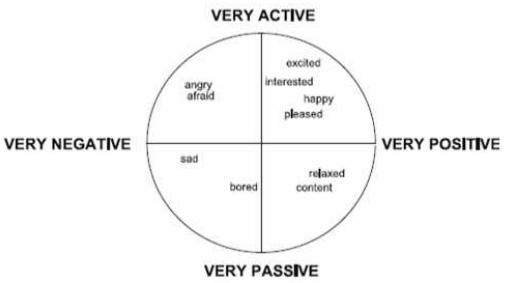
A AML possui a vantagem de definir marcações para animação segundo os parâmetros de baixo-nível do padrão MPEG-4. A desvantagem existente é a complexidade em especificar os arquivos de animação segundo as linguagens AFML e ABML. O ideal seria ter uma abstração em mais alto-nível para os FAPs, sendo possível definir suas propriedades e valores dentro do próprio arquivo AML, não precisando de "linguagens derivadas".

Seguindo essa mesma linha de linguagens de marcação para definição das animações faciais, o grupo de trabalho de emoções (*Emotion Incubator Group*) ligado ao consórcio W3C produziu um relatório que descreve os principais requisitos para a definição de uma linguagem de marcação para emoções. No entanto, a definição dessa linguagem padronizada ainda é um tópico de pesquisa em aberto (W3C07).

No contexto de geração de emoções e expressões faciais, o trabalho "*Mixed Feelings*" (Albr05) apresenta um algoritmo para geração de expressões faciais

PUC-Rio - Certificação Digital Nº 0321125/CA

para um contínuo de emoções puras e misturadas de intensidade variada. Para alcançar esse objetivo, o algoritmo proposto por Tsapatsoulis et al. (Tsap02) foi adaptado para ser aplicado em um sistema de animação baseado em músculos e em um único modelo de emoção integrado. O sistema de animação baseado em músculo foi combinado com um sistema de síntese *text-to-speech* baseado no mesmo modelo de emoção para gerar um *talking head* capaz de expressar emoções não-básicas com intensidades variadas.



Category	Activation	Evaluation	Power
Joy	17.3	42.2	12.5
Distress	−17.2	−40.1	−52.4
Happy-for	17.3	42.2	12.5
Gloating	40.0	30.0	30.0
Resentment	0.0	−40.0	−20.0
Sorry-for	−17.2	−40.1	−52.4
Hope	20.0	20.0	−10.0
Fear	14.8	−44.4	−79.4
Satisfaction	−14.9	33.1	12.2
Relief	3.0	33.0	−3.0
Fears-confirmed	−30.0	−50.0	−70.0
Disappointment	2.4	−24.9	−37.2
Pride	30.0	40.0	30.0
Admiration	27.0	53.0	17.0
Shame	4.6	−26.3	−62.3
Reproach	−3.0	−30.0	43.0
Liking	−14.9	33.1	12.2
Disliking	15.0	−35.0	−10.0
Gratitude	20.0	40.0	−30.0
Anger	34.0	−35.6	20.0
Gratification	−14.9	33.1	12.2
Remorse	4.6	−26.3	−62.3
Love	1.2	33.3	14.9
Hate	60.0	−60.0	30.0

Figura 6.4: Definição das categorias de emoções.

O sistema *talking head* usado no artigo "*mixed feelings*" (Albr05) é formado por dois componentes principais: um sistema TTS (*text-to-speech*) e um sistema de animação facial. O TTS usado nesse artigo é o "MARY" (ScTr03) (Seção 2.5). No instante em que o texto é fornecido como entrada para o módulo TTS, ele é transformado em um esqueleto XML básico que agrupa o texto com marcações relacionadas à emoção. A estrutura final do XML não é usada apenas pelo módulo TTS, mas também pelo sistema de animação facial. As expressões faciais são definidas através de rótulos (categorias de emoção). Cada rótulo é deduzido a partir da combinação de dimensões de emoção, como ilustrado pela Figura 6.4.

Ainda relacionado à geração de novas emoções, Ruttkay et al. (RuNT03) organizaram as seis emoções básicas de forma eqüidistante sobre a borda de um círculo de acordo com a similaridade. Para cada ponto do disco uma expressão facial é associada sendo computada através da interpolação linear entre as emoções básicas mais próximas, como ilustrado pela Figura 6.5. A distância do centro à extremidade do círculo descreve a intensidade.

Uma outra área interessante de trabalhos relacionados diz respeito aos movimentos dos componentes não-verbais, como apresentado na Seção 5.7. Em (PoPe02) é proposta a abordagem *meaning-to-face*, que permite a simulação de geração automática das expressões faciais sobre o dado semântico.

PUC-Rio - Certificação Digital Nº 0321125/CA

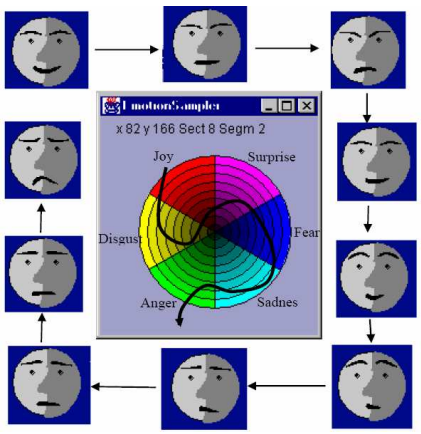


Figura 6.5: Disco emocional para combinação de emoções.

