República Federativa do Brasil
Ministério do Desenvolvimento, Indústria e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0903935-0 A2**

(22) Data de Depósito: 13/05/2009
(43) Data da Publicação: 15/02/2011
(RPI 2093)

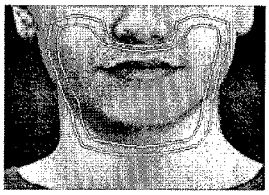
(51) Int.Cl.:
G06T 17/00

(54) Título: **SISTEMA DE SÍNTESE DE ANIMAÇÃO FACIAL POR COMPUTADOR BASEADA NA MANIPULAÇÃO DE IMAGENS**

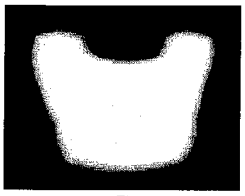
(73) Titular(es): Universidade Estadual de Campinas - Unicamp

(72) Inventor(es): José Mario de Martino, Paula Dornhofer Paro Costa

(57) Resumo: Sistema de síntese de animação facial por computador baseada na manipulação de imagens. A invenção refere-se a um sistema de síntese de animação facial por computador sincronizada com a fala e videorealista. Neste sistema a animação facial é gerada através da seleção, manipulação e apresentação de um conjunto reduzido de imagens, representando visemas dependentes do contexto fonético, que reproduz a movimentação articulatória visível na face durante a produção da fala, incluindo os efeitos da coarticulação. A abordagem adotada permite que seja possível gerar animações com alto grau de videorealismo, mesmo em plataformas com reduzida capacidade de armazenamento, como os dispositivos móveis celulares. Entende-se como animação videorealista uma animação que pode ser confundida com um vídeo real.

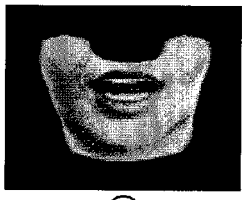


1



2

3



4

PI0903935-0

RELATÓRIO DESCRITIVO

Sistema de síntese de animação facial por computador baseada na manipulação de imagens.

CAMPO DA INVENÇÃO

A presente invenção refere-se a sistema de animação facial sincronizada com a fala baseada na manipulação de imagens fotográficas, o qual reproduz de forma videorealista a movimentação articulatória visível na face do locutor. A invenção gera animações que podem ser confundidas com o vídeo de um pessoa real e pode ser utilizada em diversas aplicações, entre elas máquinas de atendimento automático, sistemas de reserva de vôos ou de consultas médicas e em plataformas com capacidade de armazenamento limitada, como telefones celulares e computadores de bolso.

FUNDAMENTO DA INVENÇÃO

A face ocupa um papel de destaque na comunicação humana. Já nos primeiros momentos da infância, aprendemos a reconhecer e a identificar indivíduos por suas faces. Ao longo da vida, estimulada por experiências sociais, esta habilidade é aperfeiçoada, permitindo a interpretação e a identificação de estados emocionais a partir da observação da face. Várias vezes ao dia somos expostos a interações face-a-face, sendo que nestas ocasiões, não raras as vezes, a informação visual apresentada na face complementa e auxilia a compreensão da mensagem transportada pelo sinal acústico da fala.

Criar e reproduzir em faces virtuais os detalhes e nuances observados em faces reais é um dos objetivos da animação facial gerada por computador. Conferir a uma face virtual a aparência, a movimentação e o comportamento de uma face real permite reproduzir e explorar na interação humano-computador o estilo de comunicação ao face-a-face com o qual todos nós estamos habituados.

A realização do interfaceamento humano-computador através de uma personagem virtual torna a interação mais interessante, adicionando valor à aplicação. Além disso, a interação com uma personagem virtual é mais intuitiva e não exige treinamento especial, reduzindo a barreira de acesso para os indivíduos com pouca familiaridade com os mecanismos das interfaces tipo WIMP (*Windows, Icons, Menus e Pointing device*). Dependendo da aplicação, estas personagens virtuais podem desempenhar papéis variados, tais como instrutores, assistentes, avatares, apresentadores, atendentes ou vendedores. Em particular, a utilização da tecnologia embarcada em dispositivos móveis é uma opção promissora para soluções de videotelefonia móvel devido à potencial redução da taxa de transmissão. Neste cenário, em vez da transmissão do vídeo do interlocutor, bastaria a transmissão dos poucos parâmetros necessários para especificar a locução e as expressões faciais associadas. Em videotelefonia baseado em animação facial, além da aparência da face realista, a qualidade da reprodução da movimentação facial é fundamental. Também para aplicações que necessitam garantir a percepção de profissionalismo como, por exemplo, em máquinas automáticas de atendimento bancário é necessária uma aparência e movimentação realista para transmitir confiança. Para estas aplicações, animações caricatas ou com estilo de desenho animado não são apropriadas.

O realismo de uma animação por computador é expresso por sua capacidade de ser confundida com o vídeo de uma situação real. Neste contexto, é possível identificar duas componentes principais do realismo de uma animação: o fotorealismo e o videorealismo. O fotorealismo indica que os objetos da cena podem ser confundidos com objetos reais. O videorealismo indica que a animação, além de apresentar objetos de forma fotorealista, também apresenta movimentação que pode ser confundida com a realidade. Desta forma, uma animação facial sincronizada com a fala videorealista, além da reprodução fotográfica da aparência da face, também reproduz de forma natural os movimentos articulatórios da fala em estreita sincronia e harmonia com a locução.

No estado da técnica é possível identificar duas abordagens distintas para a animação facial por computador: animação facial 3D (tridimensional) e animação facial 2D (bidimensional).

A animação facial 3D utiliza e manipula um modelo geométrico tridimensional, em geral expresso por uma malha de polígonos, para representar a aparência e movimentação da face. Em geral, as animações geradas com esta técnica apresentam aparência artificial que claramente as distinguem de um vídeo de uma situação real. A modelagem tridimensional, para representar a riqueza de detalhes da face humana e permitir deformações realistas em sua superfície durante o processo de animação, é tarefa complexa, que demanda sofisticado e dispendioso processo de modelagem e animação, realizados por meticuloso trabalho de animadores experientes e habilidosos, não raras às vezes, auxiliados por caros equipamentos de captura de movimento.

Já a animação facial 2D, para a produção de videorealismo, utiliza imagens fotográficas de uma pessoa real como substrato para a geração da animação. A característica inerentemente realista das fotografias empresta às animações 2D um alto grau de fotorealismo, apenas aproximado com grande esforços pela animação 3D. A movimentação realista, entretanto, é um aspecto que necessita atenção cuidadosa e tratamento elaborado.

Na animação facial, além da aparência da face, a qualidade da reprodução da movimentação facial é fundamental. Na animação facial sincronizada com a fala, o principal objetivo perseguido é a reprodução realista da movimentação articulatória visível em estreita consonância com o sinal acústico da fala.

Para a geração de uma animação facial em sincronia e em harmonia com a locução, é imperativa a reprodução dos movimentos articulatórios associados à realização dos vários segmentos sonoros da língua. Para tanto, além da identificação das posturas características dos gestos articulatórios associados aos segmentos sonoros, faz-se necessária a

representação das transições entre estas posturas considerando os efeitos da coarticulação.

Os efeitos da coarticulação se manifestam pela alteração do padrão articulatório de um determinado segmento sonoro pela influência da articulação de outro adjacente ou, e em menor grau, próximo na cadeia da produção sonora. Os efeitos da coarticulação fazem com que, por exemplo, o "p" da palavra "para" seja visualmente distinto do "p" da palavra "pura". Neste último caso, o movimento articulatório necessário à produção do "u" influencia de maneira significativa os aspectos visíveis da articulação do "p". Dependendo de sua característica, a coarticulação pode ser classificada como perseveratória ou antecipatória. Na coarticulação perseveratória, a articulação de um segmento de fala é influenciada pela movimentação articulatória de um segmento que o antecede na cadeia fonética que compõe a locução. Na coarticulação antecipatória, a movimentação articulatória do segmento é influenciada pela articulação de um segmento que o sucede.

No estado da técnica é possível identificar três abordagens principais para sistemas de animação facial 2D sincronizada com a fala: 1) Metamorfose de visemas; 2) Síntese concatenativa; 3) Trajetória de visemas em espaço multidimensional.

A abordagem de "Metamorfose de Visemas" é inspirada fortemente na estratégia de interpolação entre quadro-chaves (*key-framing*) utilizada na animação tradicional. Visemas são as representações visuais das posturas articulatórias visíveis e distinguíveis dos vários fonemas de uma língua. Na "Metamorfose de Visemas", cada fonema da língua é representado por uma única imagem que caracteriza um visema distinto. A partir da transcrição temporizada da sequência de segmento sonoros de uma locução, o processo de síntese da animação da abordagem "Metamorfose de Visemas" associa visemas aos instantes de produção dos fonemas, definindo quadros-chaves. . Os quadros intermediários entre os quadros-chave são gerados com a técnica de metamorfose entre imagens. As primeiras propostas pertencentes à abordagem "Metamorfose de visemas", como as descritas no artigo *Synthesis*

of speaker facial movement to match selected speech sequences de K. C. Scott e outros, publicado nos *Proceedings of the Fifth Australian Conference on Speech Science and Technology*, 1994, páginas 620–625, no documento de patente norte-americano US6250928, no artigo *Miketalk: A Talking Facial Display Based on Morphing* de Tony Ezzat e Tomaso Poggio, publicado nos *Proceedings of the Computer Animation Conference 1998*, páginas 96-102, no artigo *Text-to-audiovisual speech synthesizer* de U. K. Goyal, A. Kappor e P. Kalra, publicado nos *Proceedings of the Second International Conference on Virtual* Worlds, 2000, páginas 256-269 e no artigo *Audio driven facial animation for audio-visual reality* de T. A. Faruquie e outros, publicado na *IEEE International Conference on Multimedia and Expo, 2001*, páginas 821-824, não tratam os efeitos da coarticulação, produzindo movimentação caricata e, portanto, com baixo videorealismo.

