

4

Espaços de Expressões Faciais no Padrão MPEG-4

Uma emoção é visualmente representada através da sua expressão facial. Na literatura de animação facial, expressões faciais podem ser geradas usando diferentes abordagens. É possível sintetizar expressões faciais a partir de fotografias digitais capturadas (Pigh05) como também através de clonagem da expressão facial já sintetizada (Pyun03).

De uma forma geral, os diferentes trabalhos na geração de expressões faciais não apresentam características em comum nas suas malhas poligonais. Cada pesquisa desenvolve seu modelo facial próprio e aplica seus resultados neste modelo. Muitas vezes o único denominador comum existente nessas pesquisas é o modelo de emoção utilizado, que em sua maioria são as seis expressões universais propostas por Ekman (Ekma71).

Observando esse nicho, o padrão MPEG-4 (EbPe02) (MPEG02) apontou um caminho alternativo para modelagem de expressões faciais para quaisquer modelos de emoção, o qual é fortemente influenciado por estudos neurofisiológicos e psicológicos (Raou02). O padrão propôs um conjunto de pontos de controle para definição de um modelo facial, propondo assim uma malha facial poligonal que pode ser considerada universal. Esses pontos de controle, uma vez agrupados, formam parâmetros de animação facial que são utilizados em um *framework* MPEG-4 para geração das animações faciais.

O trabalho apresentado nesta tese de doutorado opta em fazer uso do modelo de face parametrizada proposto pelo padrão MPEG-4. Uma das justificativas para essa escolha deve-se ao fato que a padronização facial do MPEG-4 é a primeira proposta para um modelo de face universal (AbPe99). Uma vez desenvolvido um modelo de geração de expressões faciais dinâmicas sobre a face MPEG-4 torna-se possível explorar uma interoperabilidade entre diferentes aplicações que façam uso de uma personagem tridimensional sintética. Uma outra vantagem em usar um modelo de face padronizado é poder importar diferentes geometrias e modelos faciais que seguem a padronização como também exportar o resultado obtido neste trabalho para outras ferramentas de animação facial. Um outro ponto interessante também para esta tese de doutorado é que o uso do padrão MPEG-4 para animação facial foi pouco ex-

plorado, podendo este trabalho contribuir de forma eficiente e inovadora para o padrão. Enfim, vislumbrando trabalhar em um modelo universal, o padrão MPEG-4 foi estudado de forma detalhada para ser utilizado na síntese das animações faciais geradas.

Este capítulo comenta, de forma genérica, o espaço de expressões faciais. Uma vez definido que as expressões faciais a serem geradas estão conforme com a parametrização do padrão MPEG-4, o restante do capítulo apresenta esse padrão e como a estrutura da face é definida. Na Seção 1.3 é apresentada uma padronização que é tida como ideal para os sistemas de animação facial; este capítulo descreve como a especificação de face do padrão MPEG-4 se enquadra na parametrização facial tida como ideal. Ainda neste capítulo, as expressões faciais e o sincronismo das expressões labiais (visemas) com a fala são abordados. Por fim, os detalhamentos para especificação de um modelo de face conforme com o MPEG-4 e as regras para animá-lo são apresentados.

4.1

Espaço de Expressões Faciais

Expressões faciais e emoções estão interligadas através de alguma forma não-determinada mas, intuitivamente, pessoas procuram por sinais emocionais nas expressões faciais (Izar97) (Ekma94) (SmSc97). Essa é uma das razões de integrar um modelo de emoção em um sistema de animação facial: selecionar diferentes sinais comunicativos em relação ao estado emocional de um agente produzindo animações emocionalmente expressivas. Torna-se pertinente avaliar qual a melhor forma de se mapear (representar) uma dada emoção na expressão facial do personagem.

Quando uma pessoa está explicitando alguma emoção os músculos da face estão em completa atividade, podendo relaxar ou contrair. É interessante para um animador ter a percepção que nenhum músculo facial atua sozinho (Ratn03). Quando um músculo se contrai, o outro músculo, que se opõe ao da ação, se torna ativo para ajustar a contração do primeiro. A Figura 4.1¹ ilustra os músculos da face responsáveis pelas expressões faciais.

A face humana é capaz de formar cerca de 7000 expressões faciais (Ratn03). Existem 44 músculos na face mas nem todos eles são utilizados para formar uma expressão facial. Os músculos da expressão facial podem ser divididos como músculos superiores e inferiores. Os músculos superiores são responsáveis por mudar a aparência das sobrancelhas, da testa e das pálpebras superiores e inferiores. Já os músculos inferiores da face podem ser organiza-

¹Figura retirada do livro (PaWa96).

