Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e Computação

Implementação e Avaliação de um Apresentador Virtual Tridimensional

Autora: Denise Hippler
Orientador: José Mario De Martino

Tese de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Engenharia de Computação.

Banca Examinadora:
Prof. Dr. José Mario De Martino - DCA/FEEC/UNICAMP
Prof. Dr. Léo Pini Magalhães - DCA/FEEC/UNICAMP
Prof. Dr. Márcio Lobo Netto - USP

Campinas, SP
Novembro/2008

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

H612i

Hippler, Denise
Implementação e avaliação de um apresentador virtual tridimensional / Denise Hippler. – Campinas, SP: [s.n.], 2008.

Orientador: José Mario De Martino.
Dissertação de Mestrado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Computação gráfica. 2. Animação por computador. I. De Martino, José Mario. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título

Título em Inglês: Implementation and evaluation of a three-dimensional virtual presenter
Palavras-chave em Inglês: Computer graphics, Facial animation
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca Examinadora: Márcio Lobo Netto, Léo Pini Magalhães
Data da defesa: 11/11/2008
Programa de Pós Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidata: Denise Hippler

Data da Defesa: 11 de novembro de 2008

Título da Tese: "Implementação e Avaliação de um Apresentador Virtual Tridimensional"

Prof. Dr. José Mario De Martino (Presidente): _____

Prof. Dr. Marcio Lobo Netto: _____

Prof. Dr. Léo Pini Magalhães: _____

Resumo

Um apresentador virtual é um agente representado por uma forma antropomórfica que permite a transmissão de informações usando uma combinação de formas de expressão que pode envolver elementos como fala, gestos, expressões faciais e mudanças de postura, podendo ainda incluir a interação com objetos. Apresentadores virtuais possuem um amplo leque de aplicações em potencial, como guias e instrutores. Este trabalho relata a implementação e avaliação de um apresentador virtual de notícias que realiza movimentos verbais sincronizados com o sinal acústico da fala e movimentos não-verbais que complementam a fala, estruturando a apresentação. A animação é gerada a partir de informações acerca do texto a ser apresentado oralmente e dos movimentos não-verbais que devem acompanhá-lo, descritas pelo usuário em um arquivo de entrada. De modo a gerar animações convincentes, estudou-se gravações de apresentações de notícias de um canal televisivo e derivou-se regras de comportamento, um conjunto de movimentos que se repetem em dadas situações. Esse padrão comportamental insere alguns movimentos automaticamente na animação, complementando os movimentos dados pelo usuário. Por fim, realizando experimentos com voluntários, avaliou-se a contribuição dos movimentos não-verbais à apresentação.

Palavras-chave: Computação Gráfica, Animação Facial, Agente Virtual, Modelo Comportamental, Avaliação de Personagem Virtual.

Abstract

A virtual presenter is an anthropomorphic graphical representation, that transmits informations using a combination of communication possibilities like speech, gesture, facial expression, posture and interaction with virtual objects. Embodied Conversational Agents (ECAs) have several potential applications like guides and instructors. This work reports on the implementation and evaluation of a virtual news presenter that makes verbal movements in synchrony with the acoustic speech signal and non-verbal movements that complement the speech, structuring the presentation. The input file contains the description of the text that should be presented and the movements to be performed. In order to generate convincing animations, a real news presenter has been studied allowing the definition of behavior rules, a group of movements that occur in certain situations. This behavior model inserts some movements automatically in the animation, complementing the movements given by the user. Experiments were made with volunteers to evaluate the contribution of the nonverbal movements to the presentation.

Keywords: Computer Graphics, Facial Animation, Virtual Agent, Behavior Model, Virtual Character Evaluation

Agradecimentos

aos meus pais, Virginia e Euclides pelos incentivos ao longo da vida, por compartilharem sua sabedoria e me apoiarem

aos meus irmãos pelo carinho, apoio, paciência e irreverência

aos meus avós Noelina e John, pelo importante papel que tiveram

ao Everton pelas dicas, pelo tempo, pelo carinho e inúmeras vezes em que me apoiou

aos amigos Mary, Mirelle, Cauê, Valéria, Schammass, Lucía e “ondevamo” por existirem

ao orientador José Mario De Martino, pela paciência, empenho, ensinamentos, valiosas sugestões e pela orientação

ao Cláusius e ao Charles pelas dicas sobre o modelo do apresentador virtual, ao Jeremias e ao Alex pelas conversas

aos colegas da FEEC que participaram do teste de avaliação, pelo tempo e boa-vontade

ao Felipe Lopasso pelo exaustivo trabalho de segmentação e ao Manuel Schindler pelo programa de detecção do olho

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro

Aos meus queridos pais,
ao meu querido Everton

Sumário

Lista de Figuras **xiii**

Lista de Tabelas **xv**

Lista de Abreviações e Termos Técnicos **xvii**

Trabalhos Publicados Pelo Autor **xix**

1 Introdução **1**

2 Revisão Bibliográfica **7**

2.1 Agentes conversacionais 9

2.1.1 Cenários de atuação 9

2.1.2 Estudos do comportamento humano 12

2.2 Animação facial por computador 15

2.2.1 Abordagens para realizar animações faciais 15

2.2.2 Sistemas de codificação de movimentos faciais 18

2.2.3 Linguagens de marcação para animação facial 22

2.3 Elaboração de testes de avaliação 34

2.3.1 Relatos de testes de avaliação 35

2.4 Considerações finais 39

3 Sistema de animação do apresentador virtual **43**

3.1 Apresentador de notícias 44

3.1.1 Análise do comportamento de um apresentador real 45
3.2 Implementação do sistema 49
3.2.1 *Script* 50
3.2.2 Modelagem dos movimentos faciais 54
3.2.3 Processamento 63
3.3 Considerações finais 64

4 Testes e Avaliação 67
4.1 Protocolo de teste 70
4.1.1 Procedimento 72
4.2 Análise dos resultados 74
4.2.1 Naturalidade 76
4.2.2 Compreensão 78
4.3 Considerações Finais 80

5 Conclusão 83

A Informações complementares sobre a avaliação 95
A.1 Notícias apresentadas e questões correspondentes 95
A.2 Resultados dos testes de avaliação 98
A.2.1 Resultados da questão de naturalidade 98
A.2.2 Resultados da questão de palavras-chave 99
A.2.3 Resultados das questões específicas 116

Lista de Figuras

2.1 Filmes inovadores a seu tempo: a) Tin Toy e b) Final Fantasy 7
2.2 Sistemas de apresentação de acordo com a interação entre usuário e ECAs 10
2.3 Sistemas de apresentação quando há diversos usuários 11
2.4 Arqueamento de sobrancelhas 14
2.5 Pontos-chave definidos no padrão MPEG-4 21
2.6 Arquitetura do sistema BEAT 29
2.7 Gramática do MIMIC 31
2.8 Exemplo de seqüência de animação gerada por um programa MIMIC 32
2.9 Personagem apresentando controle remoto 36

3.1 Modelo do apresentador virtual 45
3.2 *Template* do olho fechado. 46
3.3 Arquitetura do sistema 49
3.4 Espectograma de uma notícia 51
3.5 Inspeção da onda sonora no Cool Edit 51
3.6 Delimitação dos primeiros segmentos sonoros da frase 51
3.7 a) modelo do apresentador, b) textura da face do apresentador virtual 55
3.8 Rotação do pescoço ao redor do eixo x 56
3.9 Rotação do pescoço ao redor do eixo y 56
3.10 Rotação do pescoço ao redor do eixo z 56
3.11 Rotação de olhos: a) ao redor do eixo x, b) ao redor do eixo y 57
3.12 Abertura da pálpebra 58

3.13 Apresentador com: a) olhos fechados e b) olhos semi-abertos e c) olhos abertos . . . 58
3.14 Representação aramada mostrando: a) olhos abertos e b) apertando os olhos 59
3.15 Apresentador com: a) sobrancelhas relaxadas e b) sobrancelhas elevadas 59
3.16 Explicação dos movimentos de sincronismo labial 62
3.17 Apresentador falando os fones “p” e “a” 63

4.1 Modelo de pesquisa 68
4.2 Ambiente de teste 73
4.3 *Box plot* para avaliação da naturalidade 77
4.4 *Box plot* para palavras-chave 78
4.5 *Box plot* para as questões específicas 79

Lista de Tabelas

2.1 Exemplos de unidades de ação (AUs) do sistema FACS 19
2.2 Exemplo de combinações de AUs para formar expressões de emoção 20
2.3 Elementos em FAML 28
2.4 Exemplo do conjunto de falas em holandês e italiano (extraído de [26]) 38

3.1 Códigos dos segmentos consonantais aceitos pelo arquivo de entrada 52
3.2 Códigos dos segmentos vocálicos aceitos pelo arquivo de entrada 53
3.3 Visemas consonantais e seus contextos fonéticos 61
3.4 Visemas vocálicos e seus contextos fonéticos 61

4.1 Experimentos para avaliar o sistema 70
4.2 Média e variância das notas da avaliação de naturalidade 76
4.3 Tabela de resultados para a ANOVA 77
4.4 Média e variância do número de acertos de palavras-chave 79
4.5 Tabela de resultados para a ANOVA (palavras-chave corretas) 79
4.6 Média e variância do número de respostas corretas 80
4.7 Tabela de resultados para a ANOVA (acertos nas questões específicas) 80

A.1 Notas dadas à naturalidade dos apresentadores 99
A.2 Respostas às questões para o fragmento 1 117
A.3 Respostas às questões para o fragmento 2 118
A.4 Respostas às questões para o fragmento 3 119
A.5 Respostas às questões para o fragmento 4 120

A.6 Respostas às questões para o fragmento 5 121
A.7 Respostas às questões para o fragmento 6 122
A.8 Respostas às questões para o fragmento 7 123
A.9 Respostas às questões para o fragmento 8 124
A.10 Respostas às questões para o fragmento 9 125
A.11 Respostas às questões para o fragmento 10 126

Lista de Abreviações e Termos Técnicos

AU - Unidade de ação usada no sistema FACS.

ECA - Agente conversacional personificado, nome em português para a expressão em inglês *Embodied Conversational Agent*.

FACS - Do inglês *Facial Action Coding System*, sistema de codificação de ação facial.

Framework - No desenvolvimento do software, um framework ou arcabouço é uma estrutura de suporte definida em que um outro projeto de software pode ser organizado e desenvolvido (trecho extraído de [71]).

Tag - comandos envoltos por "<" e ">" usados para criar marcações em arquivos baseados em XML.

TTS - do inglês *Text-To-Speech*, expressão usada para sistemas que convertem texto para fala.

Trabalhos Publicados Pelo Autor

1. Hippler, D.; De Martino, J. M. Virtual Presenter. *Anais do VIII Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais – IHC 2008,* Porto Alegre/RS, Brasil, p. 296-299, 21 a 24 de outubro 2008, ISBN: 978-85-7669-203-4

Capítulo 1

Introdução

A interação com máquinas passou, ao longo dos anos, por evoluções contínuas com o intuito de facilitar o trabalho do usuário. Houve muito sucesso nessas modificações, comparando os gigantescos computadores operados com cartões perfurados nos anos 1970, controlados por especialistas, com os computadores de mesa acessíveis a pessoas com pouco treino que se tornaram corriqueiros como máquinas pessoais nos anos 1990.

Atualmente o ambiente de trabalho oferecido pelo computador baseia-se na metáfora da mesa de trabalho. A idéia é que a máquina funcione como uma superfície sobre a qual várias ferramentas, tais como caderno, calculadora, rádio e relógio, por exemplo, estão disponíveis. Ao trabalhar com uma mesa, basta escolher a ferramenta que está sobre ela, pegá-la e usá-la. De forma semelhante, no computador, clica-se no ícone de programa e ele já se abre para uso. Como evolução dessa abordagem, pesquisas vêm sendo feitas com o objetivo de facilitar ainda mais o uso de máquinas.

Por que não ser recebido por um personagem que faz as vezes de um assistente em um ambiente de trabalho? Esse agente virtual saudaria o usuário e lhe perguntaria quais tarefas deseja executar. Dependendo das configurações do usuário, poderia logo apresentar as principais notícias do dia ou mesmo fazer a previsão do tempo. Esse tipo de interação é parecido com a que ocorre entre humanos e tem potencial de tornar mais agradável o tempo passado diante da máquina. Chama-se essa nova abordagem de metáfora da conversação face-a-face.

A conversação, afinal, é uma habilidade aprendida desde a infância entre os humanos, sendo bastante intuitiva e uma forma poderosa de humanos interagirem entre si e, conseqüentemente com

suas máquinas. O agente atua então como um intermediário entre sistema e usuário. Pode-se delegar a ele as tarefas que se deseja desempenhar.

Para funcionar como uma interface amigável, fazer uso dos mecanismos de comunicação face-a-face usadas entre humanos é de primordial importância. Não são apenas os movimentos articulatórios da boca, chamados expressões verbais, que passam informações: existem também as formas de expressão chamadas não-verbais, que não são discursivas, mas são relevantes à comunicação. Gestos com braços, expressões faciais e mudanças de postura, que configuram uma comunicação multimodal, são contribuições à transmissão verbal de informações e não apenas uma forma de deixar a animação mais divertida.

Não usar expressões não-verbais em uma animação de um agente virtual, ou usá-las de forma incorreta, pode causar estranhamento e dificultar o entendimento da informação. Para tanto, pesquisas envolvendo tais expressões são fundamentais. Fatores culturais devem ser levados em consideração, afinal em diferentes países alguns gestos têm significados que não coincidem com o de outros. Na Grécia, por exemplo, levantar e abaixar a cabeça significa 'não', enquanto que no Brasil, tem justamente o significado oposto. Também a sincronização de movimentos não-verbais com a fala deve estar correta para permitir um acertado entendimento da informação que se deseja passar e tornar mais claros os pontos importantes de um discurso.

Alguns exemplos de expressões não-verbais são elevação de sobrancelhas, piscar de olhos e movimentos com a cabeça. A aplicação desses movimentos a um personagem virtual pode ser feita através de um script que indica as falas, todos os movimentos que devem ser feitos e o momento exato de sua execução. Outra maneira é usar como entrada apenas as falas, de forma textual ou sonora, para gerar, a partir destas, movimentos sincronizados com determinadas palavras.

Na primeira abordagem, entra-se com o texto a ser lido pelo apresentador e indica-se em que momento será cada piscar de olhos, cada aceno com a cabeça, cada movimento com sobrancelhas, enfim, detalha-se os movimentos não-verbais. A grande desvantagem desse caso é que todos os movimentos que devem ser executados, sem exceção, devem ser descritos pelo script, o que pode tornar trabalhosa e longa a tarefa de escrevê-lo.

Na segunda abordagem, gera-se a animação automaticamente a partir do texto que o personagem irá proferir. Determinadas palavras são mapeadas a movimentos específicos. Para palavras afirma-

tivas, por exemplo, como “sim”, “claro”, “lógico”, a cabeça se move para cima e para baixo para reforçar esta idéia. Nesse caso, não é necessário escrever um script e o trabalho do usuário é muito facilitado devido à automatização. É possível, no entando, que a animação resultante pareça muito artificial, uma vez que não se pode enfatizar palavras diferentes das já mapeadas com movimentos. Um apresentador de notícias, por exemplo, enfatiza números e nomes de cidades. A alteração das configurações do sistema poderia resolver essa dificuldade. Pode-se, por exemplo, indicar que todas as vezes em que aparecer “São Paulo” no texto, deve haver um levantar de sobrancelhas. Ainda assim, mesmo que se adicione uma palavra indicando qual movimento deve ativar, pode ser que não se deseje que esta palavra sempre cause tal movimento. Enfatizar toda vez “São Paulo” faz com que o movimento não-verbal perca o sentido. Enfatiza-se uma palavra quando ela contribui com informações novas [9]. Quando é repetida em outra frase, não se chama mais atenção para ela, como no exemplo “Você veio de onde?” “Vim de *São Paulo*.” “Ah, sei...São Paulo é uma cidade muito rica culturalmente.”

Este trabalho trata do desenvolvimento de um apresentador virtual de notícias controlado pelo usuário através de um *script*. Vale comentar que o termo apresentador virtual de notícias é usado aqui para um agente virtual que transmite notícias de forma não-tendenciosa, usando a fala e expressões não-verbais para enfatizar e tornar mais clara a informação transmitida.

Além do uso de *script*, determinados movimentos associados a um padrão de comportamento identificado como típico de um apresentador de notícias são incorporados automaticamente pelo sistema à animação. Assim o usuário define a fala do personagem e os movimentos que ele deve executar no script e o sistema acrescenta automaticamente certos movimentos, como o piscar de olhos.

Para definir os movimentos inseridos automaticamente, diversas apresentações foram gravadas e um apresentador de notícias real de um canal de televisão foi analisado, de modo a obter informações acerca do seu comportamento. Um exemplo é o piscar de olhos, que ocorre com grande freqüência e, sendo inserido automaticamente na animação, poupa o usuário.

Como resultado do trabalho, foi criado um modelo comportamental para o apresentador de notícias e foi desenvolvido um protótipo de apresentador virtual que faz, além de movimentos labiais, movimentos faciais para reforçar seu discurso, de acordo com um arquivo de entrada dado pelo usuário. A animação é sincronizada com um arquivo de áudio contendo a fala do apresentador.

O uso de personagens virtuais tem um potencial muito grande, seja em ambientes de ensino, de entretenimento ou como interface humano-máquina. Tais aplicações beneficiam-se de avaliações que permitem verificar como agentes conversacionais personificados (*embodied conversational agents* - ECAs) se comparam às estratégias alternativas, por exemplo, se realmente contribuem atuando como assistente no uso de um *software*, complementando um manual em comparação com o uso exlusivo do manual. Avaliações também permitem definir quais as maneiras de implementação e os recursos usados por ECAs geram resultados preferidos pelos usuários.

Neste projeto, foram feitas experiências em que voluntários assistiram apresentações do personagem virtual exibindo diferentes usos de movimentos não-verbais. Além dessas, foram mostradas apresentações reais, obtidas de um canal televisivo de notícias e também apresentações envolvendo apenas áudio. Verificou-se a contribuição dos movimentos-não verbais para a compreensão a notícia dada e para o quesito "naturalidade" da apresentação.