Ainda pertencente à abordagem "Metamorfose de Visemas", a proposta descrita no artigo *Image-based talking heads using radial basis functions* de J. D. Edge e S. Maddock, publicado nos *Proceedings of the Theory and Practice of Computer Graphics*, 2003, páginas 74-80, busca melhorar as soluções anteriores procurando acomodar os efeitos da coarticulação utilizando o modelo teórico de coarticulação proposto por M. M Cohen e D. W. Massaro descrito no artigo *Modeling Coarticulation in Synthetic Visual Speech*, publicado no livro coletânea *Models and Techniques in Computer Animation*, Springer-Verlag, 1993, páginas 139-156. Nesta abordagem, inicialmente são definidos visemas puros, sem levar em conta os efeitos da coarticulação. No processo de síntese da animação, as representações dos visemas puros são combinadas com o auxílio de funções de influências que levam em conta os efeitos da coarticulação. O grande problema desta solução é a caracterização dos visemas puros e das funções de influência através dos vários parâmetros que as definem. A determinação dos parâmetros das funções e a definição dos visemas puros são questões complexas que, na prática, acabam por dificultar, se não inviabilizar, a aplicação do modelo.

Para contornar estes problemas, a abordagem "Síntese Concatenativa" procura tratar os efeitos da coarticulação de forma mais direta e efetiva. Enquanto a abordagem baseada em metamorfose entre visemas tem suas origens na tradicional estratégia de animação por interpolação entre quadros-chave, a abordagem da "Síntese Concatenativa" inspira-se no paradigma utilizado nos sistemas de síntese concatenativa de áudio, que geram fala sintética através da concatenação de fragmentos de fala natural. No caso da "Síntese Concatenativa" são utilizados fragmentos de vídeo. Como resultado a abordagem acaba por necessitar de extensa base de dados composta de várias sequências de vídeo, tornando inviável a sua utilização em computadores de bolso ou dispositivos móveis celulares com baixa capacidade de processamento e de armazenamento. Os artigos *Triphone based unit selection for concatenative visual speech synthesis* de F. J. Huang, E. Cosatto e H. P. Graf, publicado nos anais da *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002. v. 2, páginas, 2037–2040, *Reusing motion data to animate visual speech* de J. Edge, M. Sanches e S. Maddock, publicado nos anais do *Symposium on Language, Speech and Gesture for Expressive Characters*, 2004, páginas 66–74 e *Visyllable based speech animation* de S. Kshirsagar e N. Magnenat-Thalmann, publicado na revista *Computer Graphics Forum*, v. 22, n. 3, 2003, página. 632–640 são exemplos de propostas desta abordagem, assim como os documentos de patente norte-americanos US6112177, US7209882 e US7388586.

O surgimento de sistemas computacionais com capacidade de processamento e armazenamento de dados cada vez maiores viabilizou a implementação de sistemas de animação facial 2D baseados em elaborados e complexos algoritmos de análise e processamento das imagens e fragmentos de vídeo de corpus audiovisual cada vez mais abrangentes e volumosos. Uma das características compartilhadas por sistemas deste tipo é que a totalidade dos milhares de quadros resultantes da captura do corpus audiovisual é analisada e processada durante o processo de construção da base de imagens. O resultado deste processo é a constituição de uma base de imagens organizada

em um espaço multidimensional, no qual as imagens são indexadas através de parâmetros específicos relacionados às suas características visuais. Nestes sistemas, a síntese da animação consiste na definição de uma trajetória dentro deste espaço multidimensional de visemas.

A abordagem "Trajetória de Visemas em Espaço Multidimensional", além da volumosa base de dados que impossibilita a sua aplicação em plataformas com capacidade de armazenamento reduzida, utiliza sofisticado sistema de busca e seleção da imagens, aumentando ainda mais o custo computacional de sua implementação. De maneira bastante ilustrativa, os autores T. Ezzat, G. Geiger e T. Poggio do artigo *Trainable videorealistic speech animation*, publicado nos *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques – SIGGRAPH 2002*, páginas 388-398, afirmam que enquanto o processo de captura do corpus audiovisual de seu sistema dura apenas 15 minutos, a execução dos algoritmos de processamento automático das imagens capturadas leva vários dias até ser concluída. Exemplos de propostas que utilizam esta abordagem são descritos nos artigos *Photo-realistic talking-heads from image samples* de E. Cosatto e H. P. Graf, publicado na revista *IEEE Transactions on Multimedia*, v. 2, n. 3, 2000, páginas 152–163, 2000, *Trainable videorealistic speech animation* de T. Ezzat, G. Geiger e T. Poggio, publicado nos *Proceedings of the 29th annual conference on Computer graphics and interactive techniques - SIGGRAPH* 2002. p. 388–398, *Face analysis for the synthesis of photo-realistic talking heads* de H. P. Graf, E. Cosatto e T. Ezzat, publicado nos *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, páginas 189–194, assim como nos documentos de patente norte-americanos US6735566, US6366885, US6504546 US6654018 e US7168953.

BREVE DESCRIÇÃO DA INVENÇÃO

A presente invenção gera animação facial sincronizada com fala a partir da descrição fonética temporizada de uma locução. A animação facial é gerada através da seleção, manipulação e apresentação de um conjunto

reduzido de imagens, que representam visemas dependentes do contexto fonético de sua produção. A utilização de um conjunto de imagens que, embora reduzido, permite representar os efeitos da coarticulação, supera as problemas das soluções do atual estado da técnica ao permitir que seja possível gerar animações com alto grau de videorealismo mesmo em plataformas com reduzida capacidade de armazenamento, como os dispositivos móveis celulares. Adicionalmente, a presente invenção também aperfeiçoa a estratégia de "Metamorfose de Visemas", ao adaptar a técnica de metarmorfose para contemplar os efeitos de aceleração e desaceleração entre visemas, reproduzindo de maneira mais realista a movimentação articulatória visível.

A atividade inventiva associada à tecnologia pode ser parcialmente caracterizada pela adaptação e extensão do conceito de visemas dependentes de contexto fonético, inicialmente estabelecido para animação 3D na tese de doutorado "Animação Facial Sincronizada com a Fala: Visemas Dependentes do Contexto Fonético para o Português do Brasil" de J. M. De Martino, defendida em julho de 2005 na Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas. É importante observar que este trabalho trata exclusivamente de animação facial 3D, sendo que o conceito de visemas dependentes de contexto fonético é puramente geométrico e estava restrito a contexto de até 3 segmentos. Na invenção em questão, os visemas dependentes de contexto fonético são representados por imagens fotográficas, considerando contextos fonéticos com até 7 segmentos. Adicionalmente a invenção utiliza uma adaptação da técnica de metamorfose entre imagens para contemplar efeitos de aceleração e desaceleração entre visemas, característicos da movimentação articulatória visível. Estes avanços no estado da técnica resultaram, além da aparência e movimentação mais realista, em um expressivo e surpreendente ganho de inteligibilidade de fala aproximadamente 6 vezes maior em relação à animação 3D utilizando visemas dependentes do contexto fonético, em situações de forte contaminação do áudio por ruído.

Em relação ao estado da técnica da animação facial 2D, a tecnologia em questão apresenta a vantagem de utilizar um conjunto reduzido de imagens que contempla os efeitos da coarticulação antecipatória e perseveratória. Este conjunto reduzido de imagens permite a geração da animação da movimentação articulatória visível de qualquer locução e viabiliza a realização da tecnologia também em dispositivos com capacidade limitada de processamento e armazenamento como os telefones celulares e computadores de bolso. Soluções videorealistas do estado da técnica adotam a concatenação de segmentos de vídeo e, portanto com um custo significativamente maior de armazenamento e processamento. Adicionalmente, o conjunto reduzido de imagens da tecnologia permite uma fácil personalização das animações. As poucas imagens podem ser facilmente capturadas em uma curta sessão de fotos, que, aliás, pode ser caseira, permitindo facilmente a geração de animações personalizadas de acordo com a preferência do usuário, incluindo a própria imagem, a de um modelo fotográfico, a do garoto ou garota propaganda da empresa, etc..

A animação facial videorealista propiciada pela presente inovação indica a sua aplicação em situações em que seja necessário transparecer seriedade e confiabilidade, como por exemplo, em máquinas de atendimento bancário, sistemas de reserva de vôo ou de consulta médica. Para estas aplicações, não são apropriadas animações pouco realistas, caricatas ou com estilo de desenho animado, já que estas podem facilmente ser associadas a entretenimento, diversão, brincadeira ou ainda a falta de profissionalismo.

BREVE DESCRIÇÃO DAS FIGURAS

A Figura 1 apresenta um diagrama de blocos que ilustra a arquitetura do sistema de animação facial que caracteriza a presente invenção: O bloco (1) representa a etapa de Geração da Fala e Informação Associada; o bloco (2) representa a etapa de Animação Facial 2D; o bloco (3) representa a Base de Imagens utilizada pelo sistema de animação facial; o bloco (4) representa a etapa de sincronização dos quadros da animação com o áudio; o

bloco (5) representa a apresentação da animação em um dispositivo de saída; a seta (6) representa o fluxo do sinal de áudio de uma locução gerado pelo bloco (1); a seta (7) a transcrição fonética temporizada desta locução gerada no bloco (1); a seta (8) representa as fluxo das imagens de referência extraídas da Base de Imagens (3) que serão utilizadas no processo da síntese da animação realizado em (2); a seta (9) representa o fluxo das imagens da animação gerada em (2); a seta (10) representa o fluxo da informação que constitui uma animação sincronizada com a fala (vídeo com áudio) que será apresenta em (5).

A Figura 2 apresenta o conjunto de imagens que constituem a base de imagens de uma realização da presente invenção. Cada imagem representa um visema dependente do contexto fonético de sua produção. As imagens de (1) a (34) estão associadas respectivamentes aos seguintes visemas dependentes do contexto fonético: silêncio (ou repouso) – representado pelo símbolo #, <p_1>, <p_2>, <f_1>, <f_2>, <t_1>, <t_2>, <s_1>, <s_2>, <l_1>, <l_2>, <l_3>, <l_4>, <□$_1$>, <□$_2$>, <□$_1$>, <□$_2$>, <□$_3$>, <k_1>, <k_2>, <k_3>, <□$_1$>, <□$_2$>, <i_1>, <i_2>, <a>, <u>, <□>, <□>, <□>, <e>, <□>, <□>, <o>.

A Figura 3 apresenta os pontos de referência utilizados para o registro das imagens em uma realização da presente invenção.

A Figura 4 ilustra o processo de criação e aplicação da máscara de transparência: a imagem (1) apresenta os contornos que definem a máscara de transparência; a imagem (2) apresenta um exemplo de máscara de transparência; a imagem (3) apresenta uma imagem da qual se deseja extrair a região de interesse; na imagem (4) é apresentado o resultado da aplicação da máscara (2) à imagem (3).

