

FaceAnim: Um framework Interativo para a Geração de Animações Expressivas de Face e Olhos

Amanda Gabriela G. Vieira^{1*}

Rossana B. Queiroz^{2†}

Daniel Camozzato^{3‡}

Universidade do Vale do Rio dos Sinos (UNISINOS), Escola Politécnica, Brasil



Figura 1: Alguns dos *keyframes* utilizados no FaceAnim.

RESUMO

Este artigo apresenta uma proposta de *framework* extensível e interativo para animações de face e olhos para mídias interativas, tais como jogos eletrônicos e simuladores. A principal motivação para o desenvolvimento do *framework* FaceAnim é permitir que usuários sem conhecimentos de animação ou programação sejam capazes de gerar expressões realistas e empáticas em uma face 3D com facilidade. Para tanto, primeiro aplica-se uma máscara de pontos de controle inspirada na CANDIDE-3 [1] ao modelo. Em seguida, os pontos de controle referentes à máscara são ajustados. Finalmente, as expressões são geradas através da aplicação e manipulação de parâmetros gerados para cada parte do rosto. Será integrado também um modelo de geração automática de olhares expressivos, que leva em consideração o estado emocional do personagem. As animações geradas serão avaliadas qualitativamente com usuários, analisando a percepção da expressividade dos modelos. O *framework* FaceAnim está sendo desenvolvido na ferramenta Unity 3D¹ com o objetivo de facilitar a geração de animações expressivas em personagens de jogos eletrônicos e outras aplicações interativas.

Palavras-chave: animação facial, animação de olhos, *framework* extensível, Unity 3D, CANDIDE, face parametrizada, *blendshape*, *keyframe*.

1 INTRODUÇÃO

Agentes virtuais animados estão cada vez mais presentes em todos os tipos de mídia existentes atualmente. Com a crescente utilização

*e-mail: amygaby.gv@gmail.com

†e-mail: rossanaqueiroz@unisinis.br

‡e-mail: danielcamozzato@unisinis.br

¹Unity 3D. Plataforma de criação de jogos digitais. Em <https://unity3d.com/pt/unity>

de agentes virtuais, tanto em mídias interativas quanto em não-interativas, desencadeou-se uma busca por maior realismo. Novas técnicas de animação facial de agentes virtuais são desenvolvidas ou melhoradas frequentemente, com o objetivo de automatizar ao máximo os processos de animação. Porém, para animações faciais mais expressivas e realistas, ainda é necessária uma grande intervenção de profissionais animadores.

Este trabalho propõe o desenvolvimento de um *framework* de animação facial e de olhos robusto, porém, fácil de usar. O principal objetivo deste trabalho é permitir que usuários sem nenhum conhecimento prévio em programação e animação sejam capazes de gerar animações faciais empáticas e realistas em um agente virtual. Como base para o desenvolvimento, pretende-se usar e adaptar a metodologia proposta por Queiroz *et al.* [11, 13], utilizando o motor de jogos Unity 3D e provendo uma interface fácil e intuitiva para a geração das animações, podendo ser disponibilizada como uma *asset* de Animação Facial.

A arquitetura proposta para o *framework* compreende três estágios principais: *i*) aplicação de uma máscara de pontos de controle inspirada na CANDIDE-3 [1] na face neutra do agente virtual, *ii*) mapeamento dos parâmetros de animação e, por fim, *iii*) geração de expressões através da manipulação dos pontos de controle. As animações serão geradas através de *blendshapes* entre as expressões.

Este artigo está organizado da seguinte maneira: Seção 2 apresenta os principais conceitos teóricos para o projeto do *framework*. Em seguida, os trabalhos inspiradores e embasadores para o modelo proposto são brevemente apresentados na Seção 3. Os estágios do *framework* FaceAnim serão descritos na Seção 4 e alguns resultados preliminares obtidos do atual estágio de desenvolvimento são apresentados na Seção 5 e discutidos na Seção 6.