O trabalho foi dividido nas seguintes etapas:

• análise do comportamento de um apresentador de notícias real a partir de gravações feitas de um telejornal. A partir dessa análise e de estudos sobre comportamento humano encontrados na literatura, é possível determinar quais movimentos faciais devem ser implementados e com quais funções.

• desenvolvimento de um apresentador virtual que permite ao usuário entrar com um script que seja seguido pelo apresentador virtual. Além dos comandos descritos pelo script, o apresentador virtual executa alguns movimentos automaticamente, de acordo com regras definidas a partir da análise de comportamento feita previamente. Esta etapa envolveu também a implementação de movimentos faciais.

• definição do formato do *script* de animação. Tal arquivo inclui o texto que o apresentador informará e informações sobre os movimentos que são realizados em sincronia com ele.

• avaliação do sistema por voluntários. Verifica-se, assim, a contribuição dos movimentos faciais à apresentação de notícias pelo personagem virtual.

Este projeto dá continuidade ao trabalho de animação sincronizada com a fala para o português brasileiro desenvolvido por De Martino [16] [17]. Esse algoritmo foi aplicado a um novo modelo de cabeça no qual foram implementados também movimentos não relacionados com a fala (não-verbais): rotação de pescoço, piscar e apertar de olhos, rotação dos globos oculares, elevar de sobrancelhas.

Este trabalho está estruturado da seguinte maneira:

• Capítulo 2: Revisão Bibliográfica - apresenta-se uma revisão detalhada dos estudos que contribuíram para o desenvolvimento do projeto. São abordados trabalhos sobre agentes virtuais, estudos feitos sobre movimentos faciais, os métodos usados para animação facial e diversas linguagens de controle para gerar o script para a animação.

• Capítulo 3: Sistema de animação do apresentador virtual - discute-se o conceito de apresentador de notícias e são expostas as análises feitas do comportamento de um apresentador de notícias real. É feita a descrição da implementação do apresentador virtual, englobando o desenvolvimento da animação facial sincronizada com a fala, a aplicação do modelo de comportamento e o formato do arquivo de controle.

• Capítulo 4: Testes e avaliações - expõem-se os testes feitos com voluntários para validar o sistema e as avaliações dos resultados.

• Capítulo 5: Conclusão - faz-se as colocações finais a respeito do projeto e propõem-se trabalhos que podem complementá-lo futuramente.

Capítulo 2

Revisão Bibliográfica

Tendo começado na década de 1970, os estudos sobre animação facial por computador já evoluíram muito. Os modelos de face animáveis tridimensionais, que começaram como uma representação poligonal rudimentar da cabeça com possibilidades de movimento bastante restritas [47], hoje já propiciam uma sensação de realismo considerável, com movimentações bastante naturais que já não causam mais tanta admiração.

O primeiro filme com animação de humanos fotorealistas, Final Fantasy, de 2001 [70], mostra personagens modelados com grande grau de detalhamento e animados de forma bastante realista, evidenciando a evolução nessa área ao ser comparado com o curta-metragem "Tin Toy" de 1986, considerado inovador a sua época. A figura 2.1 mostra imagens dos dois filmes.

Também a indústria de jogos e simuladores usufrui e contribui para o desenvolvimento de técnicas nessa área [39] [58] [75]. A necessidade de criar *engines*, programas ou conjuntos de bibliotecas que simplificam o desenvolvimento de jogos, gera conhecimento para balancear a demanda dos consumi-

(a) (b)

Fig. 2.1: Filmes inovadores a seu tempo: a) Tin Toy [65] e b) Final Fantasy [66].

dores por jogos mais realistas e a necessidade dos programadores de criar bons produtos com menor custo.

No mundo acadêmico também há investimento em pesquisas relacionadas à animação facial, com os mais diversos objetivos. O projeto europeu NECA (Net Environment for Embodied Emotional Conversational Agents) [34], por exemplo, visa desenvolver um novo conceito de ambiente virtual habitado por agentes virtuais que se comunicam utilizando a fala e sinais não-verbais como movimentos de sobrancelhas e acenos de cabeça. Gerou-se, para avaliação, "the Socialite", uma aplicação web para relacionamentos sociais e "eShowroom", um ambiente de demonstração de produtos para comércio na internet.

Todos esses investimentos em pesquisa e desenvolvimento vêm se somando, buscando, sempre mais, emprestar características humanas aos personagens criados. Tem-se então, como base para este trabalho, um conhecimento acumulado respeitável que inclui diversas possibilidades para criar o modelo de cabeça, formas de aplicar expressões faciais a ele, estudos sobre comportamento humano e de como formalizar isso através de uma linguagem, além de pesquisas sobre interação de personagens virtuais com humanos, como será mostrado a seguir.

O presente capítulo irá apresentar os principais estudos pertinentes ao assunto tratado neste trabalho. Ele está dividido em três grandes blocos: primeiramente será detalhada a questão dos agentes conversacionais, em seguida, aspectos de implementação dos mesmos e, por fim, discute-se a avaliação de tais agentes.

Inicialmente, na seção 2.1, faz-se uma revisão de diferentes abordagens relativas a agentes conversacionais personificados (ECAs - *Embodied Conversational Agents*), classificando-as, na subseção 2.1.1, de acordo com as possibilidades de interação suportadas. Em seguida, na subseção 2.1.2, são abordados estudos sobre comportamento humano que fornecem embasamento para a criação de personagens virtuais convincentes.

A seção 2.2 discute a questão da implementação dos movimentos faciais. A subseção 2.2.1 trata de métodos para animar a face virtual. Na subseção 2.2.2 é apresentada a questão do controle da animação por meio de *scripts*, com exemplos de especificações de animação facial, sendo seguida pela discussão de formalismos para criação de scripts tratada na seção 2.2.3.

Na seção 2.3, comenta-se sobre estratégias de avaliações de sistemas envolvendo ECAs e mostram-

se alguns exemplos ilustrativos. Tais avaliações visam verificar a contribuição dos sistemas e sua eficácia, dado seu objetivo.

Para finalizar, a seção 2.4 expõe as escolhas feitas para o presente projeto, considerando os estudos apresentados anteriormente neste capítulo.

2.1 Agentes conversacionais

Agentes conversacionais personificados (ECAs) são personagens virtuais que procuram reproduzir os mecanismos utilizados na comunicação face-a-face humana: gestos usando mãos e braços, expressões faciais e mudanças na postura, além da fala. Usando ECAs como interface, é, potencialmente, possível facilitar a interação de uma pessoa sem treinamento com uma máquina qualquer, afinal, a interação é a mesma que entre humanos, funcionando de maneira bastante intuitiva.

ECAs também encontram aplicação como avatares, personagens virtuais que representam seus usuários reais em ambientes virtuais, como no sistema Second Life® da Linden Lab [40], que vem ganhando crescente atenção como forma de interação social ou mesmo como sistema de ensino à distância. De fato, há diversas iniciativas nesse sentido que procuram explorar o potencial educacional existente: há uma grande lista de instituições educacionais de vários países com seu espaço no Second Life [72]. No Brasil, o projeto Cidade do Conhecimento 2.0 liderado pela USP [67] explora esse ambiente para o ensino.

As possibilidades de aplicações para ECAs são muitas, podendo envolver um número variado de contextos. O presente trabalho trata especialmente da função de apresentador, que consiste em transmitir informações sem envolvimento emocional usando os recursos disponíveis em um ECA, como a fala, gestos e movimentos faciais.

2.1.1 Cenários de atuação

Sistemas de apresentação de informação podem ser caracterizados pela interatividade de seus personagens com o usuário e pela quantidade de personagens envolvidos [57], podendo ser divididos em quatro cenários, conforme a figura 2.2 (da esquerda para a direita):

1. Apenas um personagem apresenta informações. Bastante similar ao que ocorre quando se as-

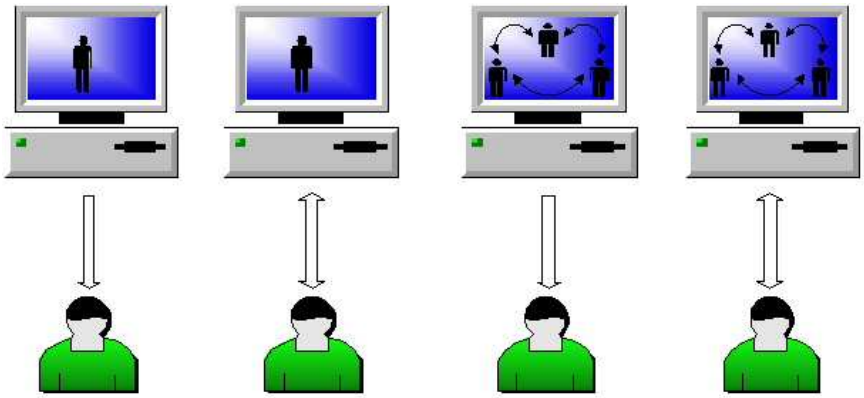


Fig. 2.2: Sistemas de apresentação de acordo com a interação entre usuário e ECAs

siste apresentações de notícias, não há busca de interação com o apresentador quando ele está falando. O PPP Persona [57], por exemplo, envolve um apresentador humanóide caricato que guia usuários através de uma página na internet. Pode ser usado, por exemplo, para ajudar o usuário a escolher seu local de férias em um sítio de turismo. Outro exemplo é o sistema descrito por van Welbergen e seus colaboradores [69], onde um personagem virtual é controlado por um arquivo de entrada que define a fala, os gestos, as interações com uma tela virtual ou com o ambiente virtual, o que o torna apropriado para falar sobre a previsão do tempo ou apresentar uma casa virtual.

2. Tem-se um personagem virtual que se comunica com o usuário, transmitindo e recebendo informações. Um exemplo é Cyberella, uma recepcionista virtual de um sítio da internet que entende um número limitado de questões de visitantes [57]. É possível também que a interação ocorra dentro de um ambiente virtual. Nesse caso, o usuário também assume a posição de um personagem virtual, como no caso em que, um aluno interage com um tutor virtual em um navio [8]. A imersão no ambiente virtual facilita o aprendizado e também o diminui os custos, afinal, o aluno não precisará entrar realmente no navio.

3. Têm-se dois ou mais personagens conversando entre si na tela para passar informação aos espectadores, sem que estes intervenham. O eShowroom [57] mostra dois vendedores conversando entre si com o objetivo de passar ao usuário informações sobre carros. Esse tipo de abordagem pretende tornar mais agradável a experiência do espectador, explorando o meca-

nismo de argumentação e contra-argumentação para expor, explicar e fixar conceitos.

4. Os múltiplos personagens na tela, além de terem a habilidade de conversar entre si, interagem com o usuário. Um exemplo é a plataforma MIAU, uma extensão do eShowroom que permite que o usuário interaja com os personagens [57].

Pode-se ainda considerar mais quatro cenários, semelhantes aos expostos acima, mas envolvendo vários usuários e não apenas um, tal como mostrado na figura 2.3. Nesse caso, o apresentador muda a forma de se dirigir aos espectadores e reconhece que há diferentes interlocutores, nos casos em que há interação. As pesquisas nesse sentido ainda estão começando.

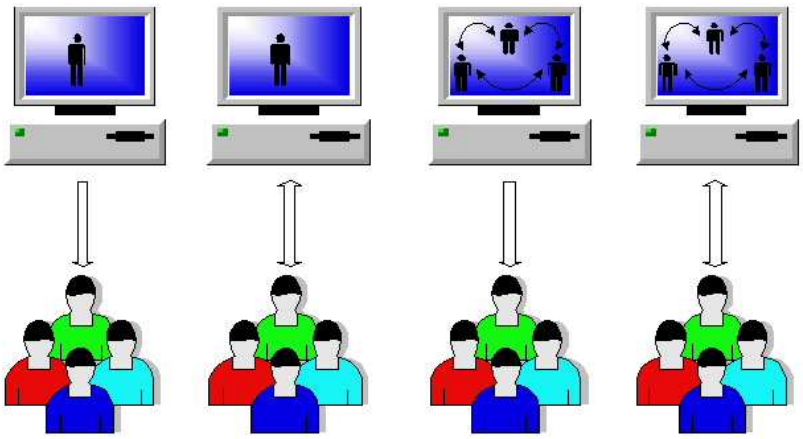


Fig. 2.3: Sistemas de apresentação quando há diversos usuários

Este projeto se ocupa do caso em que não há interação e existe apenas um apresentador, equivalente ao primeiro cenário da figura 2.2 e também da figura 2.3, afinal, um apresentador de notícias se comporta da mesma maneira não dependendo do número de espectadores. Para que ele seja convincente e transmita a sensação de naturalidade, é necessário pesquisar os movimentos não-verbais que acompanham a fala, de forma a aplicá-los ao ECA, garantindo uma maior aceitação por parte do espectador.

O comportamento não-verbal que acompanha a fala consiste em movimentos que não são diretamente associados à produção da fala, mas que podem contribuir para uma melhor comunicação de informações. Abordam-se, a seguir, aspectos do comportamento humano inerentes à comunicação face-a-face.

2.1.2 Estudos do comportamento humano

Diversos estudos apontam que o realismo gráfico do modelo virtual combinado à expressão realista de comportamentos humanos torna um personagem virtual mais convincente, facilitando a interação do usuário com o mesmo [5] [46]. Sinais faciais e corporais acompanhando a fala, além disso, melhoram a compreensão do que está sendo dito bem como o conforto do espectador, que passa a não se ver à frente de um agente robótico/humanóide, mas sim humanizado [55].

Ao ouvir alguém falando, extrai-se da fala um significado e, uma vez que este tenha sido obtido, podem-se descartar as palavras. O mesmo ocorre com gestos, que ilustram o que está sendo dito: freqüentemente os sinais que foram feitos são esquecidos, mas eles auxiliam a guardar a informação transmitida [7].

Há diversos estudos que procuram caracterizar o comportamento humano durante a fala, incluindo movimentos com a cabeça, expressões faciais executadas, mudanças na postura e também gestos realizados com as mãos. Como este trabalho trata apenas da animação da face, não são desenvolvidas aqui considerações acerca dos outros movimentos.

As expressões faciais foram caracterizadas nos seguintes grupos [51]:

- reguladores: regulam a interação orador/ouvinte, e controlam o fluir de um diálogo, ajudando a decidir qual será a próxima pessoa a falar (*turn-taking*). Interromper contato visual com o ouvinte e afastar o olhar em relação a ele são movimentos que auxiliam na interação durante uma conversa [20].

- manipuladores: correspondem às necessidades biológicas, como piscar para umedecer os olhos.

- sinais conversacionais ou ilustradores: enfatizam a fala, tornando-a mais clara. Movimentos de sobrancelhas aparecem freqüentemente como sinais conversacionais [21] [26], bem como movimentos rápidos com a cabeça e piscar de olhos.

- pontuadores: correspondem a ações faciais que ocorrem durante pausas. Essas ações reduzem a ambigüidade da fala, separando ou unindo seqüências de palavras [12]. Piscar de olhos ou movimentos de cabeça, por exemplo, podem acentuar uma pausa.

A região dos olhos é uma das partes mais expressivas do rosto - ao olhar uma imagem humana, o observador passa cerca de 60% do tempo reparando nos olhos e 15% na boca [50]. Além disso, todos os tipos de expressões faciais definidos acima envolvem a região dos olhos. Assim sendo, para emprestar maior realismo à animação, a região dos olhos deve ser trabalhada.

Diversos estudos afirmam que uma pessoa real dificilmente fica com a cabeça e os olhos totalmente parados em situações normais, indicando que a implementação de movimentos randômicos com a cabeça e olhos é uma estratégia fortemente recomendável para tornar mais natural uma animação facial [1] [2]. Para tanto, alguns estudos pertubam a movimentação dos olhos e da cabeça com ruído, como o de Kunc e Kleindienst [36], que usa o ruído de Perlin para gerar movimentos pseudo-randômicos controláveis à cabeça do personagem.

Ekman [21] está entre os primeiros a estudar a importância da movimentação das sobrancelhas em um diálogo. Segundo ele, a movimentação depende do contexto: ele notou que a elevação de sobrancelhas é usada para enfatizar palavras e trechos de frases e também indica que se está fazendo uma pergunta, enquanto que abaixar as sobrancelhas está relacionado a perplexidade e dúvidas. De acordo com Massaro e seus colaboradores [44] a movimentação de sobrancelhas está fortemente correlacionada com a entonação.

Apesar da existência de uma grande quantidade de estudos sobre o uso de sobrancelhas, sua ocorrência como comportamento não-verbal ainda é inconclusiva [33], não tendo sido definido ainda um padrão para sua incidência na fala. Há estudos que afirmam que esse movimento dependeria de emoção [51], enquanto outros defendem que dependeria da introdução de um objeto novo na frase [9]. De qualquer maneira, sabe-se que as sobrancelhas são usadas para enfatizar trechos da fala acompanhando a entonação, apesar de não ser este o único movimento que pode ter esta função.

No estudo relatado por Granström e House [26], pontos-chave na face de dois atores foram monitorados com auxílio de equipamento de captura de movimento durante processo de conversação. A partir dos dados capturados, os movimentos das sobrancelhas e da cabeça foram analisados. Verificou-se a diferença no formato do arco formado pelas sobrancelhas dependendo da emoção apresentada (figura 2.4). Ao acompanhar uma frase de confirmação, as sobrancelhas se elevam um pouco e, ao mostrar alegria e dúvida, elas se elevam ainda mais. Quando há incerteza, as sobrancelhas se abaixam. O formato das sobrancelhas quando se faz uma afirmação sobre algo de que se tem certeza

é muito próximo ao da posição neutra.

A diferença da forma das sobrancelhas dependendo da situação é uma amostra de quão sutis podem ser as modificações da face. Conseguir reproduzir tais sutilezas é uma das belezas e desafios envolvidos na criação de um ECA.