A Figura 5 apresenta os pontos-âncoras utilizados em uma realização da presente invenção na metamorfose de imagens, sendo que:

- P1 (posição lábio superior): ponto pertencente ao contorno externo do lábio superior, com coordenada "x" igual a do ponto médio do segmento determinado pelos pontos P3 e P4;
- P2 (posição lábio inferior): ponto pertencente ao contorno externo do lábio inferior, com coordenada "x" igual a do ponto médio do segmento determinado pelos pontos P3 e P4;
- P3 (canto direito da boca): ponto que caracteriza o canto direito da boca;
- P4 (canto esquerdo da boca): ponto que caracteriza o canto esquerdo da boca;
- P5 (posição queixo): ponto extremo inferior (coordenada "y" máxima) do contorno do queixo.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Os exemplos aqui descritos têm somente o intuito de exemplificar uma das inúmeras possíveis formas de realizar a invenção, de forma que este relatório descritivo deve ser encarado de forma ilustrativa, e não restritiva.

A invenção refere-se a um sistema de síntese de animação facial por computador sincronizada com a fala e videorealista. Neste sistema a animação facial é gerada através da seleção, manipulação e apresentação de um conjunto reduzido de imagens, que reproduz a movimentação articulatória visível na face durante a produção da fala, incluindo os efeitos da coarticulação. A abordagem adotada permite que seja possível gerar animações com alto grau de videorealismo, mesmo em plataformas com reduzida capacidade de armazenamento, como os dispositivos móveis celulares. Entende-se como animação videorealista, uma animação que pode ser confundida com um vídeo real.

A arquitetura da presente invenção é apresentada na Figura 1.

A etapa Geração da Fala e Informação Associada (1) engloba a geração do áudio (6) de uma locução e sua respectiva transcrição fonética temporizada (7). A locução pode ser natural, produzida por um locutor real, ou

mesmo sintética, gerada por um conversor texto-fala. A transcrição fonética temporizada é composta pela sequência de segmentos sonoros, ou fones, que compõem a locução e suas respectivas durações. A transcrição fonética temporizada pode ser gerada como subproduto de processo de conversão texto-fala ou utilizando sistema de segmentação fonética automática. Alternativamente, a segmentação fonética pode ser efetuada manualmente.

A etapa Animação Facial 2D (2) utiliza a sequência de fones temporizados para controlar a animação facial e gerar uma sequência de imagens (9) que reproduzem na face virtual a movimentação articulatória visível. A sequência de imagens, que constitui a animação, é produzida pela seleção e manipulação das imagens de referência (8) armazenadas na Base de Imagens (3).

O bloco Base de Imagens (3) representa, em uma realização do sistema, uma base de dados constituída de 34 imagens fotográficas correspondentes a posturas articulatórias dependentes do contexto fonético, representantes dos diversos fonemas do português do Brasil, que são utilizadas para a geração da animação. As imagens estão devidamente rotuladas e caracterizadas por parâmetros para auxiliar a busca, seleção e manipulação durante a síntese da animação.

A etapa Sincronização (4) inclui os mecanismos de controle da apresentação do áudio e das imagens da animação de forma a garantir o sincronismo entre o áudio e a informação visual. Animação facial sincronizada com a fala (10) produzida ao final de sincronização pode ser apresentada em dispositivo de saída (5), como um monitor de vídeo, ou ainda ser armazenada para futuras apresentações.

As contribuições inovadoras da presente invenção estão associadas à Base de Imagens (3) e ao processo da Animação Facial 2D (2), cujas descrições detalhadas se seguem.

BASE DE IMAGENS

A criação da Base de Imagens (3) é realizada a partir do processameno de um corpus audiovisual constituído de segmentos de vídeo da

face de um locutor real produzindo um conjunto de palavras sem sentido, denominados logatomas. Este conjunto de logatomas abrange representantes de todos os fonemas do português do Brasil. Do corpus audiovisual foram extraídas, em uma realização do invento, 34 imagens para formar a Base de Imagens (3). As imagens foram selecionadas levando-se em conta os grupos homofema identificados por De Martino na tese "Animação Facial Sincronizada com a Fala: Visemas Dependentes do Contexto Fonético para o Português do Brasil" defendida em 2005 na Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas. Um grupo homofema agrupa segmentos sonoros que não são possíveis de serem distinguidos visualmente. Cada imagem da Base de Imagens representa uma postura labial característica associada a um grupo homofema considerando os efeitos da coarticulação do contexto fonético de sua produção. Às posturas labiais características são dadas a denominação de visemas. Desta forma, os visemas adotados são dependentes do contexto fonético de sua produção.

Nas tabelas 1 e 2 são mostrados os grupos homofemas, o fone representante do grupo homofema, os respectivos visemas e contextos fonéticos associados, tomados como referência para formação da Base de Imagens em uma realização da invenção. Nas tabelas os segmentos sonoros, ou fones, estão representados com os símbolos do Alfabeto Fonético Internacional, ou símbolos IPA (IPA – *International Phonetic Alphabet*). A Figura 2 apresenta a região em torno da boca das imagens selecionadas para a Base de Imagens. Observa-se que, para este exemplo de realização da invenção, foram selecionadas 22 imagens correspondentes a visemas consonantais (Tabela 1), 11 imagens correspondentes a visemas vocálicos (Tabela 2) e uma imagem representante da postura labial de repouso (ou silêncio), totalizando 34 imagens. As images (1) a (34) da Figura 2 representam respectivamente, os seguintes visemas dependentes do contexto fonético: silêncio (ou repouso) – representado pelo símbolo #, $<p_1>$, $<p_2>$, $<f_1>$, $<f_2>$, $<t_1>$, $<t_2>$, $<s_1>$, $<s_2>$, $<l_1>$, $<l_2>$, $<l_3>$, $<l_4>$, $<\square_1>$, $<\square_2>$, $<\square_1>$, $<\square_2>$, $<\square_3>$,

<k_1>, <k_2>, <k_3>, <□$_1$>, <□$_2$>, <i_1>, <i_2>, <a>, <u>, <□>, <□>, <□>, <e>, <□>, <□>, <o>.

TABELA 1: Visemas consonantais dependentes de contexto.

Grupo Homofema	Fone Representante	Visemas	Contextos Fonéticos
[p, b, m]	[p]	<p_1>	[pi] [p□] [pa] [p□] [ipi] [ip□] [□pi] [□p□] [ipa] [ip□] [□pa] [□p□] [ipu] [ip□] [□pu] [□p□] [api] [ap□] [□pi] [□p□] [apa] [ap□] [□pa] [□p□] [apu] [ap□] [□pu] [□p□] [upa] [up□] [□pa] [□p□]
		<p_2>	[pu] [p□] [upi] [up□] [□pi] [□p□] [upu] [up□] [□pu] [□p□]
[f, v]	[f]	<f_1>	[fi] [f□] [fa] [f□] [ifi] [if□] [□fi] [□f□] [ifa] [if□] [□fa] [□f□] [ifu] [if□] [□fu] [□f□] [afi] [af□] [□fi] [□f□] [afa] [af□] [□fa] [□f□]
		<f_2>	[fu] [f□] [afu] [af□][□fu] [□f□] [ufi] [uf□] [□fi] [□f□] [ufa] [uf□] [□fa] [□f□] [ufu] [uf□] [□fu] [□f□]
[t, d, n]	[t]	<t_1>	[ti] [t□] [tu] [t□] [iti] [it□] [□ti] [□t□] [ita] [it□] [□ta] [□t□] [itu] [it□] [□tu] [□t□] [ati] [at□] [□ti] [□t□] [atu] [at□] [□tu] [□t□] [uti] [ut□] [□ti] [□t□] [uta] [ut□] [□ta] [□t□] [utu] [ut□] [□tu] [□t□]
		<t_2>	[ta] [t□] [ata] [at□] [□ta] [□t□]
[s, z]	[s]	<s_1>	[si] [s□] [sa] [s□] [isi] [is□] [□si] [□s□] [isa] [is□] [□sa] [□s□] [asi] [as□] [□si] [□s□] [asa] [as□] [□sa] [□s□]

		<s_2>	[su] [s□] [isu] [is□] [□su] [□s□] [asu] [as□] [□su] [□s□] [usi] [us□] [□si] [□s□] [usa] [us□] [□sa] [□s□] [usu] [us□] [□su] [□s□]
[l]	[l]	<l_1>	[li] [l□] [ili] [il□] [□li] [□l□] [alu] [al□] [□lu] [□l□] [uli] [ul□] [□li] [□l□] [ula] [ul□] [□la] [□l□]
		<l_2>	[la] [l□] [ila] [il□] [□la] [□l□] [ali] [al□] [□li] [□l□] [ala] [al□] [□la] [□l□]
		<l_3>	[lu] [l□]
		<l_4>	[ilu] [il□] [□lu] [□l□] [ulu] [ul□] [□lu] [□l□]
[□, □]	[□]	<$□_1$>	[□i] [□□] [□a] [□□] [i□i] [i□□] [□□i] [□□□] [i□a] [i□□] [□□a] [□□□] [i□u] [i□□] [□□u] [□□□] [a□i] [a□□] [□□i] [□□□] [a□a] [a□□] [□□a] [□□□] [a□u] [a□□] [□□u] [□□□] [u□i] [u□□] [□□i] [□□□] [u□a] [u□□] [□□a] [□□□]
		<$□_2$>	[□u] [□□] [u□u] [u□□] [□□u] [□□□]
[□, □]	[□]	<$□_1$>	[□i] [□□] [□a] [□□] [i□i] [i□□] [□□i] [□□□] [i□a] [i□□] [□□a] [□□□]] [a□i] [a□□]] [□□i]] [□□□] [a□a] [a□□] [□□a] [□□□]
		<$□_2$>	[□u] [□□] [u□i] [u□□] [□□i] [□□□] [u□a] [u□□] [□□a] [□□□]
		<$□_3$>	[i□u] [i□□] [□□u] [□□□] [a□u] [a□□] [□□u] [□□□] [u□u] [u□□] [□□u] [□□□]
[k, g]	[k]	<k_1>	[ki] [k□] [iki] [ik□] [□ki] [□k□] [ika] [ik□] [□ka] [□k□] [aki] [ak□] [□ki] [□k□] [uki] [uk□] [□ki] [□k□] [uka] [uk□] [□ka] [□k□]
		<k_2>	[ka] [k□] [aka] [ak□] [□ka] [□k□]
		<k_3>	[ku] [k□] [iku] [ik□] [□ku] [□k□] [aku] [ak□] [□ku] [□k□] [uku] [uk□] [□ku] [□k□]

[□], [□]	[□], [□]	<□$_1$>	[□i] [□a] [i□i] [i□□] [□□i] [□□□] [i□i] [i□□] [□□i] [□□□] [i□a] [i□□] [□□a] [□□□] [i□a] [i□□] [□□a] [□□□] [a□i] [a□□] [□□i] [□□□] [a□i] [a□□] [□□i] [□□□] [a□a] [a□□] [□□a] [□□□] [a□a] [a□□] [□□a] [□□□] [u□a] [u□□] [□□a] [□□□] [u□a] [u□□] [□□a] [□□□]
		<□$_2$>	[□□] [□u] [i□u] [i□□] [□□u] [□□□] [i□u] [i□□] [□□u] [□□□] [a□u] [a□□] [□□u] [□□□] [a□u] [a□□] [□□u] [□□□] [u□i] [u□□] [□□i] [□□□] [u□i] [u□□] [□□i] [□□□] [u□u] [u□□] [□□u] [□□□] [u□u] [u□□] [□□u] [□□□]

TABELA 2: Visemas vocálicos dependentes de contexto.