2 FUNDAMENTAÇÃO TEÓRICA

A nomenclatura *blendshape* foi vista a primeira vez no *software* de modelagem 3D Maya². Entende-se por *blendshape* a

²Autodesk Maya. *Software* de modelagem 3D.

implementação de uma técnica que permite deformar uma malha 3D, até esta se assemelhar a formas previamente definidas³. De maneira mais formal, entende-se por *blendshape* a combinação entre duas ou mais malhas 3D para gerar uma malha intermediária [6]. Essa combinação é dada associando-se pesos de o quanto cada *keyframe*⁴ contribui para gerar a malha final. A Figura 2 mostra a combinação de quatro *keyframes* e a face resultante desta combinação.

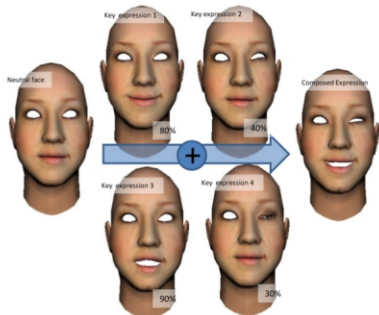


Figura 2: Exemplo de combinação de quatro *keyframes* e seus pesos em relação à malha de expressão neutra (à esquerda). A face resultante (à direita) é dada pela soma ponderada de todos os *keyframes*. Fonte: [10]

Em animações computadorizadas, entende-se por *keyframe* um desenho que define o início e o fim de uma animação. Neste artigo, *keyframes* serão tratados como "expressões-chave" pré-definidas da face do agente virtual 3D (ver Figura 9 na Seção 5). A quantidade de *keyframes* utilizados para a obtenção de animações pode variar, normalmente seguindo a lógica de "quanto mais *frames* intermediários, mais suave a animação resultante será".

Para a obtenção de expressões faciais realistas em agentes virtuais, diversos estudos sobre o comportamento e a fisionomia da face humana foram realizados. O mais famoso é o estudo feito por Paul Ekman [3], que definiu um conjunto de Unidades de Ação (*Action Units* – AUs), que representam conjuntos de músculos que mais contribuem na movimentação das faces durante a execução de uma expressão facial. Esse modelo, chamado de FACS (Facial Action Coding System), é a base da parametrização de muitos modelos de Animação Facial, assim como de algoritmos para rastreamento de faces em Visão Computacional [10]. Inspirado nas AUs de [3], Ahlberg propõe um modelo simplificado de face, chamado CANDIDE [1]. A Figura 3 ilustra esse modelo. Essa face simplificada pode ser usada como uma estrutura auxiliar em sistemas de animação [13].



Figura 3: Representação gráfica da máscara CANDIDE-3 [1], que é uma simplificação dos principais pontos que caracterizam e movimentam uma face.

³<http://www.autodesk.com.br/products/maya/overview>

⁴*Keyframes* ou, no contexto deste trabalho, "expressões-chave", também conhecidos por *morph targets*.

O modelo FACS [3] descrito por Paul Ekman e Wallace V. Friesen é uma metodologia que visa medir qualquer expressão observável em uma face humana. O modelo propõe que cada movimento observável da face seja tratado como uma "unidade de ação" (AU), e, que cada unidade de ação pode ser combinada com outras, permitindo que sua manipulação seja feita separadamente ou em grupos, como mostra a Figura 4.

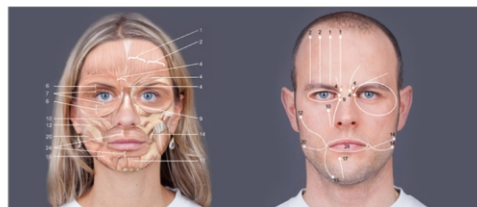


Figura 4: Exemplo das unidades de ação descritas pelo modelo FACS. Fonte: [4]

3 TRABALHOS RELACIONADOS

Atualmente, existem diversos trabalhos que propõem modelos computacionais para produzir animações faciais rapidamente, tanto por motion-capture e transferência de expressões do usuário para o agente virtual [14, 8, 13], quanto diretamente por *keyframes* e *blendshapes* [7, 12]. De fato, nos últimos anos, a indústria tem investido em prover soluções semi-automatizadas para a criação de animações faciais, como o Facerig⁵, Face Plus⁶, Faceware Retargeter⁷ e o Faceshift⁸. A maioria dos programas de animação facial requer conhecimentos prévios em animação ou programação por parte do usuário. No entanto, o modelo computacional Responsive Face de Ken Perlin [9], que é a maior inspiração para o desenvolvimento do *framework* FaceAnim, não requer que o usuário tenha conhecimentos prévios de animação ou programação para usá-lo.