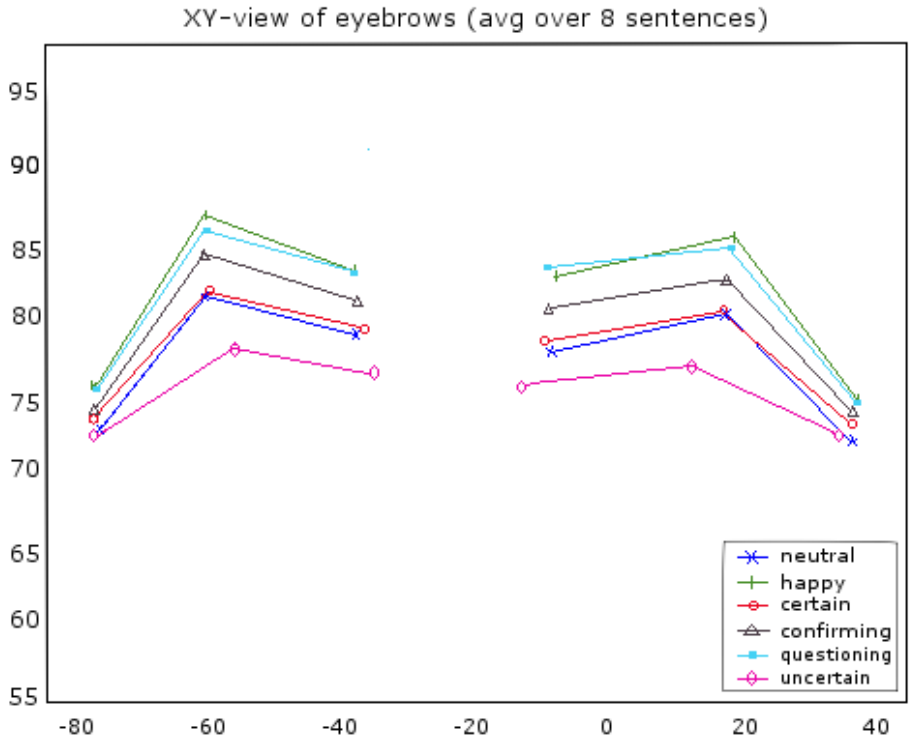


Fig. 2.4: Arqueamento de sobrancelhas (adaptado de [26])

Também se verificou em [26], através do monitoramento de um ponto marcado no nariz, que ao enfatizar uma certa palavra, faz-se um movimento com a cabeça. Esses acenos com a cabeça estão relacionados principalmente com a função de regular a conversação, mas também podem estar relacionados com a ênfase de palavras e respostas afirmativas. Através dos dados analisados foi possível observar que os acenos de cabeça atingem seu deslocamento máximo na sílaba tônica da palavra enfatizada.

A escassez de estudos comparativos ainda não permite determinar em que extensão os movimentos não-verbais que acompanham a fala são independentes de aspectos culturais e regionais. No relato de Krahmer e Swerts [33] tem-se uma comparação entre italiano e holandês, por serem línguas de ori-

gens bastante diferentes, uma de origem latina e outra germânica. Verificou-se através de testes que nos dois casos preferencialmente o movimento de sobrancelhas coincide com palavras de entonação mais forte, apesar de haver diferenças na maneira de colocar a entonação. Mais sobre este teste na seção 2.3 sobre avaliação de sistemas de animações.

2.2 Animação facial por computador

Tendo abordado as possíveis classificações de ECAs em relação às possibilidades de interação e os estudos sobre movimentos que permeiam a comunicação face-a-face, são introduzidos nesta seção conceitos relacionados à implementação de sistemas de animação facial por computador.

Inicialmente, na seção 2.2.1, são abordadas as principais estratégias para a especificação da movimentação e da manipulação do modelo geométrico tridimensional da face. Em seguida, detalham-se os formalismos baseados em script usados para especificar animações de personagens: na seção 2.2.2 abordam-se diferentes especificações existentes para os movimentos faciais e na seção 2.2.3, como criar arquivos que funcionam como roteiro da animação, indicando em que seqüência e com que intensidade os movimentos ocorrem.

2.2.1 Abordagens para realizar animações faciais

Um aspecto de implementação que deve ser discutido é a criação da animação da face. Como criar uma seqüência de movimentos na malha facial para que se tenha a impressão de ver o personagem virtual falando ou mostrando uma variedade de expressões? Há três estratégias para isso [16]:

- sistemas interativos: o animador, usando programas interativos de modelagem e animação, tais como Autodesk® 3DStudioMax®, Autodesk® Maya® e Blender, manipula os vértices de faces virtuais, criando diversas expressões faciais e movimentos na boca, de forma semelhante ao que se faria com um boneco de massa de modelar.

- sistemas baseados em captura de movimento: captura-se e reproduz-se o movimento realizado por um ator. Pode-se obter os dados marcando a face do ator e fazendo a captura com sensores. Por ser derivado de movimentos executados por uma pessoa real, o resultado é muito bom,

apresentando as nuances do comportamento humano. A desvantagem desta estratégia é que todos os movimentos desejados devem ter sido previamente capturados.

- sistemas baseados em script: os movimentos desejados são dados como entrada textual ao sistema, de forma semelhante a um roteiro de teatro. Os comandos dados nesse arquivo ativam a manipulação da geometria facial. Esta estratégia permite a criação de diversas animações diferentes, variando a intensidade dos movimentos e alterando a seqüência deles, por exemplo.

A estratégia escolhida para este trabalho é de sistema baseado em script, que será abordada em maiores detalhes na seção 2.2.3.

Para gerar os movimentos que são chamados pelos comandos dados no *script*, deve-se definir como os vértices do modelo em questão devem ser deslocados. Afinal, é isso que dá a impressão de movimento. Dependendo da aplicação, têm-se necessidades diferentes, então vale considerar as diversas possibilidades. De acordo com De Martino [16], pode-se dividir os métodos para realizar animações faciais nos seguintes:

- interpolação de poses-chave (ou keyframing): cada pose da cabeça é chamada de um quadro-chave e a animação consiste na transição suave entre dois quadros-chave. Cada quadro-chave guarda a posição de todos os vértices que compõem a cabeça naquele instante. Para obter uma transição suave, faz-se uma interpolação suavizada por uma função cosseno, por exemplo. A implementação é bastante simples, mas criar todas as diferentes expressões faciais pode ser muito trabalhoso.

- parametrização geométrica: controla a forma da pele manipulando diretamente a geometria da superfície. As parametrizações são comumente inspiradas por conhecimentos anatômicos: vetores e funções radiais foram usados por Waters [68] para aproximar deformações causadas por musculaturas faciais.

  A parametrização geométrica aplica procedimentos específicos como rotação, interpolação de vértices e translação para obter movimentos faciais e deformações. Isso permite uma maior flexibilidade que *keyframing*, evitando a complexidade de especificar um quadro-chave para cada posição facial [48], [49]. Entretanto, demanda o conhecimento de quais vértices devem ser afetados por cada procedimento e portanto não é portável para modelos faciais diferentes.

Outra forma de controlar regiões da malha da cabeça por parametrização geométrica baseia-se na deformação de superfícies da pele usando *Freeform Deformation* (FFD). FFD é uma técnica que deforma modelos geométricos sólidos e corresponde a envolver a região de interesse com um paralelepípedo cuja deformação afeta diretamente a região contida por ele. Um exemplo pode ser encontrado em um estudo de Kalra e seus colaboradores [29], que busca simular ações musculares através dessa técnica.

- parametrização *data-driven*: recebe como entrada dados coletados da performance de um ator e a animação reproduz então os movimentos que eles descrevem. Os dados utilizados são obtidos através da captura de movimentos da face de um ator usando técnicas como, por exemplo, sensores magnéticos ou ópticos ou gravação de vídeo usando duas câmeras [25].

  Deng e Noh [19] explicam que para gerar animações de um personagem falando usando uma abordagem *data-driven* tem-se a seguinte seqüência. Inicialmente, dados de movimentação facial são gravados e, em seguida, há duas alternativas para trabalhar com eles: abordagem baseada em aprendizado, em que modelos estatísticos para controle de movimentos faciais são treinados usando os dados obtidos e abordagem baseada em amostragem, em que a base de dados de movimentos faciais é organizada e processada. Por fim, dado um novo arquivo de som ou entrada de texto, animações de fala correspondentes são geradas a partir dos modelos estatísticos ou dos quadros de movimento escolhidos da nova base de dados de movimentos faciais.

- modelo biomecânico: modelo que busca reproduzir a estrutura dos tecidos da face [63], [38], [14] definindo camadas: tecido cutâneo, de gordura e a camada de músculos. Os músculos são representados por molas com parâmetros de elasticidade individuais. Inserindo ainda limitantes de preservação de volume é possível obter resultados realistas. Há simplificações desse modelo, que definem apenas duas camadas - a de gordura e a de músculos [31].

  A abordagem de redes de tensão, proposta por Platt [54], simula a face como um conjunto de molas interconectadas. Tem-se como estrutura mais simples um ponto, que fica na pele, no músculo ou no osso. A conexão entre dois pontos é feita por arcos que também têm informação de elasticidade da área entre os dois. Ao aplicar força a um ponto, trata-se a face como uma

rede de molas, unidas nos pontos.

O modelamento dos músculos do pescoço realizado por Lee e Terzopoulos [37] também se baseou na estrutura humana. Incluiu-se a definição de três camadas de músculos e movimentação voluntária e involuntária, resultando em um modelo bastante sofisticado com movimentos realistas.

Antes de abordar questões referentes ao *script* de animação, deve-se tratar da definição dos movimentos que serão usados no mesmo. Afinal, a forma como diferentes regiões do rosto são movimentadas afeta o entendimento das informações que estão sendo transmitidas. São informações relativas ao conteúdo do discurso, assim como acerca do orador, de forma que estas sutilezas devem ser descritas adequadamente no *script*.

Uma pessoa com olhos arregalados pode estar assustada, enquanto que alguém com pálpebras semi-cerradas passa a impressão de estar com sono. A diferença na abertura das pálpebras tem um papel importante para permitir uma compreensão correta da situação da pessoa. Da mesma forma, responder com uma negativa sem acrescentar nenhum movimento à fala é muito menos enfático que abanar a cabeça e elevar as sobrancelhas ao dizer o 'não'.

É necessário definir os parâmetros que podem ser aplicados a cada movimento, de forma que a animação tenha o efeito desejado - qual a abertura das pálpebras, por exemplo. É primordial também assegurar que a sincronização dos movimentos entre si e com a fala esteja correta, caso contrário, o resultado poderá confundir e causar estranhamento no usuário. Na seção seguinte apresentam-se inicialmente diversas abordagens para formalizar os eventos de movimentação facial e na seção 2.2.3, como construir um *script* que descreva uma animação usando tais movimentos.

2.2.2 Sistemas de codificação de movimentos faciais

Para criar uma animação usando *script* é necessário que exista uma definição dos movimentos que podem ser realizados e quais parâmetros eles podem receber. Deve-se especificar também qual composição de movimentos forma uma determinada expressão facial. Há diversas tentativas de descrever as possíveis mudanças na face. Uma das mais utilizadas para animação facial é FACS, que originou outras abordagens.

Facial Action Coding System (FACS) é um sistema definido por Ekman [22], para facilitar estudos psicológicos de expressões humanas que pretende codificar toda mudança visível na face humana. Apesar de não ter sido criado com o propósito de ser utilizado em animação facial, esse sistema foi muito apropriado para esta finalidade, por detalhar os movimentos musculares e definir qual composição deles origina uma determinada expressão facial de emoção.

Foram determinados em FACS os músculos que podem ser controlados voluntariamente e de forma independente, e como eles alteram a aparência da face. Tais mudanças foram mapeadas a unidades de ações (AUs) que compõem as expressões. Não há uma correspondência de um para um entre um músculo e uma unidade de ação: mais de uma unidade de ação pode estar associada a um músculo, pois há casos em que é difícil diferenciar entre um conjunto de músculos com base na aparência.

Montou-se uma tabela que combina as 46 AUs definidas para causar diferentes expressões faciais. Apesar da descrição de FACS ser bastante completa para sobrancelhas, testa e pálpebras, não inclui todas as ações visíveis para a parte inferior da face nem movimentos com a cabeça [24]. Abaixo, na tabela 2.1, alguns exemplos de AUs baseados nos exemplos dados por Parke e Waters [50] e na tabela 2.2, exemplos de como AUs podem ser combinadas para formar expressões de emoção.

AU	*Nome FACS*	*Base Muscular*
1	elevador de sobrancelha interna	Frontalis, Pars Medialis
2	elevador de sobrancelha externa	Frontalis, Pars Lateralis
4	abaixador de sobrancelha	Depressor Glabellae; Depressor Supercilli; Corrugator
5	elevador de pálpebra superior	Levator Palpebrae Superioris
6	elevador de bochecha	Orbicularis Oculi, Pars Orbitalis
7	apertador de pálpebras	Orbicularis Oculi, Pars Palebralis
10	elevador de lábio superior	Levator Labii Superioris, Caput Infraorbitalis
12	elevador do canto do lábio	Zygomatic Major
15	depressor do canto do lábio	Triangularis
17	elevador do queixo	Mentalis
20	esticador de lábio	Risorius
24	pressionador de lábio	Orbicularis Oris
25	separação de lábios	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris

Tab. 2.1: Exemplos de unidades de ação (AUs) do sistema FACS

AMA (*abstract muscle action*) é um modelo baseado em pseudo-músculos [43] que atuam em regiões específicas da face. Para simular a ação de músculos faciais, aplicam-se procedimentos espe-

Emoção	*AUs*
Alegria	6 + 12 + 25
Raiva	4 + 5 + 7 + 15 + 24
Tristeza	1 + 4 + 7 + 15 + 17
Medo	1 + 2 + 4 + 5 + 7 + 20 + 25

Tab. 2.2: Exemplo de combinações de AUs para formar expressões de emoção

cializados controlados por parâmetros. Alguns movimentos integrantes são abertura de boca, fechamento de lábio superior e inferior, compressão de lábios.

Desenvolvido no início dos anos 90 por Prem Kalra [30], *Minimal Perceptible Actions* (MPA) apresenta uma abordagem baseada nas AUs de FACS para representar expressões faciais. Foi definido um total de 65 MPAs para simular efeitos de ações musculares como “abrir boca” e “levantar canto de lábio”. Movimentos musculares não-faciais, que não existiam em FACS, também são definidos, como rotações da cabeça e globos oculares [24].

As MPAs influenciaram fortemente o MPEG-4, um padrão de compressão multimídia que codifica diferentes objetos audiovisuais [52]. Para suporte de animação facial, no MPEG-4, tem-se um modelo da face em estado neutro, pontos-chave na face e um conjunto de parâmetros de animação facial (FAPs). Os FAPs são definidos em termos dos deslocamentos normalizados dos pontos-chave em suas posições neutras. Existem 2 parâmetros de alto nível: visemas e emoções. Os parâmetros de baixo nível são 66 e definem ações isoladas. A figura 2.5 mostra os pontos-chave definidos na face.

As diversas maneiras de codificação de expressões faciais descritas acima podem ser usadas como base para a implementação dos movimentos faciais. Elas indicam quais regiões devem ser alteradas para um movimento facial específico. A implementação pode ser feita usando um modelo parametrizado, por exemplo, em que se cria um modelo que permite a geração de várias expressões faciais com base em um conjunto pequeno de parâmetros de controle. Pseudo-músculos usados no modelo biomecânico mencionado na seção 2.2.1 também podem ser implementados para gerar expressões faciais.

É importante estudar os modelos de codificação de expressões faciais pois são relevantes para a geração de animações por *script*: ao criar um arquivo que indique a seqüência de ações que devem ser executadas, é necessário especificar os movimentos que serão realizados e os modelos descritos acima

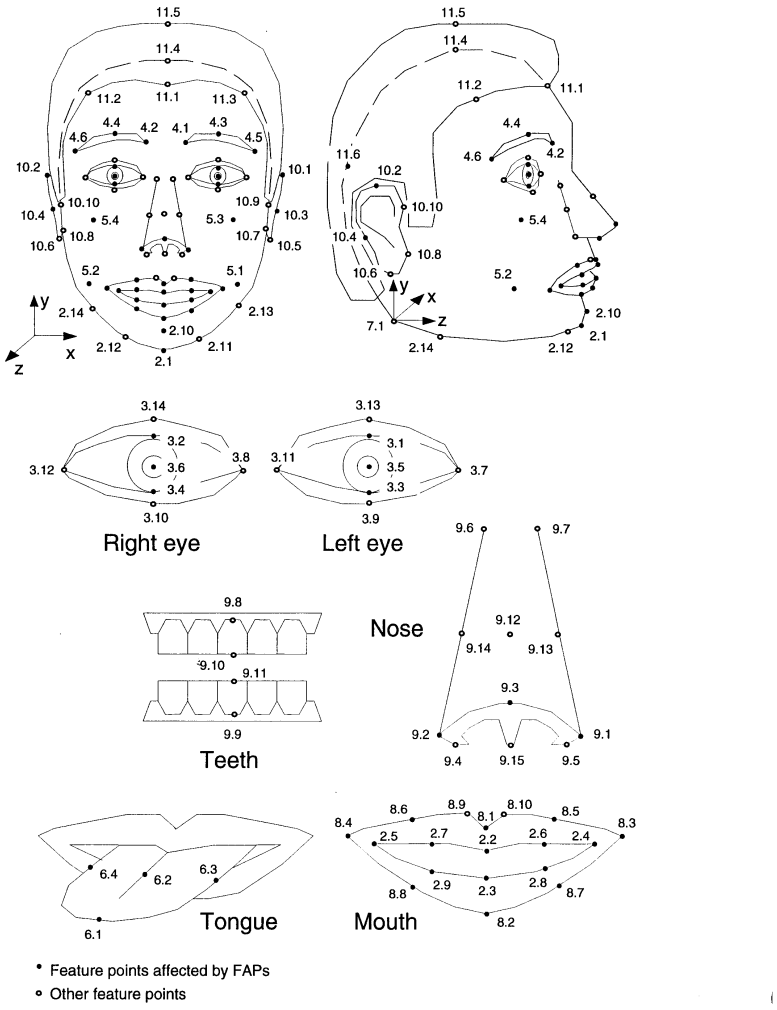


Fig. 2.5: Pontos-chave definidos no padrão MPEG-4 (extraído de [64])

permitem o uso de uma padronização que simplifica a criação do *script*. Além disso, esses modelos orientam a pessoa que cria as animações no modelamento dos movimentos que serão implementados na face virtual.

