

Pesquisadores usam a animação facial por computador para desenvolver ferramentas com fins sociais

Faces virtuais que falam por si

Fotos: Divulgação/Antoninho Perri

O professor José Mario De Martino, coordenador da linha de pesquisa: promovendo de maneira amigável e divertida a interação homem-máquina



LUIZ SUGIMOTO

sugimoto@reitoria.unicamp.br

Visitante desavisado pode pensar que o professor José Mario De Martino e seus orientados estão empenhados na criação de personagens de videogame, em frente aos computadores da Faculdade de Engenharia Elétrica e de Computação (FEEC) da Unicamp. É verdade que se trata de animação por computador, mas o que eles buscam, a partir do modelo virtual de uma face humana, é o máximo de sincronia e realismo na movimentação articulatória associada à fala, a exemplo da articulação temporomandibular (ATM) e do tecido dos lábios.

Segundo José Mario De Martino, as pesquisas no mundo voltadas para a animação facial por computador já se estendem por quase 40

Agentes transmitem notícias, aulas e palestras

anos. “Passou-se de representações hoje consideradas grosseiras, que mal conseguiam piscar os olhos e abrir a boca de forma realista, para modelos sofisticados que procuram reproduzir a realidade biodinâmica e fisiológica da face”.

O docente da FEEC explica que cada movimentação facial é descrita por um modelo implementado por um algoritmo e codificado em linguagem de programação. “Já nas técnicas tradicionais de animação, como o desenho animado e a chamada animação assistida por computador, é um animador, baseado em sua experiência, habilidade e sensibilidade, que define e controla a movimentação”.

Com o domínio de técnicas para especificar e controlar posição, forma e aparência de uma face sintética, sincronizando esta movimentação com a fala, torna-se possível estender o leque de aplicações da animação por computador para além do cinema e dos jogos eletrônicos, como por exemplo, na compressão de dados para videoconferência e em ferramentas educativas.

“Por meio desses agentes virtuais podemos transmitir notícias, palestras ou aulas, sem o inconveniente de ter que regravar a fala de uma pessoa real, em caso de mudanças no conteúdo; simplesmente, geramos a nova sequência de imagens desejada. Espera-se que no futuro eles sejam incorporados também em máquinas de atendimento automático, computadores pessoais e eletrodomésticos, promovendo de maneira natural, amigável e divertida a interação homem-máquina”, prevê Martino.

De acordo com o pesquisador da Unicamp, há duas grandes vertentes na animação facial gerada por computador. Uma é a animação baseada em modelo geométrico tridimensional (3D) e a outra em imagens fotográficas de poses-chave (2D). “Em 3D, usamos descrições matemáticas para gerar o modelo geométrico tridimensional da face e manipulá-lo de forma a reproduzir a movimentação facial. No contexto 2D, recorreremos à justaposição de imagens a fim de produzir a animação”.

Aspecto pioneiro – José Mario De Martino ressalta o caráter multidisciplinar da linha de pesquisa de seu grupo na FEEC, que trabalha com conceitos e conhecimentos da computação gráfica, da produção e percepção visual da fala e da linguística. A análise e descrição da movimentação articulatória visível para



Modelo virtual de um apresentador de notícias: movimentação dos olhos e das sobrancelhas

representar os segmentos fonéticos do português falado no Brasil foi um esforço que empresta característica pioneira a essas pesquisas.

Apenas em relação ao nosso português, já são muitas as nuances na movimentação facial. Uma sugestão de Martino é que o leitor observe a diferença na articulação dos seus lábios para pronunciar uma mesma letra, “p”, primeiro na palavra “paro”, e depois em “puro”. Juntam-se outras movimentações faciais associadas à fala, como um alçar de sobrancelhas, um balançar de cabeça, um distorcer da boca, e teremos uma idéia da complexidade do conjunto de sinais com que os pesquisadores têm de trabalhar.

“O que tentamos, no modelo virtual, é mimetizar os mecanismos de comunicação face a face que usamos habitualmente. A informação visual na face muitas vezes complementa e auxilia a compreensão da mensagem transmitida na fala. Desde a infância, somos muito bem treinados para perceber variações sutis na expressão facial. Sabemos, assim, se o outro está atento ou não, se é nossa vez de falar quando ele se cala”, explica o pesquisador.