Para que o *framework* FaceAnim atinja os resultados almejados em termos de facilidade de utilização, optou-se por criar uma interface que oferece clareza e simplicidade de uso, baseada na ferramenta *Responsive Face*, proposta por Ken Perlin [9] (ver Figura 5). Apesar de não estar no estado-da-arte, a interface para manipulação dos parâmetros deste trabalho é simples e intuitiva, permitindo que usuários inexperientes possam explorar e criar diferentes expressões faciais.

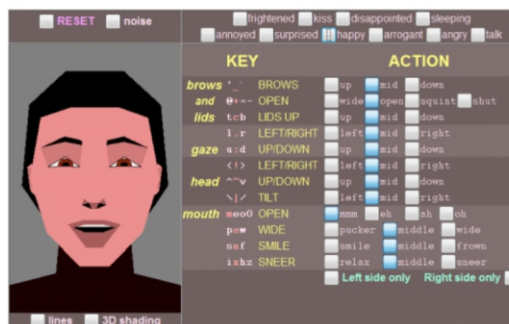


Figura 5: Responsive Face, de Ken Perlin é a principal inspiração de interface para o projeto do FaceAnim. Fonte: [9]

⁵facerig.com

⁶www.mixamo.com/faceplus

⁷facewaretech.com/products/software/retargeter

⁸www.faceshift.com

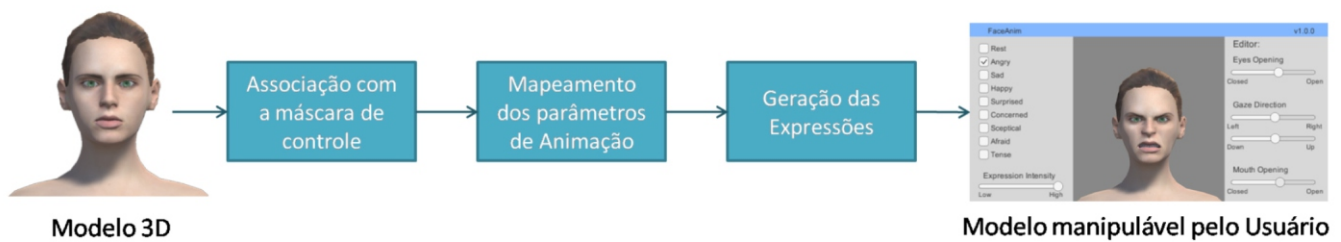


Figura 6: Arquitetura do *framework* FaceAnim.

A Face Responsiva (*Responsive Face*) de Ken Perlin [9] é uma ferramenta de animação facial que, como no *framework* FaceAnim, utiliza-se da atribuição e manipulação de parâmetros de animação em uma face 3D com poucos vértices (que formam uma estrutura similar à CANDIDE-3) para gerar animações. O FaceAnim pretende, no entanto, se adaptar em diferentes resoluções de faces 3D, como no trabalho de Queiroz *et al.* [13].

Diferentemente da ferramenta proposta neste trabalho, a Face Responsiva de Ken Perlin [9] não possui um módulo de animação automática de olhos, e sim apenas parâmetros manipuláveis pelo usuário como a direção do olhar e abertura das pálpebras. O modelo de animação do FaceAnim pretende incorporar e estender o modelo de geração de olhares expressivos proposto por Queiroz *et al.* [11], que adicionou parâmetros emocionais para o olhar, utilizando-se como motor de geração de micro-expressões visuais o modelo estatístico de Lee *et al.* [5].

