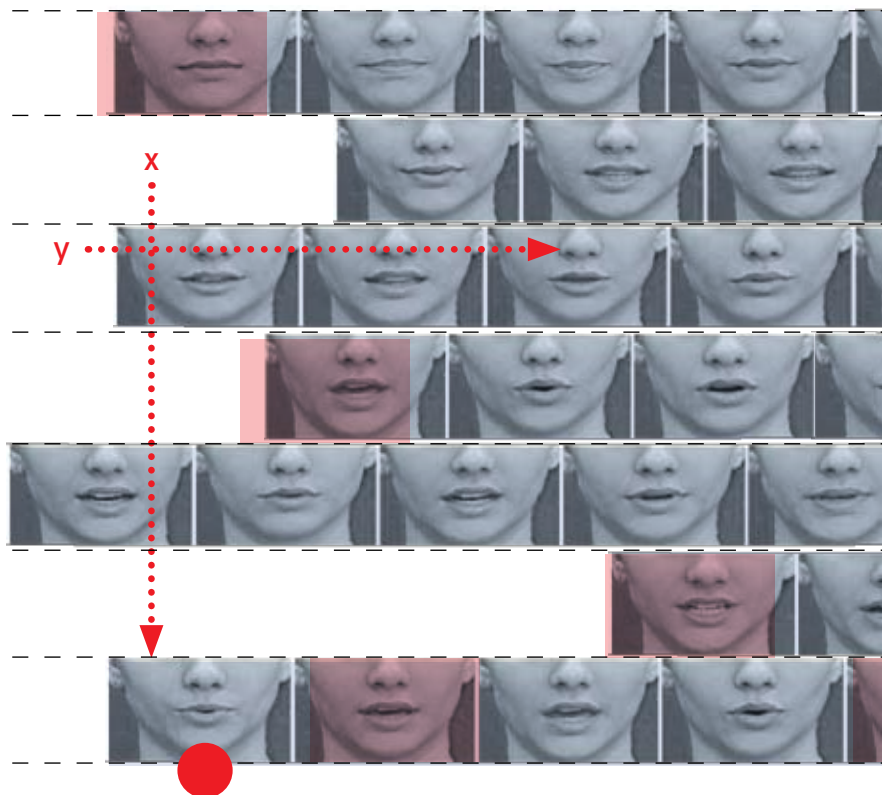


Atendentes virtuais

Sistema de animação facial pode tornar mais intuitiva interação com o computador

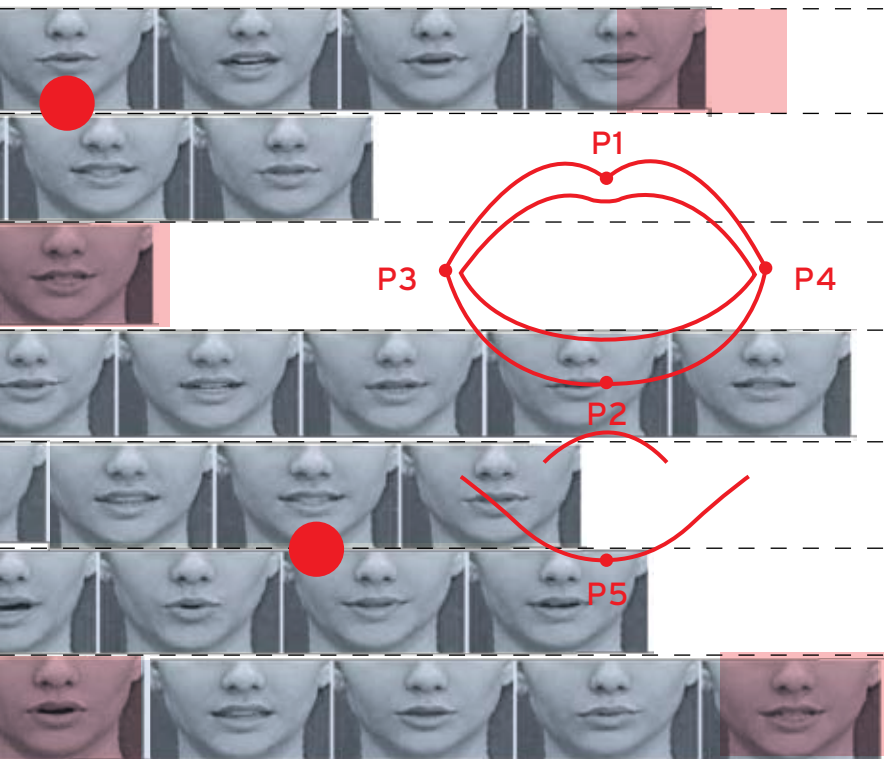
Em breve, retirar dinheiro, fazer um pagamento ou realizar qualquer movimentação bancária em uma máquina de atendimento automático poderá ser uma experiência diferente do que é hoje. No lugar de digitar comandos e interagir friamente com a tela do caixa eletrônico, o usuário terá uma vivência de atendimento próxima à de um funcionário de carne e osso da agência bancária. Ele será “atendido” pela imagem de um bancário virtual, cuja face surgirá na tela da máquina orientando-o verbalmente sobre o que fazer. Esse mesmo atendente virtual poderá estar presente no celular, verbalizando mensagens que hoje são enviadas por escrito, ou em *websites*, ajudando o usuário a fazer reservas de bilhetes aéreos, agendamento de consultas médicas e vários outros serviços realizados por meio de *menus* – uma tarefa aparentemente trivial para muitos, mas que ainda causa dor de cabeça em pessoas não familiarizadas com essas tecnologias ou com deficiência de leitura e limitações físicas, por exemplo.