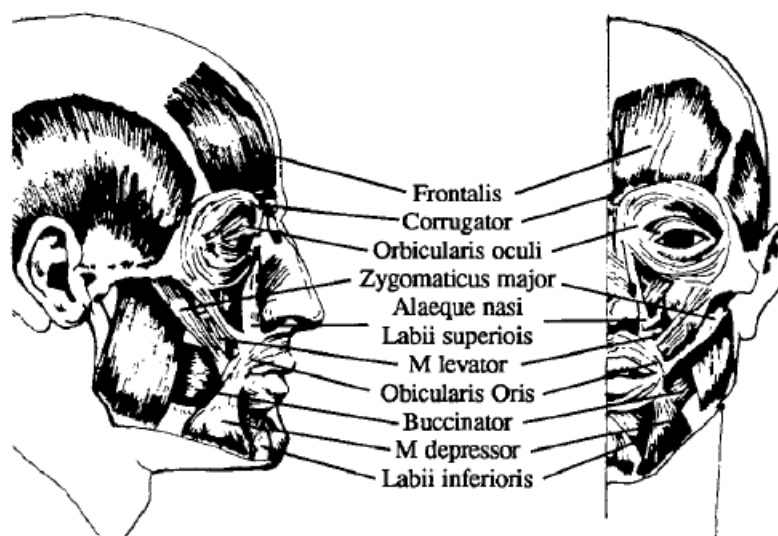


Figura 4.1: Músculos das expressões faciais.

dos em cinco grupos principais, sendo os músculos da boca os que possuem a interação muscular mais complexa (Wate87).

4.2

Apresentação do Padrão MPEG-4

O MPEG-4 é um padrão ISO/IEC pertencente à família MPEG e resultado de um esforço internacional de centenas de pesquisadores e engenheiros (EbPe02). O padrão segue o paradigma de orientação a objetos, e seu foco está na definição de documentos multimídia, estruturados através de cenas. O padrão vem sendo utilizado em aplicações gráficas interativas (conteúdo sintético) e aplicações multimídia interativas (Pere00), tendo sido adotado como padrão para o Sistema Brasileiro de TV Digital.

Uma cena MPEG-4 é dividida em objetos textuais, objetos gráficos, objetos de áudio e objetos de vídeo, podendo os dois últimos serem naturais ou sintéticos (Rodr03). Esses objetos podem ser colocados em qualquer posição da cena, tanto do ponto de vista temporal quanto do ponto de vista espacial, sendo possível também que transformações sejam aplicadas aos mesmos. Adicionalmente, é possível agrupar os objetos que compõem uma cena (também chamados de objetos de mídia) de modo a formar objetos mais complexos. Através desse agrupamento, ou composição, torna-se possível a interação entre os próprios objetos e a interação do usuário final com a cena.

Objetivando ter um padrão com grande abrangência, grupos de trabalho foram definidos para focar nos diferentes aspectos do MPEG-4. A partir do grupo de definição de objetos sintéticos *SNHC* (*Synthetic Natural Hybrid Coding*) foi criado o subgrupo de animação da face e do corpo (FBA - *The*

Face and Body Animation). Apesar do FBA compreender a especificação para o corpo e a face de uma personagem, esta tese foca na face e, portanto, o subgrupo FBA é apresentado apenas com os aspectos relevantes para animação de um modelo facial tridimensional.

4.3

Especificação da Face no Padrão MPEG-4

A especificação MPEG-4 de uma face consiste na definição de parâmetros de controle para uma animação facial. A iniciativa do grupo MPEG-4 é interessante por se tratar de uma primeira padronização para definição de um modelo de face parametrizada genérica (MPEG02). Essa iniciativa tem o propósito de permitir a interoperabilidade de fluxos MPEG-4 entre diferentes sistemas de animação. A animação facial do MPEG-4 faz uso de três importantes conceitos interligados que tornam viável uma padronização de faces virtuais e suas correspondentes animações: pontos característicos que definem uma face, a face no seu estado de neutralidade e a normalização de valores. Com o padrão MPEG-4, a animação de modelos faciais 3D pode ter seus dados gerados de forma sintética ou extraídos através da análise de faces reais.

O padrão MPEG-4 define um estado de neutralidade para uma face, como ilustra a Figura 4.2 ². Uma face no padrão MPEG-4 é considerada neutra se, e somente se, as seguintes características (propriedades) são respeitadas:

- Olhar na direção do eixo Z;
- Músculos da face relaxados;
- Pálpebras tangentes à íris;
- Pupila medindo um terço do diâmetro da íris;
- Lábios em contato, com linha horizontal e mesma altura no canto dos lábios;
- Boca fechada com os dentes superiores e inferiores se tocando; e
- Língua plana na horizontal com a ponta da língua tocando o limite entre os dentes superiores e inferiores.