Grupo Homofema	Fone Representante	Visemas	Contextos Fonéticos
[i, ĩ]	[i]	<i$_1$>	Todos os contextos exceto [tit] e [□i□]
		<i$_2$>	[tit] e [□i□]
[e, ẽ]	[e]	<e>	Todos os contextos
[□]	[□]	<□>	Todos os contextos
[a, □]	[a]	<a>	Todos os contextos
[□]	[□]	<□>	Todos os contextos
[o, õ]	[o]	<o>	Todos os contextos
[u, ũ]	[u]	<u>	Todos os contextos
[□]	[□]	<□>	Todos os contextos
[□]	[□]	<□>	Todos os contextos
[□]	[□]	<□>	Todos os contextos

As imagens selecionadas para a Base de Imagens são registradas em relação a uma imagem base. O registro tem por função corrigir variações de enquadramento da face, fazendo com que os elementos da face, tais como olhos, nariz e lábios, estejam em posições uniformes no conjunto de imagens analisado. A imagem base, ou face-base, apresenta a face completa do locutor/locutora. Na imagem da face-base, assim como nas imagens a serem registradas, em uma realização da invenção, são marcados pontos de referência para o registro, definidos por pontos de elementos faciais que não se movimentam duram a produção da fala. A Figura 3 apresenta uma escolha possível para os pontos de referência.

Em uma realização da invenção, o processo de registro é realizado através de uma transformação espacial (*warping*) que deforma a imagem a ser registrada para que seus pontos de referência passem a ter as mesmas coordenadas que os pontos de referência marcados na imagem da face-base. Nesta realização foi utilizada uma transformação de *warping* baseada em funções de base radial (RBF - *Radial Basis Function*). O mesmo tipo de *warping* foi adotado para a deformação no processo de metamorfose entre visemas, sendo que os detalhes de sua formulação são apresentados mais adiante na seção Animação Facial 2D. Após a etapa de registro, todas as imagens selecionadas do corpus audiovisual estão alinhadas com a imagem da face-base.

Durante a síntese da animação, a região da boca e do queixo de uma imagem visema é combinada com a face-base. Esta operação é realizada através da criação de uma máscara que extrai da imagem visema a região de interesse constituída da região da boca e queixo, permitindo que essa região seja posteriormente fundida à face-base. Para tanto utiliza-se uma máscara de transparência. A máscara de transparência é uma imagem do tipo *bitmap* do mesmo tamanho que as imagens dos visemas. Cada ponto da imagem, ou *pixel*, possui um valor α entre "0" (*pixel* preto – totalmente transparente) e "1" (*pixel* branco – totalmente opaco) que indica a transparência da imagem do visema. A máscara de transparência é criada através da definição de dois

contornos. Um contorno interno *CI* é desenhado ao redor da área da parte facial que deve ficar visível (região da boca e queixo). Um contorno externo *CE* define a região que não será visível. A Figura 4(1) apresenta os contornos que definem a máscara de transparência. Externamente ao contorno *CE* todos os pixels da máscara são "0" (totalmente transparentes) e na região interna limitada pelo contorno *CI* todos os pixels são "1" (totalmente opacos). Na região entre os dois contornos, a máscara é parcialmente transparente, com valores de pixel entre "0" e "1" em função da distância do pixel aos contornos externo e interno.

Considerando:

- *p* um ponto na imagem (um *pixel*);
- α_p o valor da opacidade do ponto *p* na máscara de transparência;
- *A* o conjunto de pontos internos ao contorno *CI*;
- *B* o conjunto de pontos internos ao contorno *CE*;
- *dist(C, p)* a menor distância entre o ponto *p* e o contorno *C*.

tem-se que:

$$\alpha_p = \begin{cases} 1 & se\ p \in A \\ 0 & se\ p \notin B \\ \dfrac{dist(CE,p)}{dist(CE,p)+dist(CI,p)} & se\ p \in B\ e\ p \notin A \end{cases}$$

Na Figura 4(2) tem-se um exemplo de máscara de transparência.

A aplicação da máscara é expressa pela interpolação linear entre a imagem da face-base e a imagem do visema expressa por:

$$f_p = \alpha_p\, b_p + (1-\alpha_p)\, v_p$$

onde:

- b_p é o valor da intensidade de cada canal de cor do ponto *p* da imagem da face-base;
- v_p é o valor da intensidade de cada canal de cor do ponto *p* da imagem do visema extraída da "Base de Imagens";

- f_p é a valor da intensidade de cada canal de cor do ponto p da imagem final.

A Figura 4(3) apresenta uma imagem da qual se deseja extrair a região de interesse, e em 4(4) é apresentado o resultado da aplicação da máscara 4(2) à imagem 4(3).

Uma vez que apenas a região da boca e queixo das imagens da base de imagens é utilizada durante a síntese, não existe a necessidade de se armazenar a imagens dos visemas com suas dimensões originais. Assim sendo, são apenas armazenadas imagens contendo apenas a região da boca e queixo, recortadas das imagens originais dos visemas. Estas imagens possuem dimensões menores que as originais reduzindo o espaço necessário de armazenamento da base de imagens e a quantidade de *pixels* a serem processados no momento da síntese. A Figura 2 apresenta as imagens armazenadas na Base de Imagens, em uma realização da invenção e os visemas que elas representam.

Para cada imagem da Base de Imagens estão associados os seguintes campos:

- Identificador do grupo homofema representado pela imagem.
- Coordenadas em *pixels* dos pontos P1, P2, P3, P4 e P5. Os pontos P1, P2, P3, P4 e P5 são apresentados na Figura 5.

Os pontos P1, P2, P3, P4 e P5 definem pontos-âncora que são utilizados durante a etapa de animação e correspondem a:

- P1 (posição lábio superior): ponto pertencente ao contorno externo do lábio superior, com coordenada "x" igual a do ponto médio do segmento determinado pelos pontos P3 e P4;
- P2 (posição lábio inferior): ponto pertencente ao contorno externo do lábio inferior, com coordenada "x" igual a do ponto médio do segmento determinado pelos pontos P3 e P4;
- P3 (canto direito da boca): ponto que caracteriza o canto direito da boca;

- P4 (canto esquerdo da boca): ponto que caracteriza o canto esquerdo da boca;
- P5 (posição queixo): ponto extremo inferior (coordenada "y" máxima) do contorno do queixo.

ANIMAÇÃO FACIAL 2D

A partir da transcrição fonética temporizada gerada pela etapa Geração da Fala e Informação Associada – Figura 1(1), a primeira etapa do processo de síntese da animação facial consiste na conversão fone-homofema. Para esta conversão é utilizada, para uma realização da invenção, a coluna 1 da Tabela 1, para os segmentos consonantais, e a coluna 1 da Tabela 2, para os segmentos vocálicos. Após esta etapa é efetuada o mapeamento dos fones em visemas considerando o contexto fonético de sua produção. Para tanto é efetuada uma análise dos fones vizinhos ao fone que está sendo mapeado.

Da Tabela 2 observa-se que a grande maioria dos visemas vocálicos são independentes de contexto e, portanto, na ocorrência destes fones não é necessária a análise dos fones vizinhos, exceto para os visemas $<i_1>$ e $<i_2>$.

Já para os visemas consonantais, listados na Tabela 1, observa-se que os contextos fonéticos presentes na tabela são sequências de dois ou três fones (trifones) no formato CV ou VCV, onde V representa uma vogal e C uma consoante (vide coluna "Contextos Fonéticos'"). Assim sendo, para outros tipos de contextos é necessário realizar um mapeamento para os contextos desta tabela.

A Tabela 3 apresenta o mapeamento utilizado para uma realização da invenção. A primeira coluna desta tabela apresenta o padrão vogal-consoante contemplado pelo mapeamento expresso na última coluna da tabela. Na tabela são apresentadas 7 posições de cadeia de produção fonética. O segmento sonoro da posição 1 é produzido imediatamente antes do segmento da posição 2, e assim sucessivamente. O fone de interesse para o mapeamento em um dado instante encontra-se na posição central 4. Na tabela adotou-se as seguintes convenções:

- o terceiro fone à esquerda do fone central é identificado pelo índice 1 (posição 1);
- o segundo fone à esquerda do fone central é identificado pelo índice 2 (posição 2);
- o primeiro fone à esquerda do fone central é identificado pelo índice 3 (posição 3);
- o fone central é identificado pelo índice 4 (posição 4);
- o primeiro fone à direita do fone central é identificado pelo índice 5 (posição 5);
- o segundo fone à direita do fone central é identificado pelo índice 6 (posição 6);
- terceiro fone à direita do fone centra é identificado pelo índice 7 (posição 7);
    - o símbolo "#" indica silêncio;
    - o símbolo V indica segmento vocálico;
    - o símbolo C indica segmento consonantal;
    - a notação V_i indica o segmento vocálico na posição i;
    - a notação C_i indica o segmento consonantal na posição i;
    - o símbolo "*" (asterisco) indique que não importa qual o tipo de fone;
    - a notação [i] indica o segmento sonoro [i], consoante a notação IPA.