Ainda nessa área, Garau et al. (Gara01) investigam um comportamento direcional em avatares representando pessoas engajadas em uma conversação.

Diferente dos trabalhos citados, o principal objetivo do estudo de movimentos não-verbais feito nesta tese (Rodr07) foi avaliar o impacto visual do movimento sacádico dos olhos quando eles estão conectados com animações emocionais e com o papel que a personagem assume durante a conversação. O trabalho (Rodr07) foi inspirado no projeto *Eyes Alive* (LeBB02) que apresenta um modelo estatístico de movimento dos olhos. Esse modelo é baseado tanto em um estudo empírico como em dados adquiridos de movimento dos olhos.

No que diz respeito à integração de personagens virtuais a outras aplicações, poucos ainda são os trabalhos que exploram os sistemas de animação facial. Em (Theu03) é descrito um sistema de *storytelling* com agentes semi-autônomos onde um narrador sintético virtual recebe um texto de entrada enriquecido com marcações de controle para contar a história. Diferente da integração feita nesta tese (Rodr06), o narrador desenvolvido em (Theu03) não possui emoção, comprometendo o resultado visual.

Uma outra linha de aplicação explorada nesta tese é a de apresentadores para programas de TV Digital e aplicações *web* (Rod04a) (Rod04b). Neste campo, a referência (GaMa01) descreve a implementação de uma ferramenta de suporte a aplicações *talking head* na *web*, utilizando o MPEG-4 como base para a animação das faces. Apesar de oferecer o suporte, o artigo apenas comenta algumas das possíveis aplicações *web* que poderiam fazer uso das animações faciais, deixando as suas implementações como trabalho futuro.

Por fim, uma pesquisa interessante na área de animação facial que pode inspirar trabalhos futuros a partir desta tese de doutorado é o trabalho conhecido como *reanimating face* (Blan03) (Blan04). Nesses trabalhos, é proposta a construção de uma base de formatos (*shapes*) e texturas de faces

PUC-Rio - Certificação Digital Nº 0321125/CA

a partir de um conjunto de 35 expressões faciais e expressões neutras de 200 pessoas, todas digitalizadas através de um *scanner* 3D. A partir desses dados, é gerado um espaço vetorial de expressões que possibilita a transferência de expressões entre os indivíduos. De posse desse espaço de faces é possível estimar um formato de face 3D a partir de uma imagem bidimensional, o que permite a criação de novas expressões para a imagem 2D. O método é ilustrado alterando as expressões faciais de quadros famosos como "Monalisa".

O processo também pode ser aplicado para expressões dinâmicas. Uma limitação do trabalho é o fato da base de dados concentrar-se em variações dos movimentos da boca, que no caso de vídeo produz resultados pouco realistas. Um ponto crítico do algoritmo é a geração das correspondências de pontos das faces digitalizadas para a geração do espaço de faces. A malha usada para representar as faces não é estruturada o que impossibilita a animação especificada em mais alto-nível.

6.6 Análise Comparativa

A principal diferença do trabalho proposto nesta tese é integrar em um mesmo modelo e implementação aspectos de animação facial que na maioria das vezes são tratados separadamente por cada sistema.

O modelo de emoção elaborado é flexível para a definição das emoções básicas e genérico para a construção das emoções derivadas, enquanto que a maioria dos sistemas limita-se às emoções de Ekman (Karl91) ou mesmo um subconjunto delas (DeNe06). Poucos são os que dão suporte a emoções derivadas (Albr05) (RuNT03).

A segunda vantagem é a utilização de um modelo de face parametrizado e padronizado. A descrição das emoções, visemas e outras expressões da face são realizadas com base em vetores desses parâmetros, tornando a descrição da animação independente da implementação da face. Como pode ser observado na literatura, ainda são poucos os trabalhos que convergem para um padrão de especificação parametrizada (ByBa02) (GaMa01). Existem também algumas tentativas de parametrização (Tang06) (TaWB05), porém não padronizadas.

Uma terceira diferença está na preocupação de oferecer uma interface de manipulação da animação em mais de um nível de abstração. O animador pode fazer uso direto dos parâmetros de mais baixo-nível, mas tem o recurso de abstrair os detalhes e descrever a animação através de construções mais próximas do seu entendimento (emoções básicas, palavras, humor, personalidade etc.), podendo inclusive fazer isso através de *scripts* de animação, como ocorre em (Karl91).

PUC-Rio - Certificação Digital Nº 0321125/CA

A definição de um modelo que integra "fenômenos afetivos" e incorpora mecanismos automatizados para geração dos movimentos de cabeça e dos olhos também é uma vantagem no modelo e sistema desenvolvidos nesta tese. De um modo geral, essas questões são tratadas separadamente na literatura. A combinação de visemas e emoções também é um ponto desafiador que, se ainda não resolve, esta tese ao menos oferece uma base de implementação que permite que parâmetros sejam facilmente ajustados para que análises qualitativas sejam feitas (Seção 5.5).

Por fim, um sistema de animação facial passa a ter maior apelo no momento em que é integrado a outras aplicações. A preocupação de desenvolver uma implementação reutilizável em outros sistemas acompanha esta pesquisa desde o seu início, como pode ser observado pelos trabalhos de integração realizados junto a sistemas hipermídia e de *storytelling*. Infelizmente, essa facilidade de integração ainda não é facilmente encontrada em outros sistemas de animação facial.

PUC-Rio - Certificação Digital Nº 0321125/CA