Uma face para ser conforme com o padrão MPEG-4 também precisa estar normalizada. Para que seus parâmetros alcancem a normalização desejada, o padrão define unidades de parâmetros para animação facial, chamadas de FAPUs (*Face Animation Parameter Units*). As FAPUs são computadas a partir de distâncias espaciais entre pontos faciais chaves no seu estado

²Figura retirada do livro (Pela02).

neutro (Oste02), como ilustrado na Figura 4.2. Através das FAPUs torna-se possível interpretar os parâmetros de uma animação facial sobre qualquer modelo facial de forma consistente, produzindo resultados razoáveis em termos de expressão e da fala.

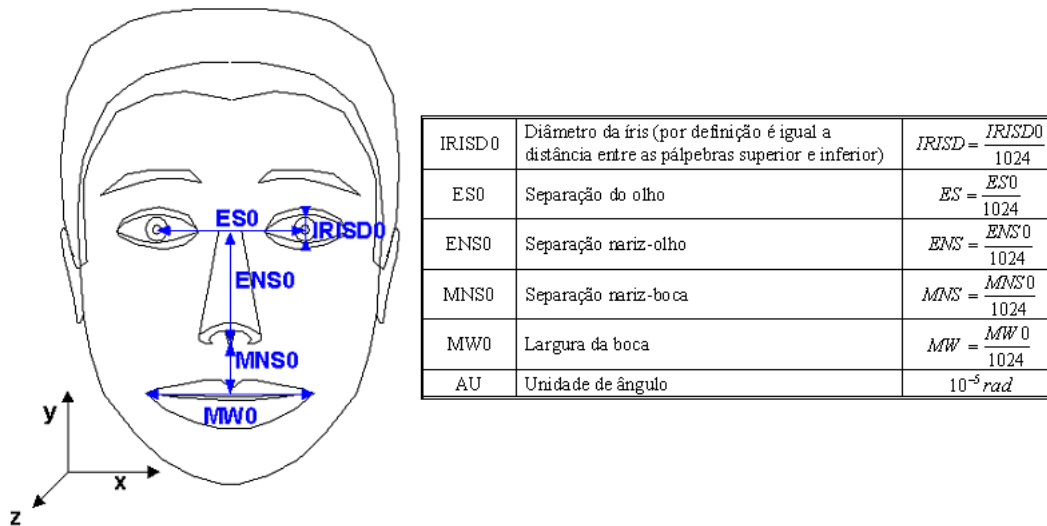


Figura 4.2: A face neutra e definição das FAPUs.

A partir da definição das FAPUs, o MPEG-4 especifica 84 pontos característicos (FPs - *Feature Points*) sobre a face em seu estado de neutralidade, como ilustra a Figura 4.3³. Os pontos característicos são usados tanto para definir a aparência de uma face quanto para animá-la, além de fornecer referência espacial para definição dos parâmetros da animação facial.

A partir da definição das FAPUs e dos FPs, o padrão MPEG-4 define dois tipos principais de parâmetros (AnPe99):

- Parâmetros da Animação Facial (FAPs - *Facial Animation Parameters*): permitem a animação de pontos chaves característicos em um modelo facial 3D, de forma independente ou em grupos, como também a representação de visemas e de emoções.
- Parâmetros de Definição Facial (FDPs - *Facial Definition Parameters*): permitem a configuração do modelo facial 3D a ser usado no receptor, através da adaptação de um modelo previamente disponível ou através do envio de um novo modelo. Ambos os tipos de modelo (novo ou já existente) são então animados através dos FAPs.

Enquanto FAPs fornecem continuamente informação visual associada ao comportamento do modelo 3D, FDPs fornecem a informação de configuração do modelo que é tipicamente definida uma única vez. Como este trabalho tem

³Figura retirada do livro (Pela02).