O modelo de parametrização do FaceAnim, assim como o do Responsive Face, é baseado no modelo FACS descrito por P. Ekman [3]. Na arquitetura proposta para o *framework* FaceAnim, utiliza-se de uma máscara de controle baseada na versão mais recente da máscara CANDIDE (versão 3) [1, 13] para fazer o mapeamento dos parâmetros de animação de uma forma mais genérica, uma vez que esta máscara é construída através da abstração das principais estruturas faciais (músculos e ossos) que movimentam a face.

Outro diferencial do *framework* FaceAnim em relação ao *framework* de Ken Perlin [9], é a possibilidade de controlar a intensidade de cada expressão, o que permite ao usuário gerar uma gama maior de animações. O controle de intensidade das expressões dá-se pela manipulação dos parâmetros atribuídos às diferentes partes da face.

4 METODOLOGIA

A metodologia proposta para o desenvolvimento do *framework* FaceAnim possui três estágios principais, seguidos de um estágio final em que o usuário interage com a interface para geração de animações, como mostrado na Figura 6.

Os três estágios principais são:

1. Associação com a máscara de controle: nesse estágio, os vértices da face serão registrados aos da máscara de controle, como em [13]. Para identificar os pontos principais da face (como olhos, sobrancelhas e boca, por exemplo) as posições dos vértices da face serão comparadas com as posições dos vértices parametrizados da máscara de controle, usando algum método de detecção/registo de pontos, como por exemplo, o de Creusot *et al.* [2]. O *framework* repetirá este processo até que todos os vértices da face do modelo 3D estejam corretamente alinhados com os vértices da máscara de controle.
2. Mapeamento dos parâmetros de animação: os parâmetros serão embasados no modelo FACS [4]. Nessa etapa, será feito

um mapeamento dos pontos da máscara para os AUs. Com isso, cada vez que o usuário manipular os parâmetros na interface, serão estabelecidas as expressões-chave que deverão ser misturadas. Essas expressões-chave poderão ser providas pelo usuário ou geradas através de algum modelo de geração de blendshapes por exemplo, como o de Queiroz *et al.* [13].

3. Geração das Expressões: a partir da interface simples e intuitiva que está sendo desenvolvida, inspirada no Responsive Face [9], o usuário poderá selecionar as expressões, suas intensidades (pesos), o comportamento dos olhos e assim definir os *keyframes*, para a geração dos *frames* intermediários.

Para facilitar a utilização por usuários inexperientes, os três estágios iniciais do *framework*, assim como a animação dos olhos do agente virtual, deverão ter intervenção mínima do usuário. Para isso, pretende-se incorporar e adaptar a metodologia proposta por Queiroz *et al.* [13] para geração da face animável.

Para a movimentação dos olhos, pretende-se incorporar os parâmetros de animação propostos por [11] para os seguintes comportamentos: concentrado, aflito, ironia e desconforto. O modelo integrado de animação automática de olhos leva em consideração o tipo de expressão selecionado pelo usuário, gerando animações de olhos realistas e empáticas [5, 11].

5 RESULTADOS PRELIMINARES

O *framework* FaceAnim está em processo de desenvolvimento na plataforma Unity 3D. Atualmente se encontra em sua primeira versão, e, já conta com algumas expressões pré-configuradas (ver coluna esquerda da Figura 7), para agilizar o processo de animação e tornar mais intuitiva a manipulação da malha 3D.



Figura 7: Interface do *framework* FaceAnim em sua primeira versão.

Além disto, o *framework* permite que o usuário manipule a intensidade das expressões, sendo Low “pouco intenso” e High “muito intenso” (ver Figuras 7 e 8).



Figura 8: Comparativo de intensidade de uma mesma expressão.

Para tornar as expressões pré-configuradas personalizáveis, além dos níveis de intensidade, é possível, na aba "Editor" (ver coluna direita da Figura 7), controlar o nível de abertura dos olhos e boca ("Eyes Opening" e "Mouth Opening") e a direção do olhar ("Gaze Direction"). Por fim, as animações serão geradas em tempo real por *blendshapes* entre as expressões escolhidas pelo usuário. No momento, para obter animações com transições mais suaves entre diferentes expressões, é utilizada uma expressão de descanso ("Rest") como o *keyframe* intermediário (ver Figuras 1 e 9).