Para criar as animações faciais, é necessário especificar um formalismo para controlar os movimentos do rosto. Ou seja, definir um roteiro de animação similar a um roteiro de teatro, que indica

o que o ator falará, como deve se mover e mesmo detalhes sobre o cenário, se isto for necessário. Isso equivale a definir, por exemplo, como ativar AUs, MPAs ou FAPs mapeados no rosto para, por exemplo, comandar que o modelo pisque a cada 4.8 segundos ou que levante as sobrancelhas e sorria enquanto fala determinada palavra. Assim, deve ser possível indicar o movimento escolhido, sua duração e mesmo sua intensidade. Para tanto, usam-se comumente linguagens de marcação, uma forma prática de realizar animações e que permite reusabilidade. Tais linguagens são formadas pelo texto que contém as informações que serão faladas e pelas marcas que o descrevem.

2.2.3 Linguagens de marcação para animação facial

Ao escrever um texto, o autor coloca diversas marcações: espaços entre palavras indicam limites entre elas, pontos indicam limites de frases [13]. Esse tipo tradicional de marcação traz maior clareza ao que foi escrito. A marcação não é parte do conteúdo do texto, mas diz algo a respeito. Quando se "traduz" o que foi escrito para a fala (lendo em voz alta), normalmente não se lê as marcações diretamente, mas elas são interpretadas usando diversos gestos paralingüísticos para transmitir a informação apropriada. Um ponto de interrogação, por exemplo pode se tornar um levantar de sobrancelhas e uma alteração na entonação e uma vírgula, uma pausa.

Um exemplo de linguagem de marcação bastante utilizada é HTML (*HyperText Markup Language*), usada para descrever páginas na internet. HTML permite que se formate a informação da página (como texto e figuras, por exemplo) da maneira desejada. A formatação é feita usando marcadores (*tags*), comandos envoltos por "<" e ">". Assim, ao escrever "Texto em itálico" tem-se como resultado "Texto em *itálico*"

Outra linguagem bastante difundida é XML (*eXtensible Markup Language*), uma especificação técnica desenvolvida pela W3C (World Wide Web Consortium - entidade responsável pela definição da área gráfica da internet), para superar limitações do HTML. XML é mais flexível que o HTML, podendo ser estendido para ser usado como base para diversas outras linguagens.

Muitas linguagens baseadas em XML foram desenvolvidas para definir um roteiro para realizar uma animação facial. A seguir, um conjunto delas que permite que se tenha uma visão geral sobre esse tipo de estratégia. Mostram-se também outras abordagens que não usam XML.

MPML - Linguagem de Marcação para Apresentação Multimodal

Multimodal Presentation Markup Language (MPML) é uma linguagem baseada em XML, desenvolvida para facilitar a descrição de apresentações multimodais na internet usando personagens [28]. Para especificar o comportamento dos personagens pode-se usar os seguintes marcadores:

- <speak> - define a fala do agente,
- <play> - define a gesticulação que será executada pelo agente ou apresenta a animação,
- <move> - define o movimento do personagem na tela.

MPML provê funcionalidades para sincronização de mídia (baseada em SMIL - *Synchronized Multimedia Integration Language*) e interação simples, usando um elemento <listen>.

A versão MPML 2.0e [73] provê funções de expressão de emoções, usando o marcador <emotion>. Isso altera parâmetros de voz (*pitch*, volume, velocidade e ênfase em certas palavras) e as ações do personagem, mas não permite especificar quais partes do rosto serão afetadas.

A seguir, um exemplo extraído do relato de Zong e seus colaboradores [74]:

```
<mpml>
   <head>
      <title> MPML Presentation </title>
      <agent id=''PD'' character=''peedy''>
   </head>
   <body>
      <page id=''first'' ref=''self_intro.html''>
         <emotion type=''pride''>
            <speak>
               My name is Zong Yuan,
               I am from Tokyo University.
            </speak>
         </emotion>
      </page>
   </body>
</mpml>
```

O resultado é o personagem Peedy dizendo a fala delimitada pelos marcadores <speak> e </speak> mostrando orgulho, que foi a emoção especificada em <emotion>.

Essa abordagem é bastante interessante para criar uma apresentação com diversos personagens, envolvendo especificação de quem está falando e quais suas emoções. As necessidades para o apresentador virtual são outras, pois se trata de apenas um personagem e é preciso que seja possível definir movimentos sutis para trechos da fala e não para a fala inteira, que é o que o marcador de emoção faz nesse caso.

AML - Linguagem de Marcação para Avatar

Avatar Markup Language (AML) [35] permite a animação sincronizada de movimentos corporais e faciais e de fala (a partir de um sistema TTS - de texto para fala). Foram usadas FAPs (MPEG-4) como parâmetros de baixo nível para a animação, mas é possível usar outro tipo de definição de baixo nível para o rosto.

A sublinguagem relativa à animação facial da AML, *Avatar Face Markup Language* (AFML) permite que diversos eventos sejam definidos de forma que se pode ter controle sobre partes distintas do rosto de acordo com os parâmetros de baixo nível escolhidos. No caso, a base de dados FAP contém as expressões disponíveis para animação. É possível aumentar essa base de dados, incluindo novos elementos.

O marcador <ExpressionsTrack> permite que se definam diversas expressões passando como parâmetros o momento de início do movimento, nome do arquivo de expressão (da base de dados) e o envelope de tempo. O arquivo de expressão, indicado por <ExName>, contém a expressão estática da base de dados de FAPs que o agente deve apresentar. Em <Shape>, define-se qual o modo de interpolação dos pontos faciais no momento anterior ao corrente, sendo que este pode ser logarítmico, exponencial ou linear e é limitado por <Point>. No marcador <Point> também são definidas a duração da transição e a intensidade do movimento, afinal é possível que o usuário não deseje um sorriso completo (de intensidade 1), mas apenas um esboço de sorriso.

AML permite que se defina detalhadamente como se dará a animação, o que traz um aumento na complexidade ao escrever o código. Abaixo, um exemplo [35], em que o agente virtual leva 0.5 segundos para sorrir com intensidade 1 e retorna à expressão anterior em 0.8 segundos realizando as transições de expressão com interpolação logarítmica:

```
<ExpressionsTrack name="Emotion">
   <Expression>
      <StartTime>00:00:000</StartTime>
      <ExName>smile.ex</ExName>
         <Envelope>
            <Point>
               <Shape>log</Shape>
               <Duration>0.5<Duration>
               <Int>1<Int>
            </Point>
            <Point>
               <Shape>log</Shape>
               <Duration>0.8<Duration>
               <Int>0<Int>
            </Point>
         </Envelope>
```

```
   </Expression>
</ExpressionsTrack>
```

AML provê um mecanismo genérico para a seleção e concatenação sincronizada de animações. Além disso, provê flexibilidade para animadores acrescentarem expressões à base de dados de FAPs para serem acessadas com AFML.

Para a aplicação ao apresentador virtual, entretanto, AML não é adequada, pois dificulta a marcação de palavras individuais, uma vez que toma como entrada um texto completo e não permite que o usuário tenha acesso a cada fonema separadamente.

CML - Linguagem de Marcação para Personagens

Character Markup Language (CML) [3], tal como AML, permite que se tenha grande controle sobre a sincronização de ações e narração. É tanto uma linguagem de representação quanto de *scripting*, por representar atributos de agentes virtuais e ter um script da animação de seu comportamento. Assim, atributos como personalidade, estado corrente e papel do agente virtual influenciam seu comportamento.

Existe uma sublinguagem referente ao movimento do corpo (CBML - *Character Body Markup Language*) e outra referente à face (CFML - *Character Face Markup Language*). Essa última é um conjunto de marcadores de baixo-nível baseado em FAPs MPEG-4 que definem representações de expressões específicas. Os elementos CFML básicos são: movimento da cabeça (inclinar, virar), gestos de cabeça (icônico ou dítico), movimento de sobrancelhas, boca, olhar e expressão.

CML define atributos de alto nível para a personalidade do personagem, emoção e comportamento que estão integrados para formar a especificação de script de animação sincronizado. Comportamentos novos podem ser criados combinando os elementos básicos mencionados acima e seus atributos, trazendo maior flexibilidade para geração de *scripts* de animação.

Esta linguagem apresenta a mesma limitação que a AML mencionada anteriormente, não permitindo aplicar movimentos verbais a palavras específicas com facilidade. Além disso, para o presente estudo envolvendo um apresentador, os atributos que o definem não são relevantes: pretende-se primeiramente criar um apresentador neutro, sem personalidade definida ou emoções que afetem sua performance.

ECAF - Encenação de Agentes Conversacionais Personificados

Embodied Conversational Agent Facade (ECAF) [36] é um *framework* que foi criado para processamento em tempo real, reagindo assim a comportamentos não previstos durante o *design* da animação.

A linguagem usada para descrever a animação é baseada em XML, sendo composta por dois principais comandos: <act>, que descreve as ações que serão executadas pelo ECA e <speak>, que trata da fala. Para complementar, aplicam-se movimentos aleatórios à cabeça de modo a emprestar maior realismo à animação.

Os comandos suportados por <act> são:

- <text>, o texto que será mostrado na tela
- <stop>, que interrompe a animação e faz a cabeça virtual voltar à sua posição neutra,
- <start_capture> e <stop_capture> que iniciam e terminam a captura de vídeo e áudio
- <gesture>, trata de diversos canais relativos à animação incluindo os movimentos que podem ser feitos e as alterações no fundo da tela.

O comando <speak> recebe o texto que será falado pelo agente virtual. Isso pode ser feito de duas formas: caso o texto venha de um arquivo de som, os fonemas e seus momentos de ocorrência são passados como informação. Se a fala for gerada por um sistema TTS, coloca-se entre os marcadores apenas o texto que será falado. Abaixo, exemplos que ilustram as duas situações, adaptadas do relato de Kunc e Kleindienst [36]:

```
<ecaf>
   <speak src="oi\_mundo.wav">
      <phoneme time="0.611">O</phoneme>
      <phoneme time="0.787">i</phoneme>
      <gesture head_angle="15">
         <phoneme time="0.845">m</phoneme>
         <phoneme time="0.925">u</phoneme>
         <phoneme time="1.06">n</phoneme>
         <phoneme time="1.170">d</phoneme>
         <phoneme time="1.252">o</phoneme>
      </gesture>
   </speak>
</ecaf>

<ecaf>
<speak>Boa tarde!</speak>
<speak>Como vai?<gesture expr="smile" expr_scale="0.8">
Eu vou bem.</gesture>Agora vou virar meus olhos e cabeça
```

```
<gesture eye_vert="20"> Olho para
<gesture head_angle="-25" persistent="true" />cima.</gesture>
</speak>
</ecaf>
```

Apesar de o objetivo deste projeto não ser realizar animações em tempo real, o formato usado por ECAF é bastante apropriado para o apresentador virtual, apesar de poder ser simplificado. Tem-se a definição do texto a ser apresentado oralmente e a movimentação a ser realizada em conjunto com o mesmo, cujos parâmetros podem ser alterados de acordo com as especificações do animador.

VHML - Linguagem de Marcação para Humanos Virtuais

Virtual Human Markup Language (VHML) [15] é uma linguagem baseada em XML para animar um personagem virtual por inteiro, incluindo produção de fala, animação de rosto e corpo, representação emocional e gerenciamento de diálogos, entre outros.

VHML contém sublinguagens com diversos focos, como por exemplo EML (*Emotion Markup Language*) para emoções, GML (*Gesture Markup Language*) para gestos, SML (*Speech Markup Language*) para a fala e FAML (*Facial Animation Markup Language*) para animação facial. Vale comentar que há linguagens que "herdam" atributos de outras, como FAML que recebe características de EML e GML. Para este projeto interessam em especial EML e FAML, que tratam das expressões faciais.

Os parâmetros padrão que podem ser atribuídos aos movimentos da FAML e da EML são:

- duração do movimento,
- intensidade do movimento,
- demarcação, pode ser usado para setar uma marca arbitrária em algum lugar do texto, para reportar que uma dada posição foi atingida,
- tempo de espera, que indica a pausa entre o início de um movimento e o próximo, permitindo a sincronização com a fala.

A tabela 2.3 apresenta os 20 movimentos definidos na FAML.

look	*eyes*	*head*	*head-roll*	*eyebrow*	*blink*	*wink*	*jaw*
-left	-left	-left	-left				close-
-right	-right	-right	-right				open-
-up	-up	-up		-up			
-down	-down	-down		-down			

Tab. 2.3: Elementos em FAML

Os elementos em EML indicam a emoção a ser apresentada pelo humano virtual e afetam a voz, a face e o corpo. As emoções definidas na EML são as seguintes: medo, raiva, confusão, estupefação, nojo, alegria, neutro, tristeza, surpresa.

Exemplo:

```
<look-left duration=‘‘2000ms’’ wait=‘‘1500ms’’/>
<eye-blink duration=‘‘40ms’’>
  Bom dia!
</eye-blink>
```

Neste caso, o personagem olha para a esquerda por 2 segundos, sendo que após 1,5 segundos ele pisca e fala “Bom dia!”.

Esta linguagem é bastante adequada para os propósitos do apresentador desenvolvido aqui, pois permite a definição dos movimentos que devem ser executados e sua intensidade, permitindo que se especifique o momento da fala em que devem ocorrer, mas seria interessante ter mais controle sobre a sincronia do movimento, com fonemas e não apenas com palavras inteiras.

Animação facial a partir de texto

Há diversas abordagens que criam animações diretamente a partir do texto a ser falado ou do arquivo de som que será usado como fala. A análise do conteúdo do texto e da variação de entonação do arquivo de som são de extrema importância nesse caso.

BEAT (*The Behavior Expression Animation Toolkit*) [9] é um sistema composto por módulos com as seguintes funções:

- analisar e marcar linguagem,
- sugerir comportamentos,

- selecionar comportamentos,
- criar a seqüência de comportamento a ser animada.

Na figura 2.6, uma ilustração simplificada do sistema.

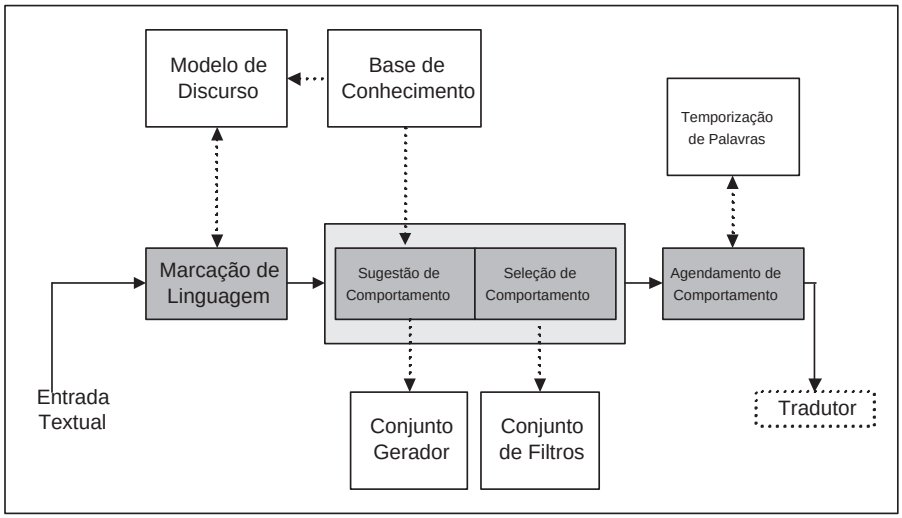


Fig. 2.6: Arquitetura do sistema BEAT (figura adaptada de [9])

O sistema recebe como entrada um texto que é analisado. A partir dessa análise, criam-se sugestões de gestos a executar e, a seguir, usando critérios de seleção como prioridade e valores limite, escolhem-se os gestos que serão efetivamente aplicados na animação. Todo esse processo é feito automaticamente, usando XML como estrutura de dados principal.

Outra linguagem baseada em XML que cria uma anotação ao texto para que depois os comportamentos sejam selecionados é RRL (*Rich Representation Language*) [53], desenvolvida para o projeto NECA [34]. Usa-se entretanto uma notação mais abstrata para a determinação do conteúdo das cenas. A seqüência de módulos é a seguinte:

1. Gerador de Cena,
2. Gerador de Linguagem natural multimodal,
3. Componente de síntese de fala,

4. Componente designador de gestos,

5. Tocador de mídia.

No estudo de Albrecht e seus colaboradores [1] o texto pode ser complementado com *emoticons*, combinações de caracteres que representam expressões faciais. Esse texto é analisado e automaticamente transformado em script baseado em XML, que traz marcações causadas pela análise do texto acrescidas das emoções passadas pelos emoticons. Abaixo, alguns exemplos de movimentos inseridos automaticamente a partir da análise do texto:

- piscar de olhos: é usado como pontuador durante pausas para estruturar o que está sendo dito. Além disso, é inserido a cada 4.8 segundos com função biológica de umedecer os olhos;

- questões: quando se usa o ponto de interrogação para indicar uma pergunta, os olhos do personagem se fixam no usuário e as sobrancelhas e a cabeça se elevam ligeiramente;

- expressões associadas à entonação: usando dados sobre entonação, a cabeça se levanta quando há máximo local de *pitch*. Para evitar movimento repetitivo, a cabeça é girada randomicamente para os lados em 75% de todos os movimentos de cabeça. Movimentos de sobrancelha nos casos associados à entonação também ocorrem aleatoriamente.

Estes métodos que criam de forma automática a animação a partir do texto pronunciado acabam não sendo interessantes para o controle de um apresentador virtual, pois reduzem drasticamente o controle da animação e o resultado final se torna fortemente dependente de inflexível análise de texto. Para um apresentador de notícias é desejável que se possa definir quais palavras enfatizar e com qual ênfase.