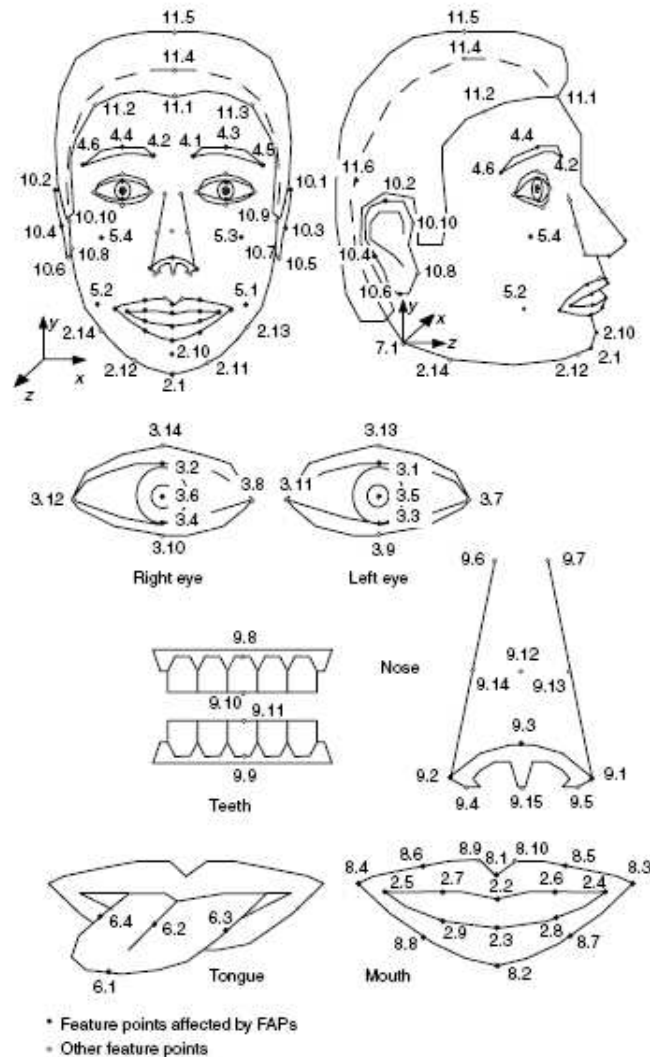


Figura 4.3: Pontos característicos do padrão MPEG-4.

seu foco no tratamento de expressões faciais dinâmicas em um modelo MPEG-4, os FAPs são vistos em maior detalhe do que os FDPs.

Apesar de alguns FPs não serem afetados por FAPs, todos os pontos característicos são imprescindíveis para definição do formato de um modelo de face (Oste02). Como é apresentado em detalhes na Seção 4.3.1, os FPs são organizados em grupos como bochechas, olhos e boca. Uma face conforme com o modelo MPEG-4 deve conhecer e respeitar a localização de cada ponto característico. Os 84 FPs com suas localizações, características e restrições estão descritos em detalhes no Apêndice A deste documento.

4.3.1 Parâmetros da Animação Facial

Os FAPs (*Facial Animation Parameters*) são baseados no estudo de ações minimamente percebidas e são fechados com relação às ações dos

músculos (MPEG02). O MPEG-4 definiu 68 parâmetros que são categorizados em 10 grupos relacionados às partes da face. Os FAPs representam um conjunto completo de ações faciais básicas, incluindo movimento da cabeça, da língua, dos olhos e controle da boca. Eles permitem a representação de expressões faciais naturais.

Os FAPs são necessários para animar faces de diferentes tamanhos e proporções. Todos os FAPs descrevem uma animação tendo como base a face neutra, que por sua vez é usada para normalizar os valores dos FAPs. A normalização dos FAPs é alcançada através das FAPUs. Para cada FAP o padrão define a FAPU apropriada, o grupo que o FAP pertence, a direção de movimento positivo e se o movimento de um ponto característico é unidirecional ou bidirecional. A Tabela 4.1 ⁴ descreve os 10 grupos de FAPs enquanto que o Apêndice B descreve, de forma mais completa, a especificação de cada um dos 68 FAPs.

Tabela 4.1: Grupo de FAPs

Grupo	Número de FAPs
1. visemas e expressões	2
2. maxilar, queixo, lábio inferior interno, canto dos lábios e meio dos lábios	16
3. globo ocular, pupilas e pálpebras	12
4. sobrancelha	8
5. bochechas	4
6. língua	5
7. rotação da cabeça	3
8. porções externas do lábio	10
9. nariz	4
10. orelhas	4

Dos 10 grupos de FAPs, o primeiro grupo (visemas e expressões) define parâmetros de alto-nível e são apresentados mais detalhadamente na Seções 4.5 e 4.4, respectivamente. Os outros nove grupos definem parâmetros de baixo-nível. Os FAPs de alto-nível são formados por conjuntos de FAPs de baixo-nível. Por exemplo, a expressão facial de alegria (feliz ou *joy*, no inglês) é representada através do valor 1 para o FAP de expressão (FAP 2) (Seção 4.4). Para compor o alto-nível da expressão facial “alegria”, vários FAPs são utilizados, por exemplo, o FAP que controla a abertura da boca, o FAP para relaxar a sobrancelha etc.