Esse é o cenário projetado por uma dupla de pesquisadores da Universidade Estadual de Campinas (Unicamp) responsável pelo desenvolvimento de um método de animação facial videorealista, elaborado a partir de imagens do rosto humano – e não desenhadas com auxílio da computação gráfica –, que cria e reproduz em faces virtuais os detalhes



e as nuances observados em rostos reais. O objetivo da inovação, que recebeu o nome de AnimaFace2D, é tornar a interação com o computador e outros dispositivos eletrônicos parecidos com a que se estabelece durante uma conversação olho no olho. “O sistema que criamos poderá suportar o desenvolvimento de interfaces humano-computador mais intuitivas, eficientes e cativantes, representando uma alternativa a aplicativos mais tradicionais que se apoiam no uso de janelas, ícones, *menus*, teclado e *mouse*”, diz o engenheiro eletricista José Mario De Martino, corresponsável pela pesquisa e professor da Faculdade de Engenharia Elétrica e de Computação da Unicamp. “Conversar com um rosto humano é um processo eficiente e intuitivo. Acredito que pessoas virtuais são uma alternativa promissora para a criação de interfaces de comunicação em diferentes tipos de dispositivos e aplicações.”

O resultado final do processo de animação facial criado na Unicamp é um conjunto de imagens fotográficas de uma face real que, ao serem processadas e apresentadas em sequência, num ritmo apropriado, transmitem a sensação de movimento, como se fosse uma pessoa falando ao vivo. Para gerar uma imagem falante videorealista, a tecnologia recorre a uma base de 34 fotografias de diferentes posturas labiais, os chamados visemas, que podem, cada um deles, ser associados a um conjunto de fonemas de



nossa língua. Cada visema é, portanto, a representação visual do posicionamento articulatório da boca para a produção acústica dos fonemas. A identificação dos visemas foi resultado de pesquisa envolvendo a análise da movimentação articulatória durante a produção dos fonemas do português falado no Brasil realizada por De Martino.

Fotografias sincronizadas - A engenheira eletricista Paula Dornhofer Costa, aluna de doutorado sob orientação do professor De Martino, utilizou a definição dos visemas para identificar o conjunto de 34 fotografias da face de uma pessoa real utilizadas para síntese da animação. Essas 34 imagens possibilitam a sincronização entre os movimentos labiais da face virtual e a fala, gerando uma animação próxima à de um interlocutor falando ao vivo. Traçando um paralelo com o universo das dublagens de filmes, a correta identificação e o perfeito sequenciamento dos visemas evitam a falta de sincronia entre fala e articulação labial. De acordo com o professor da Unicamp, o *software*

desenvolvido é inédito no Brasil. Os sistemas similares que existem no exterior, ainda não comerciais, são dirigidos para outros idiomas.

“Todo o estudo relacionado ao desenvolvimento da tecnologia foi feito no Brasil e centrado em nossa realidade, sobretudo em nossa língua e suas particularidades”, afirma De Martino. “Não contamos com a parceria de pesquisadores estrangeiros. A única e importante colaboração que tivemos foi da fundação CPqD, de Campinas, que contribuiu com a bolsa de mestrado, apoiou a implantação do sistema piloto e permitiu o uso do sintetizador de texto ‘CPqD Texto Fala.’” O *software* utilizado nesse sistema piloto foi registrado no Instituto Nacional de Propriedade Industrial (INPI) e está pronto para uso comercial. “Temos forte interesse em estabelecer parcerias com empresas interessadas em explorar a tecnologia”, diz.

“A tecnologia será apresentada a especialistas de mercado em agosto deste ano, e acreditamos que ela vá interessar às empresas que trabalham com a interface homem-máquina, como o

de automação bancária, *e-commerce* e turismo, entre outros”, diz Giancarlo Stefanuto, coordenador científico do projeto Mobilização Empresarial para Inovação (Mepi) da Inova, a Agência de Inovação da Unicamp, que administra os direitos de propriedade intelectual e busca interessados em licenciar ou criar uma empresa para comercializar esse *software*. “Percebemos que há um mercado em expansão para *softwares* de animação de expressões faciais, como o desenvolvido na Unicamp. A interface homem-máquina está ficando cada vez mais personalizada e inovações como essa tendem a ter boa procura no futuro. Para nós, essa é uma linha de pesquisa estratégica, mas para a tecnologia entrar no mercado ainda são necessários alguns ajustes e adaptações. Ele precisa ser mais amigável, ter interfaces mais intuitivas, um tutorial e manual”, diz Stefanuto. “A expansão do mercado é fruto da percepção dos especialistas da Inova comprovada pelo próprio pesquisador em fóruns internacionais onde a tecnologia é debatida.”

Sistemas de animação facial poderão ter no futuro diversos tipos de aplicação, entre eles a criação de avatares, personagens ou atores virtuais em jogos ou vídeos comerciais, a elaboração de agentes personificados virtuais, tais como vendedores, tutores, atendentes de suporte, guias virtuais e apresentadores de notícias, entre outros. Poderão ser usados também como ferramentas de estudo e treinamento de leitura labial. A tecnologia tem características que possibilitam sua adaptação a várias plataformas de processamento, como telefones celulares, *smartphones* e *tablets*. “É possível imaginar, por exemplo, uma aplicação em que o usuário de celular assinaria um serviço de notícias que seria exibido por um apresentador sintetizado em seu próprio aparelho. As vantagens são que as notícias poderiam ser enviadas em formato de texto para o celular e o sistema se encarregaria de fazer a conversão automática de texto em fala visual. Além de baratear o serviço – o envio de caracteres tem menor custo que o de imagens –, não seria necessário gravar o vídeo de um apresentador de notícias todos os dias ou a cada boletim”, explica De Martino. ■

YURI VASCONCELOS