Para completar a discussão sobre *scripts* de animação, seguem explicações sobre linguagens que não são baseadas em XML, como o MIMIC e o STEP.

MIMIC

MIMIC [23] é implementado como um compilador, que lê um programa MIMIC e gera valores de parâmetros de animação. Pode ser usado com qualquer sistema de animação que use parâmetros

de animação variantes no tempo: FAPs do MPEG-4, AUs do sistema FACS ou valores de contração de músculos, por exemplo. Na figura 2.7, a gramática do MIMIC.

program	::=	def_list MAIN < parameter_list > actiongroup
	|	def_list MAIN actiongroup
def_list	::=	ϵ
	|	def_list def
def	::=	DEF IDENTIFIER actiongroup
	|	DEFXPR IDENTIFIER FNAME
actiongroup	::=	sequence
	|	parallel
sequence	::=	{ action_list }
parallel	::=	[action_list]
action_list	::=	action
	|	action_list action
action	::=	IDENTIFIER < parameter_list > (timevalue_list)
	|	IDENTIFIER (timevalue_list)
	|	IDENTIFIER < parameter_list > ;
	|	IDENTIFIER ;
	|	< parameter_list > actiongroup
	|	actiongroup
parameter_list	::=	parameter
	|	parameter_list , parameter
parameter	::=	pname
	|	pname_arg = DECIMAL
	|	pname_arg = IDENTIFIER
pname	::=	LEFT | RIGHT | LINEAR | SIN | LOG | EXP
	|	AVERAGE | ADD | OVERWRITE
pname_arg	::=	LEFT | RIGHT | IN | HOLD | OUT | VALUE | CONNECT
	|	DURATION | FREQ | PERIOD | MIN | MAX
timevalue_list	::=	time_list ; value_list ;
time_list	::=	ϵ
	|	time_list DECIMAL
value_list	::=	ϵ
	|	value_list DECIMAL

Fig. 2.7: Gramática do MIMIC

Esse compilador permite que o usuário especifique as ações e se elas ocorrem em paralelo ou seqüencialmente. Há diretivas adicionais que controlam o alinhamento temporal de uma ação em relação a outra. Todas as restrições temporais são resolvidas automaticamente pelo compilador MIMIC. Na figura 2.8, um exemplo que ilustra esta abordagem.

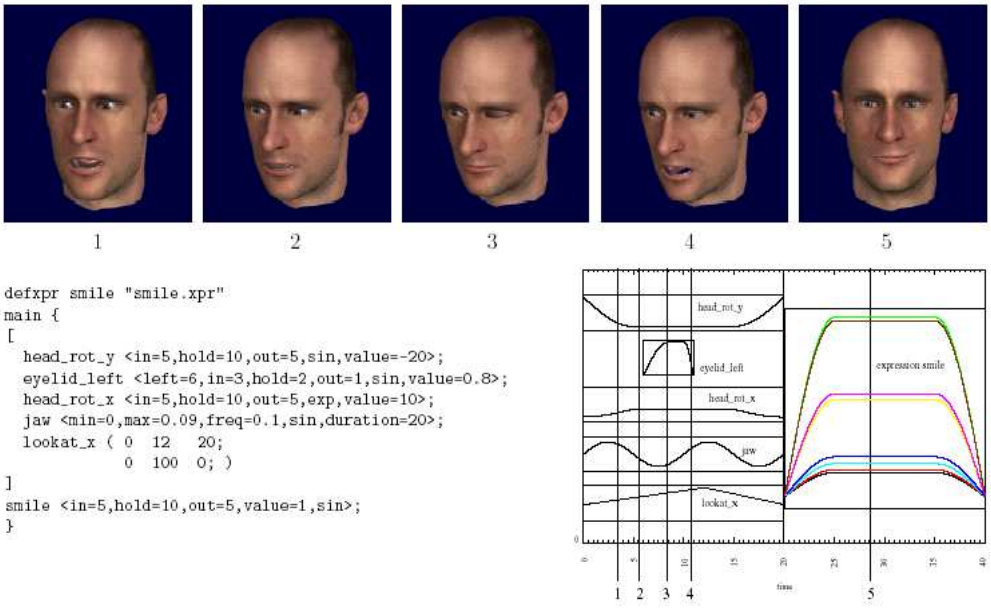


Fig. 2.8: Exemplo de seqüência de animação gerada por um programa MIMIC

STEP - Linguagem de Script para Agentes Personificados

Apesar de não tratar da face, vale comentar a linguagem STEP (*Scripting Language for Embodied Agents*) [27] que poderia ser adaptada para a face. STEP mapeia o corpo humano em nós de articulações, segmentos (parte do corpo associada à junta) e regiões (definem locais relativos ao segmento). Esses podem ser movimentados, configurando ações, sendo que as ações básicas são 'mover' e 'girar'. É possível definir diferentes ações fazendo composições de ações usando um dos operadores abaixo:

- em paralelo par([Ação$_1$, ...Ação$_N$]), as ações são executadas simultaneamente;
- em seqüência seq([Ação$_1$, ...Ação$_N$]), as ações são executadas em seqüência;
- repetições repeat(Ação, N), a ação é repetida N vezes;
- randômicas choice([Ação$_1$, ...Ação$_N$]) uma das ações especificadas é executada.

Pode haver interação com o mundo externo ou interno (como personalidade) do agente, através de comandos que verificam um certo estado:

- *teste*, testa um estado específico, se for verdadeiro, prossegue, senão, falha;
- *execução*, torna um estado verdadeiro;
- *condicionalidade*, realiza uma ação, se determinado estado for verdadeiro e realiza outra, se o estado for falso;
- *repetição*, repete uma ação enquanto um certo estado for verdadeiro.

A seguir, um exemplo para fazer personagem andar sem sair do lugar, através da movimentação coordenada de pernas e braços:

```
script(walk\_pose(Agent), Action):-
Action = seq(par([
  turn(Agent, r_shoulder, back_down2, fast),
  turn(Agent, r_hip, front_down2, fast),
  turn(Agent, l_shoulder, front_down2, fast),
  turn(Agent, l_hip, back_down2, fast)]),
           par([
  turn(Agent, l_shoulder, back_down2, fast),
  turn(Agent, l_hip, front_down2, fast),
  turn(Agent, r_shoulder, front_down2, fast),
  turn(Agent, r_hip, back_down2, fast)])
  )
```

Foi desenvolvido posteriormente o XSTEP, o STEP baseado em XML. O XSTEP é usado principalmente para construir uma biblioteca de gestos e ações.

A seguir, o mesmo exemplo mostrado acima, usando XSTEP:

```
<action name="walk(Agent)"><seq><par>
  <turn actor="Agent" part="r_shoulder">
   <dir value="back_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="r_hip">
   <dir value="front_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="l_shoulder">
   <dir value="front_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="l_hip">
   <dir value="back_down2"/>
   <speed value="fast"/></turn></par><par>
  <turn actor="Agent" part="l_shoulder">
   <dir value="back_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="l_hip">
   <dir value="front_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="r_shoulder">
   <dir value="front_down2"/>
   <speed value="fast"/></turn>
  <turn actor="Agent" part="r_hip">
   <dir value="back_down2"/>
   <speed value="fast"/></turn>
  </par></seq></action>
```

Em busca de uma padronização

A escolha de um ou outro formalismo depende muito da aplicação e muitos grupos foram, ao longo do tempo, desenvolvendo linguagens mais convenientes para seus projetos. Diversas dessas linguagens possuem, entretanto, similaridades, de forma que uma unificação de linguagens é bastante interessante. Kopp e seus colaboradores [32] relatam o início de esforço promissor buscando uma padronização internacional, o framework SAIBA (*Situation, Agent, Intention, Behavior, Animation*).

SAIBA possui três estágios, 1) planejamento de intenção, 2) planejamento de comportamento e 3) realização de comportamento. Para intermediar os dois primeiros estágios, criou-se a Linguagem de Marcação de Função (*Function Markup Language* - FML), que descreve o objetivo expressivo e comunicativo sem referências ao comportamento físico. Essa linguagem descreve aspectos relevantes para o planejamento de comportamento verbal e não-verbal, devendo definir as unidades semânticas básicas associadas com um evento comunicativo e permitir a anotação dessas unidades que descrevem a função comunicativa, tais como funções expressivas e afetivas.

Para intermediar os dois últimos estágios, a Linguagem de Marcação de Comportamento (*Behavior Markup Language* - BML) descreve detalhadamente os comportamentos multimodais que serão executados no estágio final, incluindo questões de sincronização.

O desenvolvimento de uma linguagem unificada para criar animações de ECAs será de grande valia para a comunidade internacional, pois oferece uma estrutura mais portável e robusta de animação de personagens.

Por fim, com a animação pronta, é interessante verificar quão eficaz ela é, dado um objetivo traçado. Pode ser que se deseje que o agente virtual explique o funcionamento de um equipamento ou que ele informe os procedimentos de emergência em um avião. Na seção seguinte discorre-se sobre esse assunto, apresentando alguns testes realizados por dois grupos de pesquisa.

2.3 Elaboração de testes de avaliação

Em 2000, após citar suas motivações para desenvolver um ECA, Cassell e seus colaboradores [7] afirmam, “Nossas motivações são expressas na forma de “crenças”, pois, até hoje não existe plataforma de agente conversacional personificado para testar essas alegações. Só agora que existem

implementações de ECAs “conversacionalmente inteligentes”, podemos nos voltar à avaliação de suas habilidades.” (tradução livre). Desde então, cada vez mais testes vêm sendo realizados para comprovar a aposta de que a interação humano-máquina realmente é simplificada com tal interface e que usar ECAs para diversos fins pode ser benéfico.

Christoph [11] sugere um guia sobre a realização de avaliações de ECAs. Trata-se do processo de realização de testes e avaliações, recomendando uma série de procedimentos:

- criação de um modelo de pesquisa que descreva os objetivos do estudo, as questões que se tenta responder;
- identificação e definição dos fenômenos que serão estudados;
- escolha da estratégia de pesquisa e métodos de obtenção de dados;
- escolha do tipo de voluntários que participarão nos testes;
- determinação a forma de análise estatística a usar;
- realização de mais de uma forma de teste para aumentar a confiabilidade dos resultados;
- realização de pré-testes para verificar que os voluntários entendem os termos usados e o que está sendo pedido.

A seguir, a descrição de dois testes de ECAs para ilustrar a realização de avaliações. Ambos envolvem verificar a contribuição de sinais não-verbais à fala. No primeiro, usam-se gestos de braços, enquanto que, no segundo, movimento de elevação de sobrancelhas.

2.3.1 Relatos de testes de avaliação

Buisine e seus colaboradores [6], descrevem a avaliação feita de agentes usando gestos de braços para complementar informações. O teste consistiu em apresentar agentes virtuais explicando o funcionamento de equipamentos (máquina copiadora, controle remoto e um programa de computador). Uma vez que a explicação envolve muitas informações sobre cor, formato e localização de botões, a existência de gestos com braços é relevante. Os gestos mostrados são díticos, que designam algo no

espaço (apontar para um objeto, por exemplo) e icônicos, que capturam um aspecto do objeto falado (como representar certo objeto com as mãos).

Apresentações curtas (60 a 75 segundos) usando três estratégias são mostradas aos voluntários:

1. redundância - o agente transmite as informações usando gestos e fala;

2. complementaridade - o agente transmite algumas informações usando gestos e outras com fala;

3. especialização da fala - as informações são dadas apenas usando a fala. Em cada apresentação, o personagem usado aparecia junto ao equipamento em foco, conforme mostrado na figura 2.9.

Atribuiu-e randomicamente cada estratégia a três personagens diferentes para que a aparência do personagem não influenciasse o resultado. Assim, cada um dos personagens apresentou cada objeto usando uma dada estratégia o mesmo número de vezes.

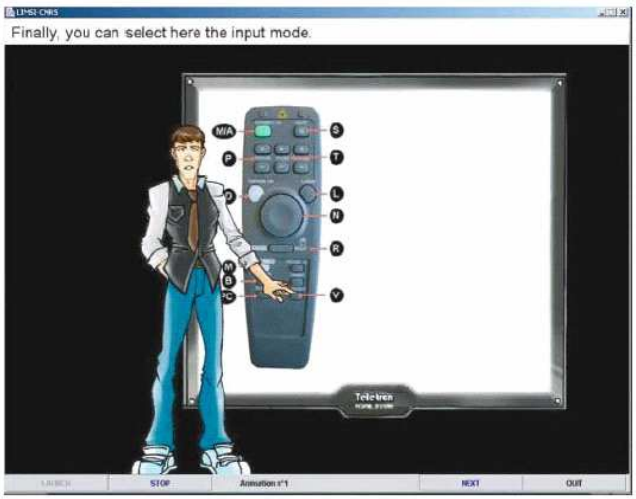


Fig. 2.9: Personagem apresentando controle remoto (imagem extraída de [6])

Cada um dos 18 participantes (9 homens e 9 mulheres) preencheu um questionário em que os ECAs eram avaliados com as seguintes questões: “qual ECA deu a melhor explicação?”, “em qual ECA você confia mais?”, “qual ECA é o mais amigável?”, “os ECAs tinham a mesma personalidade?”, “qual ECA é o mais expressivo?”. Além disso, verificou-se a persistência das informações dadas nas animações: figuras dos equipamentos explicados nos experimentos foram mostradas e o voluntário tentava se lembrar do máximo de informações possíveis.

Os resultados obtidos mostraram que as estratégias que usam gestos (tanto de forma complementar quanto redundante) tiveram maior pontuação no quesito “qualidade de explicação” e “confiança” do que a estratégia em que a transmissão de informação se dava apenas pela fala. Isso ocorreu, entretanto, apenas para os voluntários masculinos, pois para o grupo feminino, tal diferença não foi significativa. O quesito “amigável” mostrou uma relação com a aparência dos personagens e não com as estratégias usadas. Não houve efeito da estratégia usada ou da aparência dos ECAs nos quesitos “personalidade” e “expressividade”.

A verificação de persistência de informações foi bastante dependente do gênero, sendo que as mulheres se lembraram de uma quantidade ligeiramente maior de informações que os homens. A estratégia de apresentação usada não influenciou os resultados, entretanto, afirma-se que este fato não reduz a importância das estratégias multimodais: variáveis subjetivas continuam sendo cruciais para a pessoa se envolver com o ECA e determinam, até certo ponto, o sucesso de tais ferramentas multimídia.

Outro exemplo pertinente a avaliações de ECAs pode ser encontrado no relato de Granström e House [26], que avalia como as sobrancelhas influenciam a percepção do foco de uma informação (ou seja, qual informação se deseja enfatizar) e a diferença desta influência em duas línguas bastante diferentes, italiano e holandês.

Os experimentos envolveram uma cabeça virtual falando uma frase simples como “quadrado azul”. Para obter uma fala convincente, duas pessoas (A e B) seguiram um procedimento que envolvia virar cartas de uma pilha e falar as figuras geométricas nelas contidas e sua cor. Ao virar uma carta com elementos diferentes dos anteriores, o elemento diferente recebe uma entonação mais forte. Os contextos de falas existentes são 3. No primeiro contexto o elemento e a cor de uma carta para a próxima são diferentes, de forma que há contraste em ambas as palavras (CC). No segundo contexto, a primeira palavra já foi dada anteriormente e a segunda é contrastante (DC). O terceiro contexto é o inverso, sendo a primeira palavra contrastante de uma carta para a próxima (CD). A tabela 2.4 ilustra o conjunto de contextos, sendo que A indica a primeira carta descrita e B a segunda.

Especialistas analisaram as falas quanto à mudança de entonação. Verificou-se que, em holandês, há uma maior entonação nas palavras em foco (palavras contrastantes de uma carta para outra). Para o italiano, toda palavra é enfatizada, sendo que a primeira recebe uma ênfase maior que a segunda.

Contexto	*Holandês*	*Italiano*
CC	A: rode driehoek B: blawe vierkant	A: rettangolo rosa B: triangolo nero
DC	A: blawe driehoek B: blawe vierkant	A: triangolo rosa B: triangolo nero
CD	A: rode vierkant B: blawe vierkant	A: rettangolo nero B: triangolo nero

Tab. 2.4: Exemplo do conjunto de falas em holandês e italiano (extraído de [26])

Assim, a distribuição de ênfase para o italiano não depende do contexto, enquanto que para o holandês sim.

O primeiro experimento consistiu em mostrar o personagem virtual falando uma frase para os voluntários com as situações CC, DC e CD. Todas as situações tinham duas possibilidades: movimento de sobrancelhas na primeira palavra ou na segunda. Os voluntários deviam então marcar a animação mais natural do par. Os resultados mostram que os holandeses preferem o movimento de sobrancelha coincidindo com a palavra que recebe maior ênfase, que é a palavra que indica contraste. Os italianos preferem que o movimento coincida sempre com a primeira palavra, que é sempre a mais enfatizada. Assim, ambos os grupos preferem que a elevação de sobrancelhas coincinda com a palavra que recebe maior ênfase na fala, sendo que para o holandês isso depende do contexto e para o italiano, não.

O segundo experimento consistiu em mostrar aos voluntários uma animação sem movimentos de sobrancelhas e outra com movimento de sobrancelhas acompanhando a palavra enfatizada da seqüência. Definida uma palavra de interesse, os voluntários deveriam definir qual das duas animações evidenciava mais esta palavra. Verificou-se tanto para o italiano quanto para o holandês que a presença de um movimento de sobrancelhas aumenta a percepção de ênfase dada à palavra com maior entonação. Quando a palavra analisada não era a enfatizada pelo movimento de sobrancelhas, os participantes escolheram a animação sem movimento como sendo a que enfatiza a dada palavra.