⁴Tabela retirada e traduzida do livro (Pela02).

Os FAPs de baixo-nível para uma animação MPEG-4 são valores (positivos ou negativos) que descrevem um movimento unidirecional ou bidirecional de pontos característicos usando uma FAPU como unidade de medida. Como exemplo, é possível supor que o FAP 34 tenha seu valor definido para 300. Esse FAP corresponde ao ponto de controle intermediário da sobrancelha direita, como ilustrado na Figura 4.4 ⁵ e especificado no Apêndice B. Supondo que foi aplicado o valor de 300 para esse FAP, o modelo da cabeça, como resposta, irá arquear o meio da sobrancelha direita para apontar acima da posição neutra. A FAPU ENS0 é usada para normalizar o valor 300, fazendo com que o movimento se torne uma fração de ENS0.

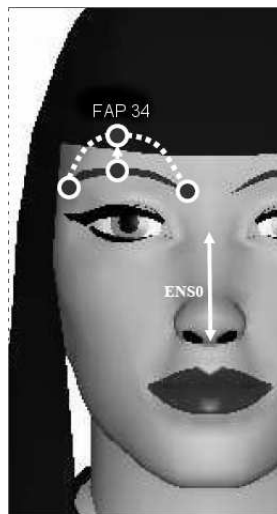


Figura 4.4: Exemplo da animação de um FAP.

É importante mencionar que usar FAPs de alto-nível (FAP 1 ou FAP 2) junto com os FAPs de baixo-nível (FAP 3 ao FAP 68), que afetam a mesma área dos FAP 1 e FAP 2, pode resultar em representações visuais não esperadas da face. Com isso, o padrão MPEG-4 determina que os FAPs de baixo-nível possuem prioridade nas deformações sobre os FAPs de alto-nível.

4.4

Expressões Faciais no Padrão MPEG-4

O padrão MPEG-4 dedica uma atenção especial para o FAP de alto-nível que representa as expressões faciais (FAP 2). O padrão define as suas expressões faciais como sendo as seis expressões faciais universais definidas por Ekman (Seção 2.2 e (Ekma71)), como ilustra a Figura 4.5 e conforme descritas na Tabela 4.2 ⁶. O uso de expressões faciais universais fortalece a credibilidade

⁵Exemplo retirado da Epitamedia, disponível em <http://www.epitamedia.com> (acesso em 08/Mar/2007).

⁶Figura e tabela retiradas do livro (Pela02).

para uma animação facial MPEG-4 através da naturalidade e do conhecimento das expressões disponíveis.

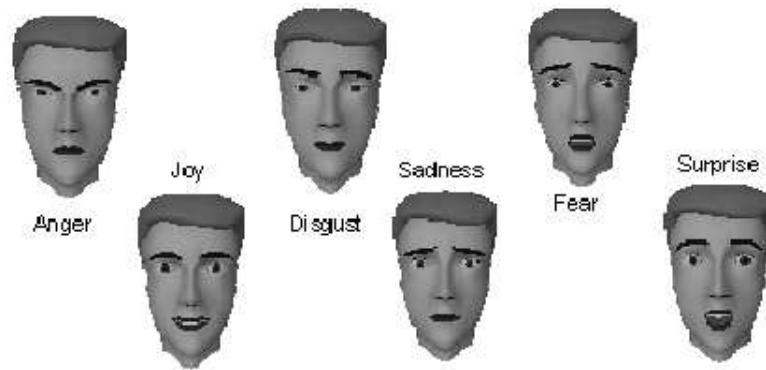


Figura 4.5: Expressões faciais no MPEG-4.

O padrão MPEG-4 define que a intensidade de uma expressão facial varia de 0 a 63 e que a combinação binária de expressões faciais é possível. Os parâmetros que especificam uma expressão facial final (*expr-final*) no MPEG-4 são: primeira expressão (*expr1*), intensidade da primeira expressão (*int1*), segunda expressão (*expr2*), intensidade da segunda expressão (*int2*), face inicializada (*init-face*) e definição da expressão (*expr-definition*) (MPEG02). O parâmetro *init-face* é usado para definir o conjunto da face neutra como expressão inicial. O parâmetro *expr-definition* é usado para adicionar os parâmetros faciais atuais às tabelas FAP com o objetivo de serem chamados novamente depois. Expressões e intensidades são os parâmetros mais importantes. A expressão facial final pode ser obtida através da Equação 4-1:

$$expr - final = expr1 \times (int1 \div 63) + expr2 \times (int2 \div 63) \quad (4-1)$$

Como indicado na Equação 4-1, a face tem a capacidade de mostrar (combinar) mais de uma expressão facial em um dado instante. Para o padrão, combinar expressões diferentes é um exercício simples de capturar as posições *in-between* de características faciais (definidas pelos pontos de controle) para duas expressões e desenhar a face resultante.