TABELA 3: Mapeamento dos segmentos consonantais para visemas.

	Posição							
Padrão	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Mapeamento**
#CV	*	*	#	C	V	*	*	$\#C_4V_5$
VCV	*	*	V	C	V	*	*	$V_3C_4V_5$
#CCV	*	*	#	C	C ∈ {l, □}	V	*	$\#C_4V_6$

	*	#	C	C ∈ {l, □}	V	*	*	$V_5C_4V_5$
	*	*	#	C ∈ {p, f, t, k}	C ∉ {l, □}	V	*	$\#C_4[i]$
	*	#	C ∈ {p, f, t, k}	C ∉ {l, □}	V	*	*	$[i]C_4V_5$
VCCV	*	*	V	C	C ∈ {l, □}	V	*	$V_3C_4V_6$
	*	V	C	C ∈ {l, □}	V	*	*	$V_5C_4V_5$
	*	*	V	C ∈ {p, f, t, k}	C ∉ {l, □}	V	*	$V_3C_4[i]$
	*	V	C ∈ {p, f, t, k}	C ∉ {l, □}	V	*	*	$[i]C_4V_5$
	*	*	V	C ∈ {s, □}	C	V	*	$V_3C_4V_3$
	*	V	C ∈ {s,r}	C	V	*	*	$V_2C_4V_5$
VCCCV	*	*	V	C ∈ {s, □}	C	C	V	$V_3C_4V_3$
	*	V	C ∈ {s, □}	C	C	V	*	$V_2C_4V_6$
	V	C ∈ {s, □}	C	C	V	*	*	$V_5C_4V_5$
	*		V	C	C	C	V	$V_3C_4V_3$
	*	V	C	C	C	V	*	$V_2C_4V_6$
	V	C	C	C	V	*	*	$\#C_4V_5$
VCCCCV	*	*	V	C	C	C	C	$V_3C_4\ V_3$
	*	V	C	C	C	C	V	$V_2C_4\ V_2$
	V	C	C	C	C	V	*	$V_1C_4\ V_1$

	C	C	C	C	V	*	*	$V_5C_4V_5$

Em uma realização do invento, antes do mapeamento expresso na Tabela 3, é efetuado primeiro o mapeamento dos representantes dos grupos homofemas vocálicos para os segmentos vocálicos extremos tônicos [i, a, u] e pós-tonicos [□, □, □] com expresso na Tabela 4.

TABELA 4: Mapeamento dos segmentos vocálicos para os segmentos extremos.

Representante Grupo Homofemas Vocálicos	Mapeamento
[i]	[i]
[e]	[□]
[□]	[a]
[a]	[a]
[□]	[□]
[□]	[□]
[o]	[u]
[u]	[u]
[□]	[□]
[□]	[□]
[□]	[□]

Após a conversão fone-visema, obtém-se a sequência de visemas correspondentes às poses-chave da animação. Observa-se que o intervalo de transição entre estas poses-chave está especificado na transcrição fonética temporizada realizada na etapa Geração da Fala e Informação Associada – Figura 1(1).

A transcrição fonética temporizada fornece ao sistema de animação a informação sobre o instante de início de cada fone e sua respectiva duração. As poses-chaves são alocadas no meio do intervalo da duração do fone. As imagens intermediárias entre duas poses-chave correspondentes aos quadros da animação são geradas através da técnica de metamorfose entre imagens.

A partir da pose-chave inicial, uma transformação espacial (*warping*) deforma a primeira imagem (origem) gradualmente até a segunda imagem (destino), a pose-chave final. De maneira simétrica, uma transformação de *warping* é realizada distorcendo a imagem final em direção à imagem inicial. A metamorfose é realizada combinando-se as duas distorções espaciais em direções opostas através interpolação de valores de cores dos pixels de cada imagem, operação comumente denominada dissolução cruzada (*cross-dissolve*). No início da nova sequência de imagens o maior peso da função de interpolação é dado ao sentido de distorção da imagem origem para imagem destino, e o mesmo é reduzido gradualmente, até o final da animação onde o maior peso é atribuído às imagens do sentido de distorção destino-origem.

O *warping* entre as imagens inicial e final pode ser obtido através da trajetória de características-âncora da imagem da face. Em uma realização da invenção foram utilizados os pontos-âncora mostrados na Figura 5. Uma estratégia de *warping* implementada em uma realização da invenção é a baseada em funções de base radial (RBF - *Radial Basis Function*).

Considere uma imagem origem que deve ser transformada em uma imagem de destino, sendo que *n* pontos-âncora da imagem de origem devem ser mapeados obrigatoriamente para *n* pontos-âncora identificados da imagem de destino. A função de transformação baseada em RBF é dada por:

$$\begin{cases} u_i = f_u(P_i) = f_u(x_i, y_i) = b_m(P_i) + \sum_{j=1}^{n} \alpha_j \, \phi_j(d(P_i, P_j)) \\ v_i = f_v(P_i) = f_v(x_i, y_i) = b_m(P_i) + \sum_{j=1}^{n} \beta_j \, \phi_j(d(P_i, P_j)) \end{cases}$$

onde:

- os pontos-âncora são identificados pelo índice $i = 1,...,n$;
- u_i e v_i são as coordenadas de um ponto $Q_i = (u_i, v_i)$ na imagem resultante;
- $P_i=(x_i, y_i)$ é um ponto da imagem origem dado pelas coordenadas (x_i, yi);
- $b_m(P_i)$ é um polinômio de grau m;
- $d(P_i, P_j)$ é a distância entre o ponto P_j e o ponto P_i;
- $\phi_j(d(P_i, P_j))$ é uma função de base radial;
- α_j é um coeficiente escalar.

Uma realização da invenção utiliza:

$$b_{m=1}(P_i) = b_1(P_i) = a_1 + a_2\ x_i + a_3\ y_i$$

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$\phi_j(d(P_i, P_j)) = \sqrt{(d(P_i, P_j)^2 + r_j^{\,2})}$$

onde:

$$r_j = \min\{\, d(P_i, P_j)\,\},\ i = 1, \ldots, n \ \text{ e } \ i \neq j$$

Os coeficientes a_1, a_2, a_3 e α_i e β_i $i=1,..n$ são determinados através da solução do sistema de equações lineares resultantes das condições de contorno impostas para o mapeamento dos pontos-âncoras.

Considerando a realização mencionada acima os coeficientes a_1, a_2, a_3 e α_i e β_i $i=1,..5$ são determinados pela resolução dos seguintes sistema de equações lineares:

$$\begin{bmatrix} 1 & x_1 & y_1 & \left(d(P_1,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_1,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_1,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_1,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_1,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_2 & y_2 & \left(d(P_2,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_2,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_2,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_2,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_2,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_3 & y_3 & \left(d(P_3,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_3,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_3,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_3,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_3,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_4 & y_4 & \left(d(P_4,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_4,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_4,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_4,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_4,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_5 & y_5 & \left(d(P_5,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_5,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_5,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_5,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_5,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 & x_4 & x_5 \\ 0 & 0 & 0 & y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_1 & y_1 & \left(d(P_1,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_1,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_1,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_1,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_1,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_2 & y_2 & \left(d(P_2,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_2,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_2,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_2,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_2,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_3 & y_3 & \left(d(P_3,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_3,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_3,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_3,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_3,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_4 & y_4 & \left(d(P_4,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_4,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_4,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_4,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_4,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 1 & x_5 & y_5 & \left(d(P_5,P_1)^2+r_1^2\right)^{\frac{1}{2}} & \left(d(P_5,P_2)^2+r_2^2\right)^{\frac{1}{2}} & \left(d(P_5,P_3)^2+r_3^2\right)^{\frac{1}{2}} & \left(d(P_5,P_4)^2+r_4^2\right)^{\frac{1}{2}} & \left(d(P_5,P_5)^2+r_5^2\right)^{\frac{1}{2}} \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 & x_4 & x_5 \\ 0 & 0 & 0 & y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Durante o processo de metamorfose entre quadros-chave, são necessárias sucessivas operações de transformação espacial para os quadros intermediários. A determinação dos pontos-âncora para os quadros intermediários é feita a partir do conhecimento dos deslocamentos das coordenadas *x* e *y* associados aos pontos-âncora das poses-chave inicial e final.

Uma vez que a definição da trajetória destes pontos interfere diretamente na movimentação articulação visível, buscou-se utilizar uma trajetória não-linear próxima da realidade. Em uma realização da invenção, a trajetória dos pontos-âncora nos quadros intermediários é realizado através de uma curva de interpolação que garanta derivada temporal (velocidade) igual a zero no instante das poses-chaves. Uma função possível é a curva de Hermite dada por:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_I & x_F \\ y_I & y_F \end{bmatrix} \begin{bmatrix} 2 & -3 & 1 \\ -2 & 3 & 0 \end{bmatrix} \begin{bmatrix} t^3 \\ t^2 \\ 1 \end{bmatrix} \quad 0 \le t \le 1$$

onde:

- $x(t)$ e $y(t)$ são as coordenadas do ponto-âncora de interesse;
- x_I e y_I são as coordenadas x e y do ponto-âncora de interesse para a pose-chave inicial;
- x_F e y_F são as coordenadas x e y do ponto-âncora de interesse para a pose-chave final;
- t é a variável independente da representação paramétrica normalizada em relação ao intervalo de tempo entre as pose-chave inicial e final.

EXEMPLO

Na Etapa de Geração e Geração da Fala e Informação Associada – Figura 1(1) foi utilizado um conversor texto fala para a conversão do seguinte texto: "O partido trabalhista brasileiro foi criado por Getúlio Vargas depois da segunda guerra mundial."

Neste exemplo, o conversor texto fala gera o áudio associado à locução e a transcrição fonética temporizada apresentada na Tabela 5.

TABELA 5: Exemplo de transcrição fonética temporizada.