Figura 9: transição entre expressões; utilização do keyframe "Rest" para tornar a transição mais suave.

6 CONSIDERAÇÕES FINAIS

Este trabalho apresenta uma proposta de arquitetura de um *framework* de Animação Facial cujo principal foco é a geração de animações empáticas de uma maneira simples e intuitiva, inspirado na ferramenta Responsive Face [9]. A implementação está em estágio inicial de desenvolvimento, e deverá incluir módulos automáticos e semi-automáticos para a parametrização da face 3D e geração de comportamentos expressivos de olhos, baseado no *framework* proposto por Queiroz *et al.* [13].

O *framework* já possui expressões pré-configuradas e a possibilidade de personalizá-las. A automatização dos estágios iniciais do *framework*, tal como a aplicação dos *blendshapes*, encontram-se em processo de desenvolvimento.

Como trabalho futuro, o *framework* será otimizado para que permita que o usuário carregue qualquer malha 3D de faces, gerando

automaticamente expressões personalizadas para o modelo carregado. Além disso, será integrado um modelo de movimentação da cabeça, que levará em consideração a fisionomia da malha carregada pelo usuário. Pretende-se também, adicionar mais parâmetros customizáveis pelo usuário, como a movimentação das sobrancelhas e expressões do nariz.

Para validar o *framework* FaceAnim serão realizadas avaliações qualitativas com usuários. Os critérios a serem avaliados serão a facilidade de uso do *framework*, o nível de realismo das expressões geradas e a qualidade das animações resultantes. Os resultados obtidos através das avaliações serão mensurados e os parâmetros avaliados no *framework* serão ajustados. Este processo se repetirá até que o nível de aprovação das expressões e animações geradas seja satisfatório.

REFERÊNCIAS

- [1] J. Ahlberg. Candide-3-an updated parameterised face. Technical Report LiTH-ISY-R-2326, Dept. de Engenharia Elétrica, Universidade de Linköping, Suécia, 2001.
- [2] C. Creusot, N. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *Int. J. Comput. Vision*, 102(1-3):146–179, Mar. 2013.
- [3] P. Ekman and W. Friesen. *Facial Action Code System*. Consulting Psychologists Press, Inc., Palo Alto, CA, 1978.
- [4] P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system – the manual. Sample available in <http://face-and-emotion.com/dataface/facs/manual/TitlePage.html>, 2002.
- [5] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. *ACM Trans. Graph.*, 21(3):637–644, July 2002.
- [6] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. Pighin, and Z. Deng. Practice and Theory of Blendshape Facial Models. In S. Lefebvre and M. Spagnuolo, editors, *Eurographics 2014 - State of the Art Reports*. The Eurographics Association, 2014.
- [7] H. Li, T. Weise, and M. Pauly. Example-based facial rigging. *ACM Trans. Graph.*, 29:32:1–32:6, July 2010.
- [8] H. Li, J. Yu, Y. Ye, and C. Bregler. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph.*, 32(4):42:1–42:10, July 2013.
- [9] K. Perlin. Responsive face. SIGGRAPH 97 Technical Sketch, 1997. Disponível em <http://mrl.nyu.edu/perlin/experiments/face demo/>.
- [10] R. B. Queiroz. *Personalizing Facial Macro- and Microexpressions*. PhD thesis, Pontifícia Universidade Católica do Rio Grande do Sul, Faculdade de Informática, Porto Alegre, Brasil, 8 2015.
- [11] R. B. Queiroz, L. M. Barros, and S. R. Musse. Providing expressive gaze to virtual animated characters in interactive applications. *Comput. Entertain.*, 6(3):41:1–41:23, Nov. 2008.
- [12] R. B. Queiroz, S. R. Musse, and N. I. Badler. Investigating macroexpressions and microexpressions in computer graphics animated faces. *Presence*, 23(2):191–208, Aug 2014.
- [13] A. B. Rossana B. Queiroz and S. R. Musse. An adaptive methodology for facial expression transfer. In *Proceedings of SBGames 2015*, pages 158–167, 2015.
- [14] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *ACM Trans. Graph.*, 30(4):77:1–77:10, July 2011.