Combinar animações de baixo-nível com expressões é possível através da especificação de todas as animações como deslocamentos dos FPs a partir das suas posições na face neutra. Dessa forma, quando uma expressão é usada ela se torna a nova expressão neutra (temporariamente até que o parâmetro *init-face* seja definido) e os pontos de controle são deslocados a partir da nova posição inicial. O uso da posição neutra permite que a combinação entre FAPs de baixo-nível e expressões seja uma tarefa simples, podendo produzir um resultado natural.

Tabela 4.2: Expressões faciais primárias

Id	Nome da Expressão	Descrição Textual da Expressão Facial
0	Na	Face neutra.
1	Feliz (<i>joy</i>)	As sobrancelhas estão relaxadas. A boca está aberta e os cantos da boca estão para trás na direção das orelhas.
2	Triste (<i>sadness</i>)	A parte interna das sobrancelhas estão inclinadas para cima. Os olhos estão levemente fechados. A boca está relaxada.
3	Com raiva (<i>anger</i>)	A parte interna das sobrancelhas estão para dentro, para baixo e próximas. Os olhos estão abertos totalmente. Os lábios estão pressionados um contra o outro ou abertos mostrando os dentes.
4	Com medo (<i>fear</i>)	As sobrancelhas estão levantadas e próximas. A parte interna das sobrancelhas está inclinada para cima. Os olhos estão tensos e em alerta.
5	Repugnância (<i>disgust</i>)	As sobrancelhas e as pálpebras estão relaxadas. O lábio superior está levantado e curvado, normalmente de forma assimétrica.
6	Surpresa (<i>surprise</i>)	As sobrancelhas estão levantadas. As pálpebras superiores estão abertas totalmente e as inferiores estão relaxadas. O maxilar está aberto.

4.5

Visema e Lip-Sync no MPEG-4

Um visema (FAP 1) é a correlação visual de um fonema. Com essa definição é possível conceituar um visema como um formato particular da boca quando uma face reproduz (fala) um determinado fonema. Ainda não há um consenso do número de visemas necessários para produzir uma animação natural e confiável. A literatura não especifica um número mínimo ou máximo de visemas, mas sugere que esse não seja menor que sete (Ratn03) ou maior que dezoito (PaWa96) para o idioma inglês. Com isso, o MPEG-4 assumiu um valor intermediário, como ilustra a Tabela 4.3⁷. Vale observar que, até o momento, não há um estudo que defina a representação de visemas no padrão MPEG-4 para o idioma português.

Com o objetivo de permitir coarticulação da fala com movimento labial (CoMa93), no MPEG-4 o formato da boca de uma pessoa falando não é influenciado apenas pelo fonema corrente mas também pelo fonema ante-

⁷Tabela retirada do livro (Pela02). Os visemas e seus respectivos exemplos estão associados à língua inglesa.

Tabela 4.3: Visemas e os fonemas relacionados.

Id	Fonemas	Exemplos
0	nenhum	na
1	p,b,m	<i>put,bed,mill</i>
2	f,v	<i>far,voice</i>
3	T,D	<i>think, that</i>
4	t,d	<i>tip,doll</i>
5	k,g	<i>call,gas</i>
6	tS,dZ,S	<i>chair,join,she</i>
7	s,z	<i>sir,zeal</i>
8	n,l	<i>lot,not</i>
9	r	<i>red</i>
10	A:	<i>car</i>
11	e	<i>bed</i>
12	I	<i>tip</i>
13	Q	<i>top</i>
14	U	<i>book</i>

rior (Pela02). Assim, no MPEG-4 a transição de um visema para o visema subsequente é definida através de uma combinação (*blending*) dos dois visemas com um fator de peso associado a cada um deles, como ilustra a Equação 4-2⁸:

$$visemafinal = visema1 \times (blend \div 63) + visema2 \times (blend \div 63) \quad (4-2)$$

Como definido na Equação 4-2, o padrão MPEG-4 especifica que a intensidade de um visema varia de 0 a 63, analogamente às expressões faciais.