Por fim, no terceiro experimento, verifica-se como os voluntários usam dicas audiovisuais ao interpretar as falas. Os voluntários ouvem e vêem o personagem virtual falando uma expressão correspondente a “quadrado azul” com certa entonação e elevação de sobrancelhas na primeira palavra e depois, na segunda. Ou seja, mesmo enfatizando “azul” com a voz por exemplo, em uma animação, o movimento de sobrancelhas ocorre com “quadrado” e em outra, com “azul”. Os participantes devem

então decidir qual expressão teria sido dita anteriormente, 1) quadrado vermelho (DC), 2) triângulo azul (CD) ou 3) triângulo vermelho (CC). Verificou-se que, para o holandês, nos casos de DC e CD, os participantes percebem o foco na palavra enfatizada com a entonação da voz, independendo da movimentação de sobrancelhas. Nos casos de CC, entretanto, a sobrancelha influi bastante na decisão e os voluntários escolheram a palavra enfatizada com a sobrancelha praticamente duas vezes mais freqüentemente do que quando não havia este movimento. Para o italiano, percebeu-se não ser possível reconstruir o histórico do diálogo com base nas propriedades audiovisuais dos estímulos.

Os experimentos descritos acima tratam de agentes virtuais que usam movimentos não-verbais para complementar o que dizem. O primeiro usa gestos de braços e mãos para complementar informações enquanto que no segundo se testa especificamente movimentos de sobrancelhas.

A avaliação de ECAs é bastante influenciada por fatores subjetivos que são difíceis de se medir, afinal cada voluntário pode ter uma percepção diferenciada do mundo. Assim, o grupo de participantes deve ser cuidadosamente escolhido e as perguntas, testadas previamente ao início da avaliação para evitar entendimentos incorretos.

2.4 Considerações finais

Como mencionado previamente, este trabalho trata de um apresentador que transmite informações, sem interações com o usuário. De acordo com as figuras 2.2 e 2.3, portanto, ele se encaixa no primeiro caso. Apresentadores podem ter diversas funcionalidades, como aeromoças que fazem a demonstração das rotinas em casos de emergência nos aviões ou instrutores para operação de caixa automático.

O apresentador exposto aqui apresenta notícias. Na abordagem escolhida para este projeto, um apresentador de notícias não expressa emoções, pois a apresentação de notícias está tipicamente associada a um discurso neutro. Tampouco há interação com o usuário, que é exclusivamente ouvinte/observador.

A apresentação de notícias é beneficiada por alguns movimentos que tornam mais claro o discurso, enfatizando palavras e acentuando pausas. Foram implementados portanto os pontuadores e sinais conversacionais. Também um movimento com função biológica foi utilizado, o piscar de olhos.

Afinal, acompanhar notícias apresentadas por uma face virtual realista faz com que o espectador espere um comportamento realista.

De forma a criar um modelo de comportamento, foram analisados os movimentos faciais de um apresentador real. Os olhos foram tratados, dada a importância que o espectador dá, mesmo que inconscientemente, a eles. Afinal, se as pessoas focam mais nos olhos que no restante da face, eles certamente não podem ser negligenciados. Também as sobrancelhas, estando na região dos olhos e tendo grande poder expressivo foram consideradas.

Não é simples transmitir a sensação de que o personagem virtual está vivo, mas aos poucos, em etapas, é possível que se chegue mais perto deste objetivo. Vale atentar que os humanos nunca estão totalmente parados: eles piscam, esfregam as mãos ou mudam de postura. Uma vez que este projeto trata apenas de uma cabeça, deve-se procurar também passar essa sensação de movimento contínuo, nunca mantendo a cabeça em uma posição estática.

O repertório de movimentos escolhidos para acompanhar a fala foi o seguinte:

- rotação de cabeça,
- rotação de globos oculares,
- elevação de sobrancelhas,
- piscar de olhos,
- apertar de olhos.

Para implementar tais movimentos, usou-se o modelo de parametrização geométrica que envolvem rotação, translação e interpolação. A fala foi implementada seguindo a abordagem proposta por De Martino [16].

Entre os formalismos utilizados para descrição de animações faciais, tanto VHML quanto ECAF permitem que se descreva a animação com o texto a ser apresentado oralmente e os movimentos que devem ser realizados, conforme se pretende para o apresentador virtual. A abordagem implementada utilizou uma estrutura mais simplificada que a de ECAF e com um controle maior sobre o texto a ser marcado que VHML, que apenas marca palavras, enquanto que na abordagem escolhida, marca-se fonemas.

O próximo capítulo aborda a implementação do sistema, os estudos sobre o comportamento do apresentador real, a descrição do arquivo de entrada e os passos necessários para a criação da animação da apresentação de notícias.

Capítulo 3

Sistema de animação do apresentador virtual

Durante uma apresentação faz-se a exposição de informações com o intuito de acrescentar algo à platéia. O apresentador faz uso de gestos, mudanças de postura, expressões faciais e modulação da voz para aumentar sua expressividade, melhorando a qualidade da apresentação. Uma apresentação feita por um orador imóvel sem mudanças na entonação da voz dificilmente manteria tanto a atenção quanto uma realizada de forma mais dinâmica.

Para realizar apresentações usam-se diversos recursos que melhoram a compreensão da fala, mas a forma de agir difere de acordo com o objetivo e com a personalidade do apresentador. Assim, professores realizam apresentações interagindo com o quadro negro, enfatizando partes mais importantes da lição com a voz e também com expressões faciais para ensinar seus alunos. Técnicos de futebol fazem discursos motivacionais antes de partidas, incentivando jogadores, batendo palmas, gritando palavras de ação e relembrando a rivalidade com o time opontente, por exemplo. Apresentadores de notícias tipicamente adotam uma postura isenta, buscando apenas informar os espectadores de fatos correntes.

De modo a verificar como os conceitos que dão vida a personagens virtuais discutidos no capítulo anterior se aplicariam especificamente a um apresentador de notícias, observou-se um apresentador de notícias real. Foram gravados e analisados trechos de apresentações de notícias, permitindo que se fundamentasse o comportamento do apresentador virtual na movimentação espontânea de uma pessoa exercendo a atividade que se deseja reproduzir. Estudos sobre o comportamento do apresentador real e considerações a respeito da aparência do apresentador virtual estão expostos na seção 3.1.

Uma vez definido como o apresentador se comportará, trata-se dos detalhes de implementação na seção 3.2. A estrutura do sistema é descrita e discorre-se sobre as diferentes etapas necessárias para a criação da animação.

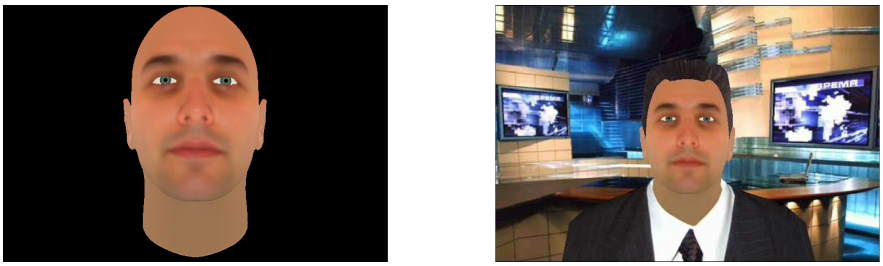
3.1 Apresentador de notícias

Um apresentador virtual de notícias é um agente representado por uma forma antropomórfica que permite a transmissão de informações usando uma combinação que envolve elementos como fala, gestos, expressões faciais e mudanças de postura, podendo ainda incluir a interação com objetos. A sua personificação pode ser bidimensional ou tridimensional, humanóide ou não, com características masculinas ou femininas, enfim, sua aparência pode ser extremamente diversificada.

Apesar de os esquemas de classificação da comunicação não-verbal na literatura serem muito variados, eles concordam que a aparência faz parte da comunicação não-verbal [45]. O uso de roupas e adornos comunica algo a respeito de quem os usa. Um apresentador de notícias real costuma ser caracterizado por vestimentas e adereços sóbrios, bem como um penteado discreto. Tradicionalmente, vê-se apenas parte do tronco de um apresentador de notícias e sua cabeça, onde se concentra a maior parte da movimentação, de modo que esta é a região de interesse para a representação de um apresentador.

Este trabalho teve como foco as atividades faciais, sendo que a movimentação se concentra na cabeça ilustrada na figura 3.1a. Para reforçar a caracterização do personagem virtual como apresentador de notícias, adicionou-se parte do tronco com textura de paletó de gravata e um modelo de cabelo. Também para melhorar a contextualização do apresentador, incluiu-se um fundo mostrando um ambiente de telejornal. Uma imagem mostrando o cenário com o apresentador pode ser observada na figura 3.1b.

A imagem apresentada na figura 3.1b deixa claro, à primeira vista, que se trata de um apresentador de notícias. O comportamento, entretanto, deve estar condizente com a aparência. Manter a face imóvel durante a fala pode passar sensação de artificialidade, mas aplicar movimentos muito enfáticos como os de um técnico procurando motivar seus jogadores também não contribui para caracterizar a função de apresentador de notícias.



(a) (b)

Fig. 3.1: a) cabeça do apresentador, a região articulada do mesmo e b) apresentador com corpo, cabelo e fundo

Para criar um padrão de comportamento convincente, é necessário identificar os movimentos do personagem que devem estar presentes durante o discurso e criar parâmetros que os descrevam. Como este trabalho cria um modelo de comportamento para um apresentador de notícias virtual, os estudos feitos envolveram um apresentador de notícias real. Tais estudos são relatados na seção seguinte.

3.1.1 Análise do comportamento de um apresentador real

Nesta seção apresenta-se a análise feita do comportamento do apresentador a partir de programas de notícia televisivos, que forneceu dados para criar regras de comportamento do modo a aumentar o realismo da animação do apresentador virtual. Tal análise foi feita, em sua maior parte, por inspeção visual de segmentos de vídeo.

Gravou-se em vídeo 42 seqüências de 3 a 30 segundos mostrando o apresentador real transmitindo informações para enriquecer o estudo sobre comportamento. Os segmentos foram digitalizados a 30 frames por segundo.

Piscar de olhos

Pisca-se involuntariamente com objetivo de umedecer os globos oculares. Piscar com função biológica ocorre em média a cada 4.8 segundos [51], mas este valor pode aumentar ou diminuir, dependendo de fatores como concentração e emoção. Devido a essa flutuação e para quantificar o padrão de piscar demonstrado por um apresentador de notícias real, o comportamento do mesmo foi analisado.

Sendo o piscar de olhos o movimento não-verbal de maior ocorrência, para permitir uma análise

mais exaustiva foi desenvolvida uma ferramenta *software* para detectá-lo automaticamente. Essa análise foi aplicada a um conjunto de 27 gravações de modo a obter a razão média do piscar e um desvio padrão do movimento.

Com base no algoritmo descrito em [10], um programa foi desenvolvido no software Matlab® para obter a freqüência do piscar de olhos.

A verificação da abertura dos olhos é feita através da comparação da região de olhos com um *template* de olho fechado (figura 3.2). Inicialmente seleciona-se uma região de olho fechado de uma das imagens da seqüência para ser o *template*. A partir do centro dessa região o programa busca pelo olho no próximo quadro. A cada imagem da seqüência, o programa atualiza o valor desse centro, de forma que são permitidos pequenos movimentos de cabeça. Considera-se que ambos os olhos piscam simultaneamente, de forma que apenas um olho é testado.

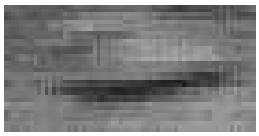


Fig. 3.2: *Template* do olho fechado.

O *template* é comparado com a região que o rodeia usando a função de correlação dada pela equação 3.1,

$$c = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A}))(\sum_m \sum_n (B_{mn} - \bar{B}))}} \tag{3.1}$$

onde A_{mn} é a matriz correspondente à região de face sendo testada e B_{mn}, o *template*. A média das regiões, $\bar{A}$ e $\bar{B}$ é subtraída para normalizar o brilho que pode variar de um quadro para outro.

O olho é considerado fechado quando o resultado da função de correlação é igual ou maior que 0.68, valor definido experimentalmente. Quando isso se dá, testa-se os próximos 5 frames e o valor máximo dentre estes é definido como piscar. Verificando a existência de piscar de olhos não detectados, obteve-se uma taxa de acerto para encontrar o olho fechado de 91,8%.

Os resultados mostram que ocorre um piscar de olhos em média a cada 46 quadros com um desvio padrão de 37. O alto valor de desvio padrão se deve à existência de piscar de olhos seguidos. Além disso, o apresentador analisado pode não piscar apenas para umedecer os olhos, mas também para complementar a fala, o que aumentaria a ocorrência do movimento.

Distribuição dos movimentos ao longo da fala

Tendo tratado dos estudos a respeito do piscar de olhos, vale apresentar algumas observações realizadas acerca dos movimentos que acompanham a fala do apresentador. Essas, de um modo geral, corroboraram as descrições dadas na literatura.

Como discutido no capítulo 2, pequenos movimentos de cabeça ocorrem durante toda a apresentação - o apresentador nunca mantém sua cabeça parada por muito tempo. Esses movimentos podem delimitar diferentes partes da frase, funcionando como vírgulas ou parênteses. Nesse caso, geralmente ocorre uma inclinação da cabeça para um lado ao entrar em novo tema e para o outro lado, ao retornar ao tema anterior.

Ao iniciar ou terminar uma apresentação também ocorrem movimentos com a cabeça que reforçam os limites da apresentação. Os que ocorrem no início, entretanto, normalmente se devem a mudanças na posição da câmera ou ao fato de o apresentador estar observando algo sobre sua mesa. Ao concluir, freqüentemente ocorre também o levantar das sobrancelhas como que para enfatizar a conclusão.

Freqüentemente, movimentos de sobrancelhas ocorrem simultaneamente a acenos de cabeça para baixo, seja com a função de pontuar, seja com a função de enfatizar trechos do texto.

A bibliografia afirma que os olhos não costumam ficar parados em situações normais, mas sim em constante movimento, como explicado na seção 2.1.2. Verificou-se que para o apresentador de notícias isso não ocorre, uma vez que ele provavelmente não se enquadra na situação considerada “normal”, por não estar se relacionando com outras pessoas. Devido à postura que se espera de um apresentador, ele mantém seus olhos sempre fixos em um mesmo ponto na tela, ou seja, quando a cabeça se move, eles se movem na direção contrária a ela para manter este foco.

O fechamento parcial de olhos, aqui denominado apertar de olhos, é o movimento menos utilizado durante uma apresentação de notícias. Ocorre apenas ao enfatizar trechos da fala e é geralmente

acompanhado de acenos com a cabeça.

Os movimentos não-verbais que atuam como ilustradores (com função de enfatizar a fala, conforme explicado na seção 2.1.2) claramente acompanham a entonação, reforçando as palavras acentuadas. Já os que atuam como pontuadores estão estreitamente relacionados às pausas realizadas.

Regras de comportamento

Considerou-se que um apresentador de notícias dá as informações mantendo um afastamento do tema, não inserindo opiniões pessoais. Assim, a partir das análises feitas, criou-se um conjunto de regras que descrevem um comportamento básico de apresentador de notícias para um personagem virtual.

Os estudos descritos deram ênfase à ocorrência do piscar de olhos e permitiram a obtenção de valores bastante precisos em relação à freqüência com que ocorrem. Os demais movimentos não-verbais da cabeça foram obtidos através de inspeção visual das gravações. As regras abaixo definem quando certos movimentos devem ser ativados:

1. Após 9 quadros sem piscar, uma função randômica gera um valor entre 0 e 74 que indica em quantos quadros será o próximo piscar. Caso haja um movimento de piscar ou apertar os olhos inserido pelo usuário, reinicializa-se o procedimento.

2. Quando a cabeça se move, os olhos são rotacionados para manter foco no mesmo ponto da tela. Isso se alcança rotacionando os olhos no ângulo oposto ao de rotação da cabeça.

3. O final de uma apresentação é reforçado por um aceno de cabeça.

4. Ao apertar os olhos ou elevar as sobrancelhas, uma função randômica é ativada para definir se acenos de cabeça devem ocorrer simultaneamente ou não. Isso aumenta a quantidade de movimentos de cabeça durante a animação, uma vez que a cabeça se move ligeiramente durante toda a apresentação.

As regras acima não levam em consideração o que está sendo dito, apenas estabelecem uma conexão entre eventos. Isso permite ao usuário controlar quais palavras devem ser enfatizadas, sem se preocupar com outros movimentos que costumam seguir um padrão e auxiliam a aumentar o realismo na performance do apresentador.

3.2 Implementação do sistema

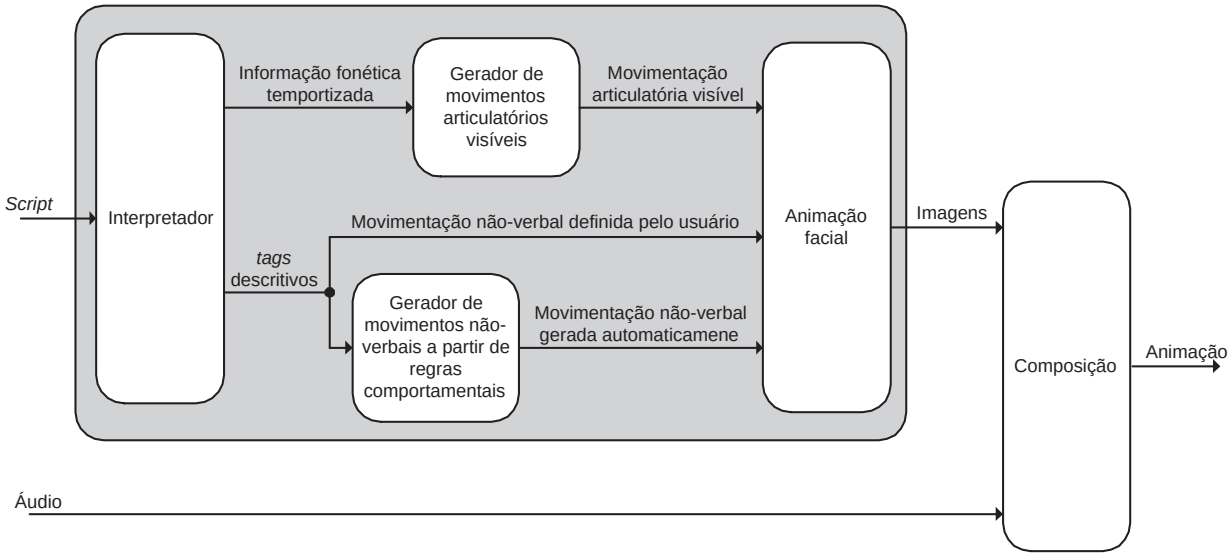


Fig. 3.3: Arquitetura do sistema

Uma visão geral do sistema é dada na figura 3.3. Como entrada têm-se o *script*, que descreve a movimentação facial do apresentador, e o arquivo de áudio, com o conteúdo sonoro.