Fone	Duração (ms)
#	278
[□]	163
[p]	78
[a]	126
[□]	56
[t]	64
[i]	93
[d]	56
[□]	77

[t]	82
[□]	59
[a]	127
[b]	49
[a]	124
[□]	75
[i]	78
[s]	54
[t]	65
[□]	78
[b]	50
[□]	60
[a]	128
[z]	58
[i]	86
[l]	38
[e]	74
[□]	55
[□]	73
[□]	76
[f]	139
[o]	127
[□]	89
[k]	77
[□]	68
[i]	80
[a]	127

[d]	66
[□]	69
[p]	113
[o]	185
[□]	58
[□]	58
[e]	106
[t]	77
[u]	97
[l]	47
[□]	78
[□]	59
[v]	69
[a]	185
[□]	48
[g]	49
[□]	47
[s]	50
[d]	61
[e]	109
[p]	70
[o]	85
[□]	55
[s]	51
[d]	66
[a]	217
[s]	89

[e]	112
[g]	52
[ũ]	119
[d]	58
[□]	76
[g]	70
[□]	157
[□]	79
[□]	41
[m]	62
[ũ]	112
[d]	60
[i]	89
[a]	74
[□]	74
#	243

Esta transcrição fonética temporizada é enviada para etapa de Animação Facial 2D – Figura 1(2), como representado pela seta (7) da Figura 1.

A etapa Animação Facial 2D – Figura 1(2), em um primeiro momento realiza o mapeamento dos fones da transcrição fonética para os respectivos representantes dos grupo homofemas como indicado nas Tabelas 1 e 2. O resultado deste mapeamento é apresentado na Tabela 6.

TABELA 6: Mapeamento para os representantes dos grupos homofemas vocálicos e consonantais.

Fone	Representante Grupo Homofema
#	#
[□]	[□]
[p]	[p]
[a]	[a]
[□]	[□]
[t]	[t]
[i]	[i]
[d]	[t]
[□]	[□]
[t]	[t]
[□]	[□]
[a]	[a]
[b]	[p]
[a]	[a]
[□]	[□]
[i]	[i]
[s]	[s]
[t]	[t]
[□]	[□]
[b]	[p]
[□]	[□]
[a]	[a]
[z]	[s]
[i]	[i]
[l]	[l]

[e]	[e]
[□]	[□]
[□]	[□]
[□]	[□]
[f]	[f]
[o]	[o]
[□]	[□]
[k]	[k]
[□]	[□]
[i]	[i]
[a]	[a]
[d]	[t]
[□]	[□]
[p]	[p]
[o]	[o]
[□]	[□]
[□]	[□]
[e]	[e]
[t]	[t]
[u]	[u]
[l]	[l]
[□]	[□]
[□]	[□]
[v]	[f]
[a]	[a]
[□]	[□]
[g]	[k]

[□]	[□]
[s]	[s]
[d]	[t]
[e]	[e]
[p]	[p]
[o]	[o]
[□]	[□]
[s]	[s]
[d]	[t]
[a]	[a]
[s]	[s]
[e]	[e]
[g]	[k]
[ũ]	[u]
[d]	[t]
[□]	[□]
[g]	[k]
[□]	[□]
[□]	[□]
[□]	[□]
[m]	[p]
[ũ]	[u]
[d]	[t]
[i]	[i]
[a]	[a]
[□]	[□]
#	#

Após o mapeamento para os representantes dos grupos homofemas, a etapa Animação Facial 2D – Figura 1(2) efetua o mapeamento dos representantes vocálicos para os visemas vocálicos utilizando a Tabela 2. O resultado deste mapeamento é apresentado na Tabela 7.

TABELA 7: Mapeamento para os visemas vocálicos.

Representante Grupo Homofema	Visema Vocálico
#	
[□]	<□>
[p]	
[a]	<a>
[□]	
[t]	
[i]	<i_2>
[t]	
[□]	<□>
[t]	
[□]	
[a]	<a>
[p]	
[a]	<a>
[□]	
[i]	<i_1>
[s]	
[t]	
[□]	<a>
[p]	
[□]	

[a]	<a>
[s]	
[i]	<i_1>
[l]	
[e]	<e>
[□]	<□>
[□]	
[□]	<□>
[f]	
[o]	<o>
[□]	<□>
[k]	
[□]	
[i]	<i_1>
[a]	<a>
[t]	
[□]	<□>
[p]	
[o]	<o>
[□]	
[□]	
[e]	<e>
[t]	
[u]	<u>
[l]	
[□]	<□>
[□]	<□>
[f]	
[a]	<a>
[□]	

[k]	
[□]	<□>
[s]	
[t]	
[e]	<e>
[p]	
[o]	<o>
[□]	<□>
[s]	
[t]	
[a]	<a>
[s]	
[e]	<e>
[k]	
[u]	<u>
[t]	
[□]	<□>
[k]	
[□]	<□>
[□]	
[□]	<□>
[p]	
[u]	<u>
[t]	
[i]	<i_1>
[a]	<a>
[□]	<□>
#	

Após a determinação dos visemas vocálicos, a etapa de Animação Facial 2D –Figura 1(2) inicia o processo de determinação dos visemas consonantais efetuando o mapeamento dos representantes dos grupos homofemas vocálicos para os segmentos vocálicos extremos como especificado na Tabela 4. O resultado deste mapeamento é apresentado na Tabela 8.

TABELA 8: Mapeamento preparatório para o mapeamento para os visemas consonantais.

Representante Grupo Homofema	Sequência para o mapeamento do visemas consonontais
#	#
[□]	[□]
[p]	[p]
[a]	[a]
[□]	[□]
[t]	[t]
[i]	[i]
[t]	[t]
[□]	[□]
[t]	[t]
[□]	[□]
[a]	[a]
[p]	[p]
[a]	[a]
[□]	[□]
[i]	[i]
[s]	[s]
[t]	[t]

[□]	[□]
[p]	[p]
[□]	[□]
[a]	[a]
[s]	[s]
[i]	[i]
[l]	[l]
[e]	[□]
[□]	[□]
[□]	[□]
[□]	[□]
[f]	[f]
[o]	[u]
[□]	[□]
[k]	[k]
[□]	[□]
[i]	[i]
[a]	[a]
[t]	[t]
[□]	[□]
[p]	[p]
[o]	[u]
[□]	[□]
[□]	[□]
[e]	[□]
[t]	[t]
[u]	[u]
[l]	[l]
[□]	[□]
[□]	[□]

[f]	[f]
[a]	[a]
[□]	[□]
[k]	[k]
[□]	[□]
[s]	[s]
[t]	[t]
[e]	[□]
[p]	[p]
[o]	[u]
[□]	[□]
[s]	[s]
[t]	[t]
[a]	[a]
[s]	[s]
[e]	[□]
[k]	[k]
[u]	[u]
[t]	[t]
[□]	[□]
[k]	[k]
[□]	[a]
[□]	[□]
[□]	[□]
[p]	[p]
[u]	[u]
[t]	[t]
[i]	[i]
[a]	[a]
[□]	[□]

#	#

Continuando, a etapa Animação Facial 2D – Figura 1(2) efetua a conversão dos segmentos consonantais para os respectivos visemas consonantais. O mapeamento é efetuado utilizando a Tabela 3. O resultado é apresentado na Tabela 9. Nesta tabela é indicado o contexto fonético considerado e a respectiva linha da Tabela 3 que estabelece o contexto e o mapeamento. A linha 1 da Tabela 3 apresentao padrão #CV; a linha 2 o padrão VCV, e assim por diante as linhas são numeradas até 22.

TABELA 9: Mapeamento para os visemas consonantais.

Sequência para o mapeamento do visemas consonontais	Mapeamento Tabela 3	Contexto	Visema
#			
[□]			
[p]	linha 2	[□pa]	<p_1>
[a]			
[□]	linha 11	[a□a]	<$□_1$>
[t]	linha 12	[ati]	<t_1>
[i]			
[t]	linha 2	[it□]	<t_1>
[□]			
[t]	linha 7	[□ta]	<t_1>
[□]	linha 8	[a□a]	<$□_1$>
[a]			
[p]	linha 2	[apa]	<p_1>
[a]			
[□]	linha 2	[a□i]	<$□_1$>

[i]			
[s]	linha 11	[isi]	<s_1>
[t]	linha 12	[it□]	<t_1>
[□]			
[p]	linha 7	[□pa]	<p_1>
[□]	linha 8	[□□a]	<$□_1$>
[a]			
[s]	linha 2	[asi]	<s_1>
[i]			
[l]	linha 2	[il□]	<l_2>
[□]			
[□]			
[□]	linha 2	[□□□]	<$□_2$>
[□]			
[f]	linha 2	[□fu]	<f_2>
[u]			
[□]			
[k]	linha 7	[□ki]	<k_1>
[□]	linha 8	[i□i]	<$□_1$>
[i]			
[a]			
[t]	linha 2	[at□]	<t_1>
[□]			
[p]	linha 2	[□pu]	<p_2>
[u]			
[□]	linha 11	[u□u]	<$□_2$>
[□]	linha 12	[u□□]	<$□_1$>

[□]			
[t]	linha 2	[□tu]	<t_1>
[u]			
[l]	linha 2	[ul□]	<l_1>
[□]			
[□]			
[f]	linha 2	[□fa]	<f_2>
[a]			
[□]	linha 11	[a□a]	<$□_1$>
[k]	linha 12	[ak□]	<k_2>
[□]			
[s]	linha 11	[□s□]	<s_1>
[t]	linha 12	[□t□]	<t_2>
[□]			
[p]	linha 2	[□pu]	<p_1>
[u]			
[□]			
[s]	linha 11	[□s□]	<s_1>
[t]	linha 12	[□ta]	<t_1>
[a]			
[s]	linha 2	[as□]	<s_1>
[□]			
[k]	linha 2	[□ku]	<k_3>
[u]			
[t]	linha 2	[ut□]	<t_1>
[□]			
[k]	linha 2	[□ka]	<k_2>

[a]			
[□]	linha 2	[a□□]	<□$_1$>
[□]			
[p]	linha 2	[□pu]	<p$_1$>
[u]			
[t]	linha 2	[uti]	<t$_1$>
[i]			
[a]			
[□]			
#			

Das Tabelas 5, 7 e 9 é possível derivar a Tabela 10 que apresenta o resultado da conversão da transcrição fonética temporizada para a sequência de visemas que será utilizada para a geração da animação pela etapa Animação Facial 2D – Figura 1(2).

Tabela 10: Resultado do mapeamento para visemas dependente do contexto fonético.