Na vida real um visema é modificado pela expressão facial corrente (Seção 4.4). Por exemplo, uma pessoa pode sorrir enquanto fala e esses visemas são diferentes dos visemas de uma expressão natural (neutra). Isso faz com que a combinação entre visemas e expressões seja importante para se ter uma animação facial realista. Com o objetivo de combinar visemas e expressões, os visemas precisam ser especificados como deslocamentos de FPs, relativos à face neutra.

É importante ter em mente que expressões não devem alterar o fechamento da boca durante a fala para garantir que os pontos característicos para cada visema sejam animados corretamente (MPEG02). Por exemplo, se uma expressão facial permite que a boca esteja aberta, então a personagem tem que tentar falar sem que seus lábios se encostem. Um outro exemplo está ilustrado

⁸Na Equação 4-2, o parâmetro *blend* é um valor inteiro no intervalo de 0 a 63 correspondendo à intensidade de cada visema em um determinado instante de fala.

na Figura 4.6⁹, onde há uma combinação correta de visemas e expressões para a expressão facial feliz (*joy*). As propriedades para a expressão feliz mostram que os cantos da boca devem ser movidos para criar um sorriso, mas a boca deve permanecer aberta.

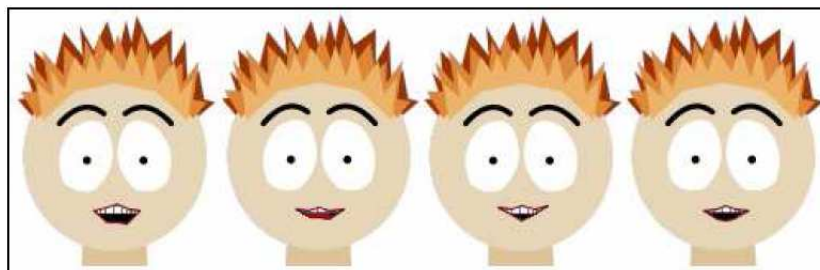


Figura 4.6: Visemas para expressão facial feliz.

4.6

Especificação de um Modelo de Face no Padrão MPEG-4

Para que uma animação facial MPEG-4 seja executada, o receptor MPEG-4 deve ser capaz de decodificar fluxos de dados que especificam e manipulam FAPs. O fluxo de FAPs precisa fornecer um modelo de face conforme com o padrão e que o receptor seja capaz de animá-lo (Oste02). Normalmente, isso é um modelo proprietário para o decodificador (Seção 4.3), enquanto que o codificador não conhece a aparência do modelo de face.

Para um modelo de face ser conforme com o MPEG-4 ele deve ser capaz de interpretar todos os FAPs como descrito na Seção 4.3. Com isso, o modelo da face deve ter um número de vértices pelo menos correspondentes aos pontos característicos necessários para a sua animação. O padrão especifica que um modelo de face conforme com o MPEG-4 deve ter no mínimo 50 vértices, mas, provavelmente, tal modelo (bastante simplista) não irá gerar um resultado visual satisfatório. Idealmente, é esperado que um modelo de face tenha pelo menos 500 vértices para se ter bons resultados (Oste02).

Um modelo de face proprietário pode ser construído em 4 passos (Oste02):

1. É necessário construir o formato do modelo da face e definir a posição dos pontos característicos (FPs) no modelo facial como especificado e ilustrado na Figura 4.3.

⁹Figura retirada do projeto *MetaFace Framework*, disponível em <http://www.metaface.computing.edu.au/> (acesso em 08/Mar/07).

2. Para cada FAP, é definido como cada ponto característico deve se mover. Para a maioria dos pontos, o MPEG-4 define o movimento em apenas uma dimensão (detalhes no Apêndice B deste documento).
3. Após a definição do movimento dos FPs para cada FAP, é definido como o movimento de um ponto característico afeta seus vértices vizinhos. Esse movimento do ponto característico para o movimento do vértice pode ser feito utilizando tabelas de busca (*lookup tables*) como FAT (OsHa97), deformação baseada em músculo (Karl92) (TeWa90) (Wate87), transformação de distância (LaPo99) ou clonagem a partir de modelos existentes (MaOs01) (NoNe01).
4. Para expressões faciais, o MPEG-4 fornece apenas dicas qualitativas de como elas devem ser desenvolvidas (Tabela 4.2). Similarmente, visemas são definidos por sons dados que correspondem a formatos de boca necessários (Tabela 4.3). FAP 1 e FAP 2 devem ser desenvolvidos com cuidado porque eles geralmente são usados para animações visualmente atraentes.

Seguindo os passos acima, um modelo de face está pronto para ser animado com FAPs MPEG-4.