O *script* contém a descrição da animação a ser gerada, com informações acerca da movimentação verbal e não-verbal e possui formato baseado em XML: o segmento do texto a ser acompanhado por movimentos não-verbais fica entre marcadores (*tags*) descritivos na estrutura <marcação> texto </marcação>. O conteúdo dos marcadores especifica o movimento que deve ocorrer durante a exposição oral do texto demarcado. Na seção 3.2.1 explica-se a obtenção dos dados para o script e discorre-se sobre sua formatação.

Uma vez comandada a leitura do *script*, os dados nele contidos são interpretados e reconhecidos como informação fonética temporizada, necessária para a geração dos movimentos de sincronismo labial ou informação não-verbal temporizada, dada pelos marcadores que descrevem os demais movimentos faciais.

A informação fonética temporizada, composta pelos segmentos sonoros e suas durações, é utilizada para gerar a movimentação articulatória visível, usando a abordagem descrita em [16]. Os movimentos não-verbais são gerados usando a estratégia de parametrização geométrica, de acordo com a informação dada pelas marcações.

Parte dos movimentos não-verbais é definida diretamente a partir da descrição da movimentação dada pelo usuário nos marcadores. Outra parte é gerada automaticamente a partir de regras comportamentais (tratadas na seção 3.1.1). Para tanto, é feita a verificação de quais regras se aplicam, o que também depende dos movimentos inseridos pelo usuário. Assim, se o usuário não inseriu nenhum piscar de olhos, este é inserido automaticamente, por exemplo.

Na seção 3.2.2 são descritos os procedimentos usados para modelar cada movimento. A região em cinza na figura refere-se ao processamento do script de entrada que, ao gerar a movimentação facial, retorna as imagens correspondentes à animação. O processamento é abordado na seção 3.2.3.

A animação final é composta pela sincronização das imagens geradas com o arquivo de áudio. Uma vez que não há dispositivo de sincronização com som embutido no sistema, é necessário fazer a sincronização fora dele. As imagens geradas anteriormente são integradas com o arquivo de som com o auxílio de software de edição de vídeo, como, por exemplo, Windows Movie Maker, originando a animação da apresentação de notícias.

3.2.1 *Script*

O usuário define no *script* de entrada como ele deseja que o apresentador virtual atue. Isso envolve definir o texto que deve ser apresentado oralmente e quais movimentos devem acompanhar a fala. Os dados do script de entrada são: os segmentos sonoros que constituem a notícia, a duração dos mesmos e os movimentos não verbais que acompanham a fala. A próxima seção descreve a obtenção de tais dados e a seção seguinte descreve sua formatação no *script* de entrada baseado em XML.

Dados do *script* de entrada

O objetivo principal do apresentador é informar uma notícia, de forma que o primeiro passo para escrever o *script* de entrada é definir o texto de tal notícia. Em seguida é necessário obter a seqüência de segmentos sonoros, ou fones, correspondentes a ela. A transcrição do discurso em segmentos sonoros pode ser obtida de forma manual ou automática. Neste trabalho foi utilizada a conversão automática com o auxílio do Conversor Texto Fala Aiuruetê [4]. Já a duração dos segmentos sonoros foi determinada pela inspeção visual do formato de onda e do espectograma do sinal de áudio. Esta inspeção visual foi feita utilizando o pacote de edição de som Cool Edit versão 96. Utilizando

informações do som e do espectograma é possível encontrar os limites entre um segmento sonoro e outro.

A figura 3.4 mostra o espectograma para a notícia que começa com “Terminou sem acordo uma reunião entre representantes...”. Na figura 3.5, é possível observar a delimitação encontrada para os dois primeiros segmentos sonoros da notícia. Por fim, a figura 3.6 ilustra a segmentação do início do arquivo.

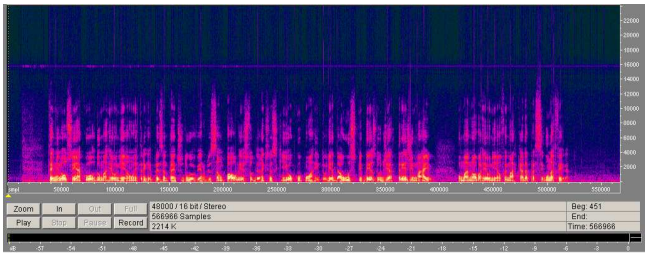
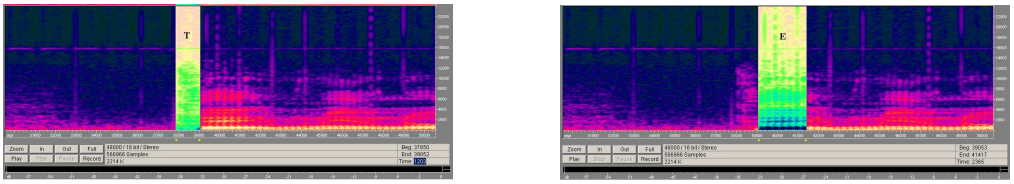


Fig. 3.4: Espectograma de uma notícia

O terceiro tipo de dado necessário ao arquivo de entrada é a movimentação não-verbal que acompanha a fala. Alguns movimentos, conforme explicado na seção 3.1.1, são gerados automaticamente pelo sistema. A ênfase que se dá a certas palavras, entretando, deve ser definida pelo animador, que



(a) (b)

Fig. 3.5: Inspeção da onda sonora no Cool Edit

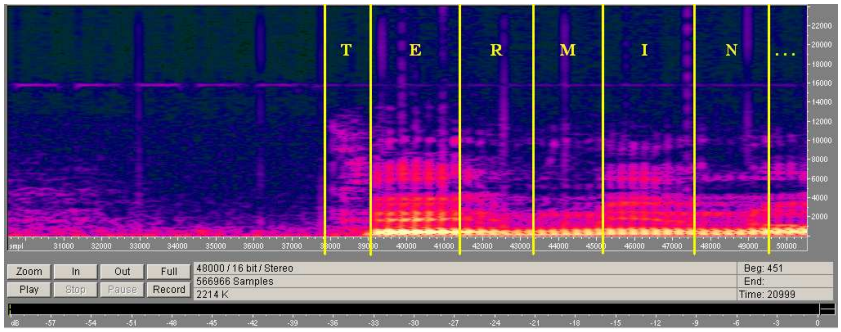


Fig. 3.6: Delimitação dos primeiros segmentos sonoros da frase

fica livre para escolher onde deve haver alguma ênfase e qual movimento deve ser usado, indicando também em qual segmento sonoro ocorre o ponto máximo do movimento.

A seção seguinte detalha a integração dos dados apresentados aqui em um arquivo baseado em XML.

Linguagem de controle

A animação é descrita por um *script* baseado em XML que contém informações acerca do texto a ser apresentado oralmente e dos movimentos que serão executados.

A descrição do texto a ser apresentado oralmente envolve a informação sobre cada segmento sonoro e sua duração:

f_1 d_1 f_2 d_2 f_3 d_3 ...

com f_i representando o segmento sonoro e d_i, a duração do mesmo. Nas tabelas 3.1 e 3.2, os valores que os fonemas podem assumir usando como base a pesquisa de [16] e o código correspondente usado para o arquivo de entrada: na coluna “fone”, os símbolos definidos pela IPA (*International Phonetic Assossiation*) e na coluna “código”, os símbolos utilizados pelo software Aiuruetê [4].

Fone	*Código*
[p]	p, P
[b]	b, B
[m]	m, M
[f]	f
[v]	v
[t]	t
[d]	d, D
[n]	n
[s]	s, S
[z]	z

Fone	*Código*
[l]	l, L
[ʃ]	sh
[ʒ]	zh
[ʎ]	lh
[ɲ]	nh
[k]	k, K, Ks
[g]	g, G
[ɾ]	R
[ɣ]	r

Tab. 3.1: Códigos dos segmentos consonantais aceitos pelo arquivo de entrada

Como exemplo, a frase “Palestinos e israelenses entram no oitavo dia de ofensiva.” seria escrita da seguinte forma:

```
/ 21375
```

Fone	*Código*
[i]	i, iN
[a]	a, eh, aN
[u]	u, uN, oN, ON, o, O
[ɪ]	I, IN
[ɐ]	A, e, E, oh, AN, eN, EN
[ʊ]	U, UN

Tab. 3.2: Códigos dos segmentos vocálicos aceitos pelo arquivo de entrada

```
p 1011 a 1809 l 1932 e 3433 S 7281 t 936 i 1962 n 1962 U 3403 S 3618
i 3005 S 3649 r 1686 a 3219 e 2085 l 3556 eN 5672 s 4506 I 1110 S 3587
eN 9780 t 1778 R 1318 AN 3342 U 1932 n 1349 u 3280
o 2115 I 5917 t 1840 a 6009 v 3955 U 4108 d 3434 i 2882 A 2759
d 1533 i 2303 o 3127 f 3740 eN 4691 s 6254 i 7849 v 3617 A 4354
/ 6101
```

onde ’/’ representa uma pausa.

Após colocar o texto no formato acima, o usuário pode determinar quais palavras ou fones devem ser enfatizados pelo apresentador e de que maneira. Isso é feito enriquecendo o texto que será lido pelo apresentador através da marcação das palavras escolhidas. Assim, o sistema recebe a informação da ação que o apresentador deve realizar enquanto ele fala os fones marcados.

Cada movimento possui um código que o identifica. A rotação de pescoço é dada por “RotPesc” e a elevação da sobrancelha por “Sob”, por exemplo. No marcador que descreve o movimento a ser executado, inclui-se o código do mesmo e os valores de seus parâmetros. Há parâmetros específicos para cada movimento e um parâmetro comum a todos, que define em qual fone o movimento atinge seu máximo, “sinc”.

Abaixo está o DTD (*Document Type Definition*) do arquivo de entrada. Um DTD descreve o formato de um arquivo XML, definindo o código dos elementos e os parâmetros exigidos por cada um.

```
<!ELEMENT RotPesc (#PCDATA| Piscar*| ApOlhos*|Sobr*)>
<!ATTLIST RotPesc pescAngX CDATA #REQUIRED
                  pescAngY CDATA #REQUIRED
                  pescAngZ CDATA #REQUIRED
                  sinc     CDATA #REQUIRED>

<!ELEMENT Piscar (#PCDATA)>
<!ATTLIST Piscar  sinc CDATA #REQUIRED>

<!ELEMENT ApOlhos (#PCDATA)>
<!ATTLIST ApOlhos sinc CDATA #REQUIRED>

<!ELEMENT Sobr (#PCDATA| RotPesc*| Piscar*| ApOlhos*)>
<!ATTLIST Sobr intensidade CDATA #REQUIRED
               sinc        CDATA #REQUIRED>
```

Os movimentos que podem ser usados para acompanhar a fala são dados por ELEMENT acima. Cada um deles pode conter outros elementos, colocados entre parênteses. O asterisco indica que um elemento pode ser contido zero ou mais vezes. Os atributos dos elementos são indicados por ATTLIST e, quando não são opcionais, estão demarcados com #REQUIRED. Assim, a rotação de pescoço, por exemplo, é indicada pelo código RotPesc e sua lista de atributos dada por ATTLIST é composta pelos ângulos pescAngX, pescAngY e pescAngZ, que indicam o grau de rotação ao redor de cada eixo e por sinc, que faz a sincronização com a fala.

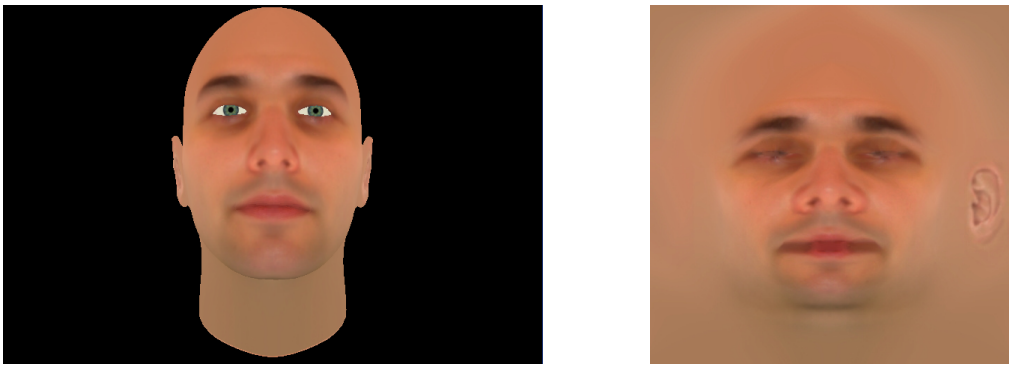
Tomando a frase dada como exemplo acima, pode-se definir que ela seja acompanhada por um aceno com a cabeça e um movimento de apertar de olhos que culminam no final da palavra “palestinos” seguidos por um levantar de sobrancelhas durante “entram” e outro aceno com a cabeça. Para gerar a animação correspondente, é necessário indicar os parâmetros desejados para cada movimento, como os ângulos de rotação para o aceno de cabeça e a intensidade da elevação de sobrancelhas. Também deve-se determinar em que fone o movimento atinge seu máximo com o parâmetro “sinc”. Assim, basta envolver a frase com as descrições dos movimentos:

```
/ 21375
<RotPesc pescAngX="15" pescAngY="0.0" pescAngZ="0.0" sinc="9">
<ApOlhos sinc ="9">
   p 1011 a 1809 l 1932 e 3433 S 7281 t 936 i 1962 n 1962 U 3403 S 3618
   i 3005 S 3649 r 1686 a 3219 e 2085 l 3556 eN 5672 s 4506 I 1110 S 3587
</ApOlhos>
</RotPesc>
<Sobr intensidade="1.0" sinc="2">
   eN 9780 t 1778 R 1318 AN 3342 U 1932
</Sobr>
<RotPesc pescAngX="-10" pescAngY="0.0" pescAngZ="10" sinc="8">
   n 1349 u 3280 o 2115 I 5917 t 1840 a 6009 v 3955 U 4108
   d 3434 i 2882 A 2759 d 1533 i 2303
   o 3127 f 3740 eN 4691 s 6254 i 7849 v 3617 A 4354
</RotPesc>
   / 6101
```

Na próxima seção trata-se dos detalhes da modelagem dos movimentos usados para dar vida ao apresentador virtual.

3.2.2 Modelagem dos movimentos faciais

A base do sistema do apresentador virtual é o modelo geométrico tridimensional da cabeça, que atua como apresentador. A cabeça virtual (figura 3.7a) é composta por olhos, língua, arcada dentária superior e inferior, cavidade da boca e pele. Ao todo, a malha da face possui 5.960 vértices e 34.476 triângulos, o que permite um nível de detalhamento do modelo bastante satisfatório. A textura (figura 3.7b) foi gerada a partir de fotos de um modelo real garantindo maior realismo.



(a) (b)

Fig. 3.7: a) modelo do apresentador, b) textura da face do apresentador virtual

Os movimentos que complementam a fala foram implementados usando parametrização geométrica, método de animação comentado na seção 2.2.1. Usou-se rotação, interpolação e translação de vértices aplicados a certas regiões da malha facial para alcançar os efeitos desejados. A implementação foi feita em C++ usando a biblioteca gráfica OpenGL.

Vale comentar que não há necessidade de tratar a textura ao realizar as operações que alteram a geometria facial, pois há um mapeamento entre cada ponto do arquivo de textura para cada vértice da face. Assim, mesmo que se desloque determinado vértice, sua textura continua associada a ele.

Os movimentos implementados foram: rotação do pescoço, rotação dos globos oculares, piscar de olhos, apertar de olhos, elevação de sobrancelhas e sincronismo labial. Em seguida descreve-se a implementação desses movimentos.

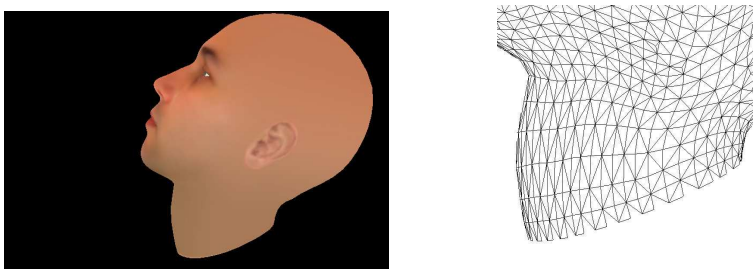
Rotação do pescoço

Código do movimento: RotPesc

Parâmetros de entrada específicos: [pescAngX, pescAngY, pescAngZ] - ângulos de rotação ao redor dos eixos x, y e z.

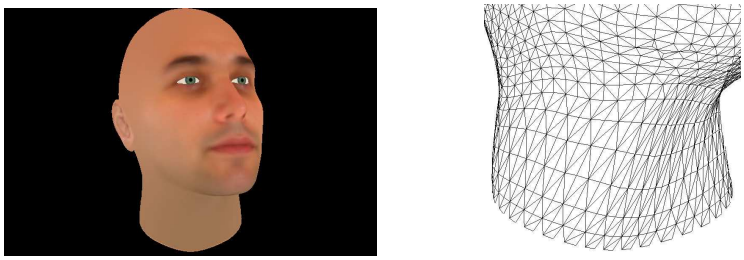
Não se pode rotacionar todo o pescoço e a cabeça com um mesmo ângulo, ou se terá uma impressão de descontinuidade, afinal, ao rotacionar o pescoço, sua base deve continuar fixa sobre o corpo. O pescoço do modelo é composto por um conjunto de círculos superpostos. Portanto, para obter uma variação suave ao longo do pescoço, rotacionou-se essas “camadas” com ângulos distintos, sendo que conforme se está mais próximo da cabeça, maior o ângulo. A rotação é feita ao redor do centro da base da cabeça.