Fone	Duração (ms)	Visema Dependente do Contexto	Imagem Figura 2
#	278	#	(1)
[□]	163	<□>	(30)
[p]	78	<p$_1$>	(2)
[a]	126	<a>	(26)
[□]	56	<□$_1$>	(22)
[t]	64	<t$_1$>	(6)
[i]	93	<i$_2$>	(25)
[d]	56	<t$_1$>	(6)

[□]	77	<□>	(30)
[t]	82	<t_1>	(6)
[□]	59	<$□_1$>	(22)
[a]	127	<a>	(26)
[b]	49	<p_1>	(2)
[a]	124	<a>	(26)
[□]	75	<$□_1$>	(16)
[i]	78	<i_1>	(24)
[s]	54	<s_1>	(8)
[t]	65	<t_1>	(6)
[□]	78	<a>	(26)
[b]	50	<p_1>	(2)
[□]	60	<$□_1$>	(22)
[a]	128	<a>	(26)
[z]	58	<s_1>	(8)
[i]	86	<i_1>	(24)
[l]	38	<l_2>	(11)
[e]	74	<e>	(31)
[□]	55	<□>	(28)
[□]	73	<$□_2$>	(23)
[□]	76	<□>	(30)
[f]	139	<f_2>	(5)
[o]	127	<o>	(34)
[□]	89	<□>	(28)
[k]	77	<k_1>	(19)
[□]	68	<$□_1$>	(22)
[i]	80	<i_1>	(24)

[a]	127	<a>	(26)
[d]	66	<t_1>	(6)
[□]	69	<□>	(30)
[p]	113	<p_2>	(3)
[o]	185	<o>	(34)
[□]	58	<□$_2$>	(23)
[□]	58	<□$_1$>	(14)
[e]	106	<e>	(31)
[t]	77	<t_1>	(6)
[u]	97	<u>	(27)
[l]	47	<l_1>	(10)
[□]	78	<□>	(28)
[□]	59	<□>	(30)
[v]	69	<f_2>	(5)
[a]	185	<a>	(26)
[□]	48	<□$_1$>	(22)
[g]	49	<k_2>	(20)
[□]	47	<□>	(29)
[s]	50	<s_1>	(8)
[d]	61	<t_2>	(7)
[e]	109	<e>	(31)
[p]	70	<p_1>	(2)
[o]	85	<o>	(34)
[□]	55	<□>	(28)
[s]	51	<s_1>	(8)
[d]	66	<t_1>	(6)
[a]	217	<a>	(26)

[s]	89	<s_1>	(8)
[e]	112	<e>	(31)
[g]	52	<k_3>	(21)
[ũ]	119	<u>	(27)
[d]	58	<t_1>	(6)
[□]	76	<□>	(29)
[g]	70	<k_2>	(20)
[□]	157	<□>	(32)
[□]	79	<□$_1$>	(22)
[□]	41	<□>	(29)
[m]	62	<p_1>	(2)
[ũ]	112	<u>	(27)
[d]	60	<t_1>	(6)
[i]	89	<i_1>	(24)
[a]	74	<a>	(26)
[□]	74	<□>	(30)
#	243	#	(1)

A Etapa Animação Facial 2D – Figura 1(2), após a conversão da transcrição fonética temporizada para a sequência de visemas, extrai da Base de Imagens – Figura 1(3) as imagens associadas aos visemas, combina com a imagem-base, efetua a metamorfose entre imagens resultantes, gerando as imagens intermediárias entre os visemas. A sequência de imagens assim produzida é então enviada para a etapa de Sincronização – Figura 1(4).

A etapa de Sincronização cuida da apresentação sincronizada das imagens com o áudio no dispositivo de saída – Figura 1(5). Alternativamente, as imagens e o áudio devidamente sincronizados e compactados podem ser armazenadas na forma de arquivo de vídeo em um dispositivo de armazenamento.

QUADRO REIVINDICATÓRIO

1. Método para geração de animação facial **videorealista** 2D a partir do áudio da fala caracterizado por compreender ao menos 34 visemas representantes dos fonemas do português do Brasil contendo:

- Geração da Fala e Informação Associada;
- Construção, manipulação e interpolação da Base de imagens de referência;
- Sequenciamento de imagens;
- Manipulação do áudio;
- Temporarização dos fones;
- Sincronização de áudio e visemas; e
- Animação 2D sincronizada com a fala. (software)

2. Método de acordo com a reivindicação 1, caracterizado pela geração da fala ser gerada por um locutor real.

3. Método de acordo com a reivindicação 1, caracterizado pela geração da fala ser gerada sinteticamente por um conversor texto-fala.

4. Método de acordo com a reivindicação 1 a 3, caracterizado por compreender transcrição fonética ser gerada dentre: um processo de conversão texto-fala, uma segmentação fonética automática e uma segmentação fonética efetuada manualmente.

5. Método de acordo com a reivindicação 1 a 4, caracterizado por compreender a seleção e manipulação das imagens de referência (fig1.8) para geração da sequência das ditas imagens (visemas) requeridas pelo usuário.

6. Método de acordo com a reivindicação 1 a 5, caracterizado por compreender uma base de dados de imagens fotográficas correspondentes às posturas articulatórias dependentes do contexto fonético, representantes dos diversos fonemas do português do Brasil.

7. Método de acordo com a revindicação 5, caracterizado por compreender mais de trinta e quatro imagens correspondentes aos visemas representantes do português do Brasil.

8. Método de acordo com a reivindicação 1 a 7, caracterizado por compreender aos menos 22 imagens correspondentes a visemas consonantais, ao menos 11 imagens correspondentes a visemas vocálicos, e ao menos 1 imagem representante da postura labial de repouso.

9. Método de acordo com a reivindicação 1 a 8, caracterizado por compreender contextos fonéticos de até sete segmentos.

10. Método de acordo com a reivindicação 1 a 8, caracterizado por compreender contextos fonéticos de sete ou mais segmentos.

11. Método de acordo com a reivindicação todas as reivindicações anteriores, caracterizado pelo rotulamento de imagem, seleção e manipulação na síntese da animação.

12. Método de acordo com a reivindicação 1, 2, 4 a 11, caracterizado por compreender processamento de segmentos de vídeos da face de um locutor real reproduzindo logatomas representantes de todos os fonemas do português do Brasil.

13. Método de acordo com todas as reivindicações anteriores, caracterizado por compreender correção de variações de enquadramento dos elementos (registro) da face, tais como olhos, nariz e lábios, da base de imagens em relação a uma imagem base.

14. Método de acordo com a reivindicação anterior, caracterizado por compreender pontos de elementos faciais estáticos durante a produção da fala.

15. Método de acordo com a reivindicação 14, caracterizado por compreender pontos de referência dentre: um ponto médio do lábio superior; um ponto médio do lábio inferior; um ponto no canto direito da boca; um ponto no canto esquerdo da boca; e um ponto no extremo inferior do queixo.

16. Método de acordo com todas as reivindicações anteriores, caracterizado por compreender a região da boca e queixo das imagens recortadas das imagens originais dos visemas (fig. 4 - necessário coordenadas mais específicas).

17.Método de acordo com todas as reivindicações anteriores, caracterizado por compreender combinação da região da boca e queixo do visema de interesse combinado com a imagem base.

18.Método de acordo com a reivindicação 16 e 17, caracterizado por conter uma máscara de transparência em um arquivo do tipo de imagem (fig. 4.4).

19.Método de acordo com todas as reivindicações anteriores, caracterizado por obter a sequência de visemas decorrente da conversão fone-homofema.

20.Método de acordo com todas as reivindicações anteriores, caracterizada por compreender visemas vocálicos independentes do contexto, excluindo os visemas $<i_1>$ e $<i_2>$.

21.Método de acordo com todas reivindicação anteriores, caracterizado por compreender um contexto fonético consonantal em sequência selecionada dentre: dois fones (CV), três fones (VCV), e ambos os fones.

22.Método de acordo com todas reivindicação anteriores, caracterizado por compreender o uso da curva de Hermite para movimentação articulatória visível.

23.Método de acordo com todas as reivindicações anteriores, caracterizado por compreender uma integibilidade até 6 vezes maior em relação à animação 3D utilizando visemas dependentes do contexto fonético.

24.Método de acordo com todas as reivindicações anteriores, caracterizado pela coarticulação ser antecipatória e perseveratória na conversão fone-homofema.

25.Método de acordo com todas as reivindicações anteriores, caracterizado por compreender uma transformação espacial da imagem registrada com a imagem face-base do tipo RBF (Radial Basis Function).

26.Método de acordo com todas as reivindicações anteriores, caracterizado por ser utilizado na reprodução de animação facial 2D em equipamentos com baixa capacidade de armazenamento e processamento.

27.Método de acordo com todas as reivindicações anteriores, caracterizado por ser utilizado na reprodução de animação facial 2D em equipamentos com baixa capacidade de armazenamento e processamento incluindo : microcomputadores, telefones celulares, assistentes pessoais digitais (PDAs), smartphones, set-top boxes digitais, , máquinas de atendimento bancário, sistema de reserva de vôos e sistema de reserva de consulta.