Martino acrescenta que os movimentos em faces reais podem ser agrupados em cinco categorias: articulatórios, emocionais, conversacionais, fisiológicos e contorcidos. Os articulatórios são produzidos pela movimentação do aparelho fonador; os emocionais expressam

tristeza, alegria, medo, raiva; os sinais conversacionais acompanham a fala enfatizando palavras e pontuando o discurso; entre os fisiológicos está o piscar de olhos para umedecer a córnea; e a contorção facial, menos usual, é a chamada careta.

Agentes virtuais – O professor apresenta um primeiro trabalho desenvolvido em 2005, uma face virtual simples, reproduzida em computador, acompanhada de voz sintética. Para a modelagem da movimentação facial foi feita uma extensa análise de imagens de vídeo de um locutor real. Foram horas e horas de gravação, com medidas fotogramétricas das trajetórias tridimensionais de pontos marcados na face. Esta análise revelou um conjunto de 29 visemas – que são os movimentos faciais feitos para pronunciar os sons existentes em uma língua.

“No intuito de avaliar a contribuição da tecnologia para a compreensão da fala, gravamos o áudio distorcido com ruídos até que ele se tornasse quase ininteligível. Reunimos um grupo de 35 pessoas e exibimos tanto o vídeo natural como o modelo virtual, a fim de que dissessem o que ouviam. O ganho de inteligibilidade foi de aproximadamente 50% para o vídeo e de 20% para o modelo, demonstrando que este é um caminho interessante a percorrer”, argumenta o docente.

O trabalho mais recente mostra o modelo virtual de um apresenta-

dor de notícias, escolha que não foi gratuita, já que na vida real, em princípio, este profissional não deve expressar emoções, mantendo-se sempre neutro. Nas mãos dos pesquisadores, o agente virtual está ganhando movimentação dos olhos e das sobrancelhas, e possivelmente terá rugas. “Ele já é capaz de reproduzir as frases que quisermos. Nos próximos dias, vamos avaliar a eficácia do modelo na transmissão da informação e sua contribuição para uma maior memorização, exibindo notícias a um grupo de voluntários”.

No celular – José Mario De Martino e seus orientados também estão desenvolvendo um agente virtual para telefones celulares. Uma aplicação mais imediata seria para diminuir os custos de videoconferências: ao invés de se transmitir um vídeo, seria possível enviar apenas o texto, gerando-se na outra ponta a imagem e o áudio em um celular. “A única exigência seria um celular com capacidade suficiente de processamento”.

No futuro, esta pesquisa, que teve financiamento da Fapesp, também permitiria a substituição das atuais mensagens de texto por um rosto virtual, capaz de ler todo texto digitado em um editor, ou mesmo informar sobre a previsão do tempo, horários de vôos nos aeroportos ou preços de produtos nos supermercados.

Pesquisa visa ajudar deficientes auditivos

O professor José Mario De Martino alimenta expectativa especial quanto à aplicação da animação facial por computador em ferramentas que auxiliem pessoas portadoras de deficiências, como de audição. “Se conseguirmos passar ao deficiente auditivo pistas visuais associadas à movimentação articulatória da fala, isto facilitaria a compreensão do áudio. Além disso, este deficiente é bem melhor treinado para a leitura labial, que todos nós fazemos, consciente ou inconscientemente, como quando estamos conversando em local barulhento”.

Martino observa que o esforço para reproduzir esses movimentos articulatórios também traz ganhos de conhecimento, principalmente em relação ao processo de produção da fala e das pistas visuais, que pode contribuir também na área médica. “Não estamos focados nisso, mas um modelo virtual refinado, que reproduza com perfeição a produção da fala, pode ser usado em exercícios para pessoas com problemas de articulação, aliviando o trabalho do fonoaudiólogo”.

O docente adianta que seu grupo está entrando em uma linha de pesquisa complementar, passando do rosto para o corpo, com o desenvolvimento de um agente virtual sinalizador de Libras (língua de sinais brasileira). “Há uma dificuldade fundamental, pois a Libras ainda não está perfeitamente compreendida, inclusive pelos lingüistas. Já na língua falada, temos todo um conhecimento, acumulado durante séculos, da fonética e da parte articulatória em que nos apoiar”.

Segundo Martino, ainda não há consenso entre os lingüistas quanto à possibilidade de se quebrar a Libras como na língua oral, que é pensada como segmentos sonoros que são justapostos para descrever a fala. “Se isto não for verdade para a língua de sinais, significa que precisaremos de um conjunto infinito de elementos para descrevê-la. Outra dificuldade é que, se na língua oral existe um alfabeto fonético inclusive para transcrever variações de pronúncia, em Libras não estamos nem perto disso; não sabemos ainda até que ponto um sinal levemente diferente continua sendo compreendido pelo outro”.