A vantagem de se trabalhar com adaptações entre modelos faciais comparando a recuperar (fazer *downloads*) de modelos de face de um codificador para um decodificador é que o decodificador pode adaptar seu modelo potencialmente bem sofisticado ao formato desejado (Oste02). Por fim, é recomendável que o codificador sempre envie o conjunto completo dos pontos característicos para adaptação de um modelo facial. Enviar apenas uma quantidade parcial dos FPs pode resultar em representações incompletas do modelo facial. Para aplicações que especificam exatamente como o conteúdo deve ser apresentado no decodificador, é preferível utilizar a metodologia de recuperação do modelo de face usando o grafo da cena (Oste02).

4.7

Análise do Padrão MPEG-4 segundo a Parametrização Ideal para Animação Facial

Para que o padrão MPEG-4 de animação facial seja considerado como possuindo uma parametrização eficiente e “universal” é necessário analisar suas características a fim de verificar quão próximo o padrão está da parametrização tida como ideal apresentada na Seção 1.3. Essa análise foi feita em (PaFo02) e é apresentada a seguir.

- *Intervalo de possíveis faces e expressões*: Concentrando no intervalo de possíveis faces, os FDPs são escaláveis e permitem modificação no formato da face através dos pontos de controle (FPs), de textura adicional, ou ainda, em um nível mais alto, da malha poligonal como um todo, ambos de complexidade arbitrária. Isso faz com que o intervalo de faces que podem ser representadas torne-se praticamente ilimitado. Para o intervalo de emoções, os FAPs cobrem todas as principais articulações em detalhe como também a possibilidade de exagerar o movimento para animações estilo *cartoon*.
- *Facilidade de uso*: No geral, a especificação MPEG-4 é razoavelmente de baixo-nível e não tem a intenção de ser de uso direto. Isso leva a entender que o controle da face através de um conjunto de *sliders* correspondendo a cada FAP levaria a um processo custoso. Essa abordagem baixo-nível foi necessária, pois o objetivo é permitir que esses parâmetros sejam uma ligação universal entre vários sistemas e processos envolvendo animação facial.
- *Sutileza*: Simular várias expressões faciais delicadas é um dos desafios mais difíceis em uma animação facial. Os FAPs MPEG-4 cobrem a maioria das articulações na face e podem ser controlados em unidades suficientemente pequenas para permitir movimentos dificilmente perceptíveis. “Dificilmente perceptível” não significa a mesma coisa que sutil, mas o primeiro é a base para o segundo.
- *Ortogonalidade*: Com a definição dos movimentos dos pontos de uma face, os FAPs são naturalmente ortogonais.
- *Capacidade de ser base para uma abstração de alto-nível*: Os FAPs já fornecem parâmetros de alto-nível (visemas e expressões faciais) e os FAPs de baixo-nível podem ser usados para definir outras expressões. Os parâmetros de alto-nível proporcionam a combinação de parâmetros que podem ser usados para transições e simples coarticulação.
- *Previsibilidade*: É muito difícil objetivamente prever o resultado de um conjunto complexo de parâmetros em um sistema tão complexo quanto a face humana. O MPEG-4 possui um compromisso através de um subconjunto previsível - os movimentos dos FPs (pontos de controle).

- *Portabilidade*: Os FAPs são normalizados em relação à face que eles são aplicados e assim eles são portáteis de uma face para outra. Experimentos durante o desenvolvimento do padrão mostraram uma reprodução razoável das mesmas expressões de alto-nível em diferentes modelos faciais (PaFo02).
- *Possibilidade de medir os parâmetros*: O valor que cada ponto característico assume durante a execução de um movimento são medidos da mesma maneira (com a mesma unidade).
- *Eficiência (largura de banda)*: O MPEG-4 define a codificação dos parâmetros em um fluxo de *bits* altamente eficiente que pode ser usado no próprio sistema ou como parte de um sistema MPEG-4. O sistema de codificação otimiza o fluxo de *bits* permitindo a transmissão dos parâmetros que apenas são realmente usados. O intervalo típico de taxas de *bits* varia de 3 a 6 kbit/s para uma animação facial completa com 25 a 30 quadros por segundo. Utilizando apenas os parâmetros de alto-nível, a taxa de *bit* reduz a aproximadamente 0.3 kbits/s. Esses resultados são apropriados para qualquer plataforma ou ambiente nos dias de hoje.

Como pode ser constatado, a especificação para animação facial do MPEG-4 não é a parametrização ideal mas está bem próxima de atender os requisitos colocados em (PaFo02) e apresentados na Seção 1.3.