FIGURAS

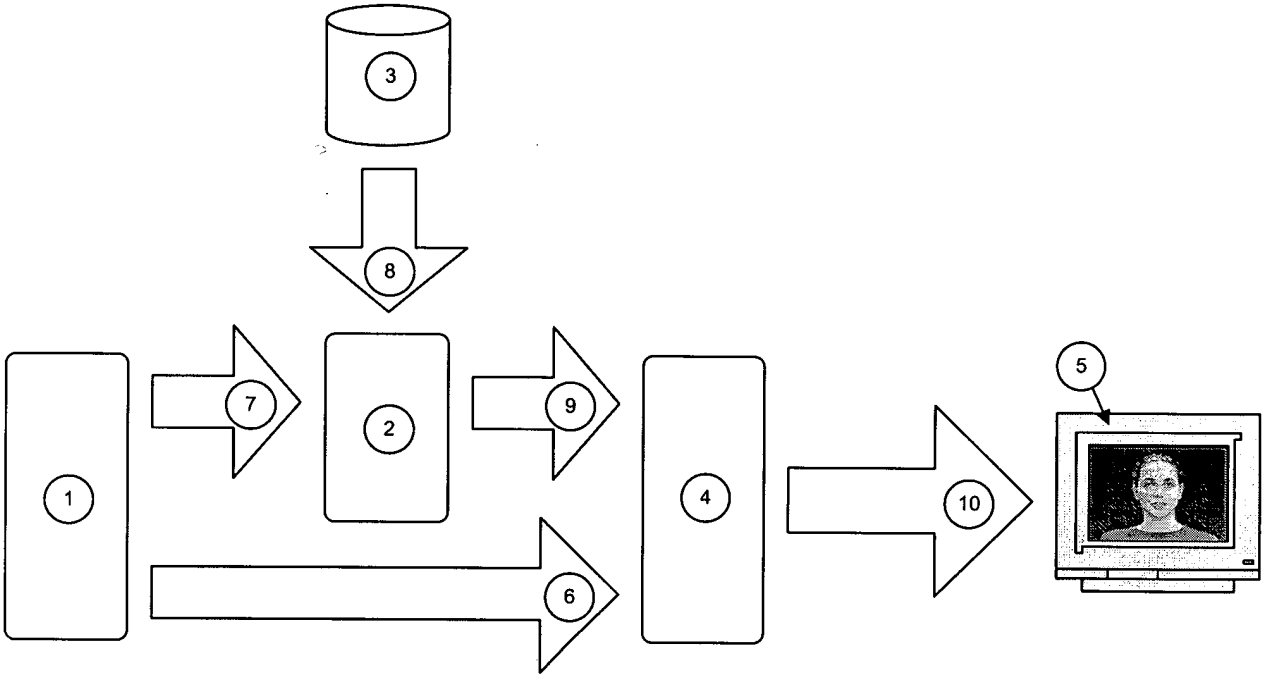


FIGURA 1

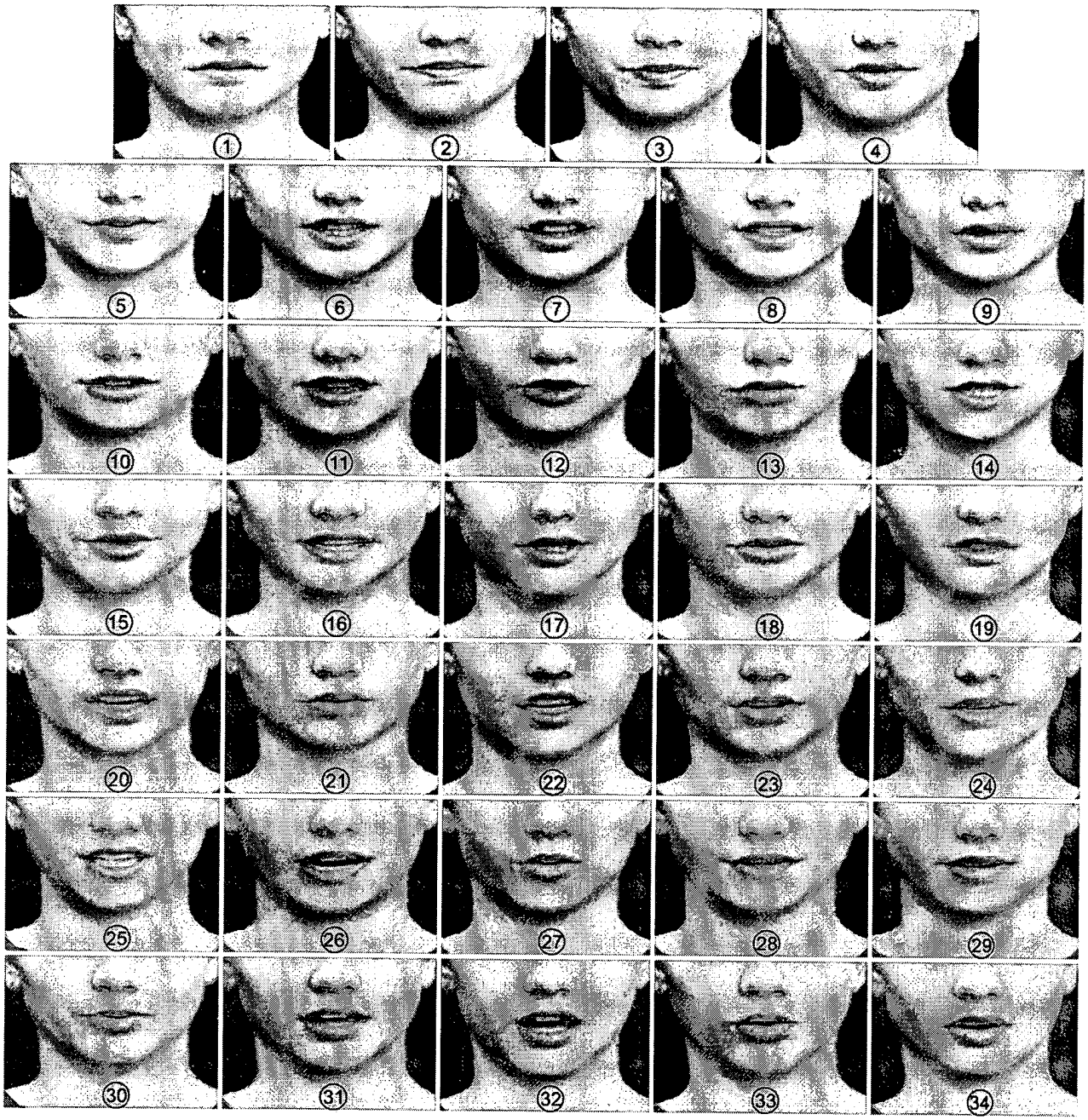

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

FIGURA 2

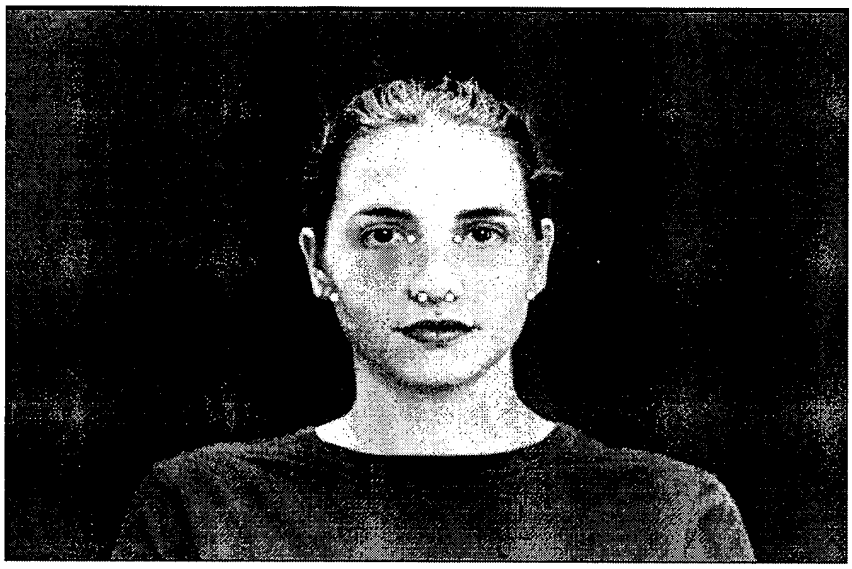


FIGURA 3

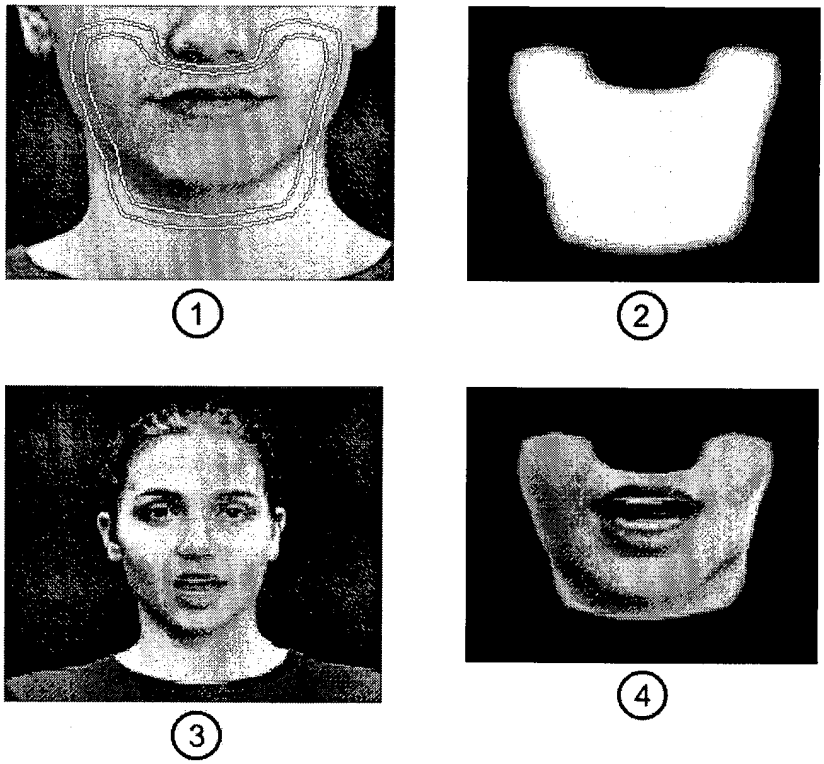


FIGURA 4

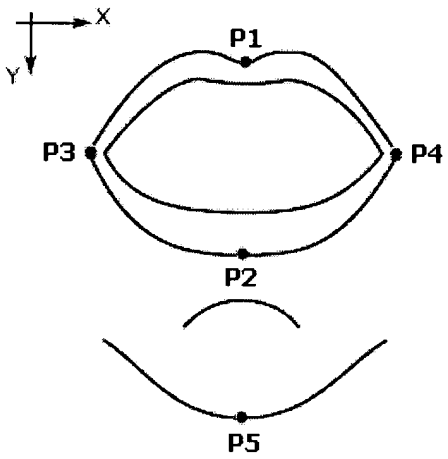


FIGURA 5

PI 0903935-0

RESUMO

Sistema de síntese de animação facial por computador baseada na manipulação de imagens.

A invenção refere-se a um sistema de síntese de animação facial por computador sincronizada com a fala e videorealista. Neste sistema a animação facial é gerada através da seleção, manipulação e apresentação de um conjunto reduzido de imagens, representando visemas dependentes do contexto fonético, que reproduz a movimentação articulatória visível na face durante a produção da fala, incluindo os efeitos da coarticulação. A abordagem adotada permite que seja possível gerar animações com alto grau de videorealismo, mesmo em plataformas com reduzida capacidade de armazenamento, como os dispositivos móveis celulares. Entende-se como animação videorealista uma animação que pode ser confundida com um vídeo real.