As figuras 3.8, 3.9 e 3.10 ilustram a rotação do pescoço ao redor dos eixo x, y e z a) na versão texturizada do modelo e b) com enfoque na versão aramada.



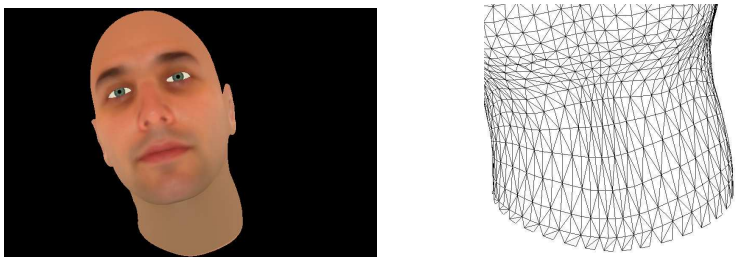
(a) (b)

Fig. 3.8: Rotação do pescoço ao redor do eixo x



(a) (b)

Fig. 3.9: Rotação do pescoço ao redor do eixo y



(a) (b)

Fig. 3.10: Rotação do pescoço ao redor do eixo z

Rotação dos globos oculares

O apresentador apenas move os olhos ao rotacionar a cabeça, com o intuito de mantê-los fixos em um ponto da tela, portanto este movimento não é evocado pelo usuário, mas sim internamente ao programa. Assim, quando o usuário comanda uma rotação da cabeça, o sistema usa os valores

(a) (b)

Fig. 3.11: Rotação de olhos: a) ao redor do eixo x, b) ao redor do eixo y

pescAngX e pescAngY, que indicam os ângulos de rotação ao redor dos eixos x e y, para realizar a rotação dos olhos na direção contrária.

Para executar o movimento, faz-se a translação dos globos oculares para a origem, para então rotacioná-los ao redor dos eixos x e y. A seguir, translada-se as duas esferas novamente para suas posições originais. Abaixo, a matriz de transformação usada para realizar as rotações ao redor dos eixos x e y, sendo que sinX e cosX são o seno e cosseno do ângulo de rotação ao redor do eixo x e sinY e cosY, ao redor do eixo y.

```
matrizDeRotacao[4][4] = {
  {              cosX,              0,                 sinX,       +xe*(1-cosX)-ze*sinX},
  {-sinX* sinY,   cosY,    sinY*cosX, sinY*(+ze*(1-cosX)+xe*sinX)+ye*(1-cosY)-ze*sinY},
  {-sinX* cosY,  -sinY,    cosX*cosY, cosY*(+ze*(1-cosX)+xe*sinX)+ze*(1-cosY)+ye*sinY},
  {                        0,              0,                 0,                      1}};
```

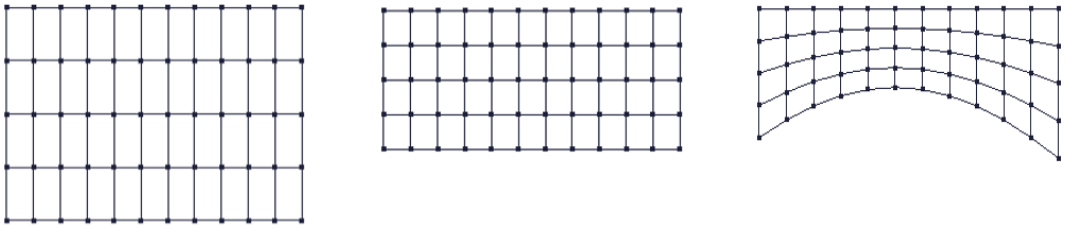
Piscar de olhos

Código do movimento: Piscar

Parâmetros de entrada específicos: não há.

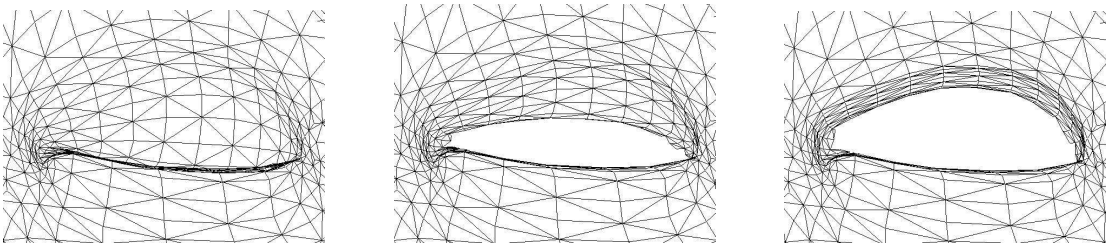
Ao piscar, tantos as pálpebras inferiores quanto as superiores se deslocam, ficando em contato quando os olhos se encontram fechados. Fez-se testes procurando reproduzir esse fato, mas o movimento das pálpebras inferiores causou um estranhamento, não contribuindo para melhorar o efeito resultante. Para este projeto, optou-se portanto por apenas tratar do deslocamento das pálpebras superiores para o piscar de olhos.

Uma vez que o modelo da face em posição de relaxamento tem os olhos fechados, é possível realizar um deslocamento dos vértices da parte inferior das pálpebras em direção aos da parte superior, dando o efeito de abertura de olhos e fazendo o oposto para realizar o fechamento.



(a) (b) (c)

Fig. 3.12: Descrição da movimentação das pálpebras: a) representação da pálpebra fechada, b) as pálpebras são abertas reduzindo-se a distância entre as camadas de vértices, c) através do deslocamento dos pontos ponderado por uma função senoidal obtém-se obtenha uma curvatura suave



(a) (b) (c)

Fig. 3.13: Apresentador com: a) olhos fechados e b) olhos semi-abertos e c) olhos abertos

Considera-se os vértices das pálpebras organizados em camadas (figura 3.12a) e calcula-se a distância entre elas. Para realizar o movimento de abertura do olho, reduz-se a distância entre camadas de forma que os vértices todos ficam próximos. A figura 3.12b ilustra essa operação. O deslocamento foi feito de acordo com uma função senoidal, para obter o formato arredondado suave que as pálpebras tomam quando abertas. Assim, cada coluna de vértices é deslocada com um peso diferente, de modo a criar uma forma arredondada, conforme mostrado na figura 3.12.

Apertar de olhos

Código do movimento: ApOlhos

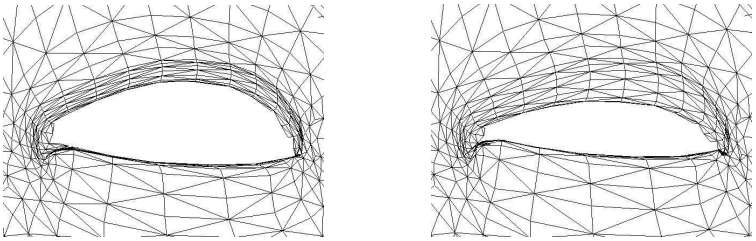
Parâmetros de entrada específicos: não há.

Este movimento consiste no abaixamento das pálpebras superiores e na elevação das pálpebras inferiores ocasionando o fechamento parcial das pálpebras, deixando ainda parte dos olhos à mostra. Ao contrário do que ocorre no piscar de olhos, o movimento da pálpebra inferior contribui visual-

mente para a percepção deste movimento, de forma que também foi tratado.

Usando o mesmo procedimento aplicado ao piscar de olhos, realiza-se um fechamento parcial das pálpebras superiores. Para realizar o movimento das pálpebras inferiores não se pôde utilizar a mesma estratégia, uma vez que não há pontos-alvo em direção aos quais se poderia deslocar os vértices de pálpebra inferior. Assim, realizou-se a sua rotação ao redor do centro do olho. Uma vez que a movimentação é relativamente pequena, não há o estranhamento do rompimento de continuidade mencionado na explicação da rotação do pescoço.

Optou-se por não permitir passagem de parâmetros pois o apertar de olhos não tem seu efeito de intensidade aumentado de forma significativa ao aumentar ou diminuir a abertura de pálpebras.



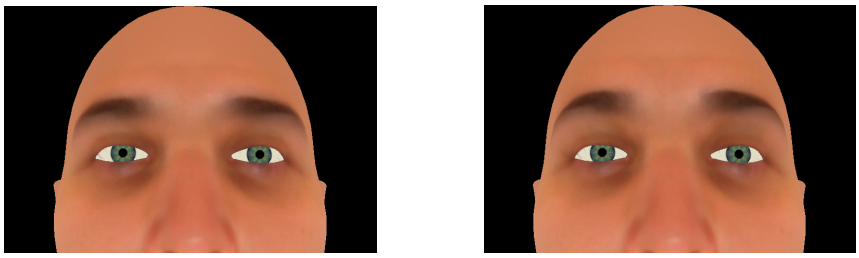
(a) (b)

Fig. 3.14: Representação aramada mostrando: a) olhos abertos e b) apertando os olhos

Elevação de sobrancelhas

Código do movimento: Sobr

Parâmetros de entrada específicos: [intensidade] - intensidade variando de 0 (relaxamento) a 1 (elevação máxima).



(a) (b)

Fig. 3.15: Apresentador com: a) sobrancelhas relaxadas e b) sobrancelhas elevadas

Este movimento é realizado de forma similar à elevação de pálpebras: tendo definido como região de interesse a fronte do modelo, que inclui sobrancelhas e testa, faz-se um deslocamento da região

inferior (sobrancelhas) em direção à superior (topo da face). Tal como na elevação de pálpebras faz-se um deslocamento ponderado por uma função: usa-se uma curva seguindo uma distribuição gaussiana para obter uma curvatura mais suave.

Sincronismo labial

Parâmetros de entrada específicos - fone e duração do mesmo

O algoritmo de sincronização labial foi implementado de acordo com a abordagem apresentada em [16], que descreve a deformação a ser aplicada à região dos lábios e também os movimentos rígidos de rotação e deslocamento do maxilar, para que se obtenha os visemas.

Visema é a configuração que a região da boca assume ao emitir diferentes sons, tendo este nome por ser o equivalente visual do termo fonema. Quando não se consegue distingüir visualmente entre um fone e outro, tem-se homofemas, visemas que são iguais para diferentes segmentos sonoros. Os homofemas adotados neste trabalho foram tomados de [16] e podem ser vistos nas tabelas 3.3 e 3.4.

A influência de um visema nos visemas vizinhos é chamada de coarticulação: o formato da boca ao pronunciar um mesmo segmento sonoro muda de acordo com os segmentos sonoros vizinhos. Afinal, ao falar uma palavra, não se emite simplesmente cada segmento sonoro que a compõe, como quando se soletra, o que faria com que os visemas fossem independentes entre si, mas faz-se a ligação de um segmento sonoro com os seguintes, o que implica em uma interdependência dos visemas. O visema do [r], por exemplo, é bastante diferente para “rasgo” e “rosto”.

Em [16] tratou-se da coarticulação usando o conceito de contexto fonético, em que cada segmento sonoro é influenciado pelo segmento sonoro anterior e pelo posterior. O [r] do exemplo acima, seria r_1 para “rasgo” e r_2 para “rosto”. A correspondência entre visemas e contextos fonéticos pode ser vista nas tabelas 3.3 e 3.4.

Para obter resultados realistas, os deslocamentos de 5 pontos-chave (mostrados na figura 3.16a) de um ator pronunciando diversos segmentos sonoros foram medidos. Além disso, mediu-se também o momento em que é atingida a posição equivalente à configuração final do visema para cada ponto. Essa posição é chamada alvo articulatório e varia de um visema para outro. Em alguns casos, os pontos-chave atingem seu alvo articulatório antes mesmo da emissão do som, que é o caso

Símbolo	*Contextos*	*Denominação*	*Homofemas*
<p1>	[pi] [pa] [ipɪ] [ipɐ] [ipʊ] [apɪ] [apɐ] [apʊ] [upɐ]	Bilabiais	[p, b, m]
<p2>	[pu] [upɪ] [upʊ]		
<f1>	[fi] [fa] [ifɪ] [ifɐ] [ifʊ] [afɪ] [afɐ]	Labiodentais	[f, v]
<f2>	[fu] [afʊ] [ufɪ] [ufɐ] [ufʊ]		
<t1>	[ti] [tu] [itɪ] [itɐ] [itʊ] [atɪ] [atʊ] [utɪ] [utɐ] [utʊ]	Alveolares Plosivos/Nasais	[t, d, n]
<t2>	[ta] [atɐ]		
<s1>	[si] [sa] [isɪ] [isɐ] [asɪ] [asɐ]	Alveolares Fricativos	[s,z]
<s2>	[su] [isʊ] [asʊ] [usɪ] [usɐ] [usʊ]		
<l1>	[li] [ilɪ] [alʊ] [ulɪ] [ulɐ]	Alveolares Laterais	[l]
<l2>	[la] [ilɐ] [alɪ] [alɐ]		
<l3>	[lu]		
<l4>	[ilʊ] [ulʊ]		
<ʃ1>	[ʃi] [ʃa] [iʃɪ] [iʃɐ] [iʃʊ] [aʃɪ] [aʃɐ] [aʃʊ] [uʃɪ] [uʃɐ]	Pós-alveolares	[ʃ, ʒ]
<ʃ2>	[ʃu] [uʃʊ]		
<ʎ1>	[ʎi] [ʎa] [iʎɪ] [iʎɐ] [aʎɪ] [aʎɐ]	Palatais	[ʎ, ɲ]
<ʎ2>	[ʎu] [uʎɪ] [uʎɐ]		
<ʎ3>	[iʎʊ] [aʎʊ] [uʎʊ]		
<k1>	[ki] [ikɪ] [ikɐ] [akɪ] [ukɪ] [ukɐ]	Velares Plosivos	[k, g]
<k2>	[ka] [akɐ]		
<k3>	[ku] [iʊ] [akʊ] [ukʊ]		
<ɾ1>	[ɣi] [ɣa] [iɾɪ] [iɾɐ] [aɾɪ] [aɾɐ] [uɾɐ]	Velares Fricativos/Tepes	[ɣ] [ɾ]
<ɾ2>	[ɣʊ] [iɾʊ] [aɾʊ] [uɾɪ] [uɾʊ]		

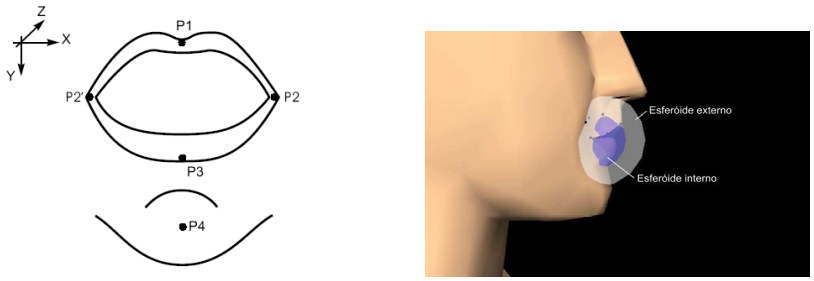
Tab. 3.3: Visemas consonantais e seus contextos fonéticos (extraído de [16])

Símbolo	*Contextos*	*Denominação*	*Homofemas*
<i1>	todos os contextos exceto [tit] e [ʃiʃ]	Alto Anterior	[i, ĩ]
<i2>	[tit] e [ʃiʃ]		
<a>	todos os contextos	Baixo Central	[a, ɛ, ɐ̃]
<u>	todos os contextos	Alto Posterior	[u, o, õ, ũ]
<ɪ>	todos os contextos	Postônico Alto Anterior	[ɪ]
<ɐ>	todos os contextos	Postônico Alto Central	[ɐ, e, ɔ, ẽ]
<ʊ>	todos os contextos	Postônico Alto Posterior	[ʊ]

Tab. 3.4: Visemas vocálicos e seus contextos fonéticos (extraído de [16])

do fone “p”, por exemplo. A partir dos dados medidos, o autor criou uma correspondência entre o deslocamento dos pontos-chave e cada visema.

O sistema recebe como entrada os fones, de acordo com o mapeamento dado nas tabelas 3.1 e 3.2



(a) (b)

Fig. 3.16: Explicação dos movimentos de sincronismo labial: a) pontos-chave monitorados, b) região de influência do lábio superior (figuras extraídas de [16])

e suas durações. Em seguida é feita a verificação do contexto fonético para cada fone. Obtêm-se os valores de cada ponto de interesse P_i e o momento em que cada atinge seu alvo articulatório - este valor é dado como um percentual da duração do fone. Calcula-se então a posição dos pontos de interesse para cada quadro da animação.

Os vértices do modelo geométrico da cabeça virtual afetados pela fala são reposicionados de acordo com os movimentos dos pontos-chave. Os deslocamentos dos vértices do modelo geométrico ocorrem devido a movimentos rígidos, correspondentes à movimentação da mandíbula, e a deformações na superfície da face, na região ao redor da boca, ocasionadas pela movimentação voluntária do tecido labial durante a produção da fala.

Os movimentos rígidos de translação e rotação da mandíbula são calculados a partir da trajetória do ponto P_4, na ponta do queixo.

As deformações ao redor da boca ocorrem em duas regiões de influência associadas aos pontos-chaves localizados nos lábios. Os pontos P_1, P_2 e P_2' influenciam a região do lábio superior. Os P_3, P_2 e P_2' influenciam a região do lábio inferior. Cada uma destas regiões são definidas por dois esferóides: um esferóide interno, que define os pontos de maior influência dos pontos-chaves no deslocamento dos vértices, e um esferóide externo que limita a região de influência. Entre os dois esferóides há um decaimento da influência sofrida pelos vértices dado pela função 3.2.

$$\Delta\bar{V} = R_i[K_i\Delta\bar{P}_2 + (1 - K_i)\Delta\bar{P}_i] \qquad \text{i=1,3} \tag{3.2}$$

(a) (b)

Fig. 3.17: Apresentador falando os fones “p” e “a”

Sendo que $\Delta\bar{V}$ é o deslocamento a ser aplicado ao vértice, $\Delta\bar{P}_2$ é deslocamente de P_2, $\Delta\bar{P}_i$ é deslocamento de P_i, R_i é um fator de decaimento esferoidal que considera a distância do vértice ao esferóide interno, variando de 0 a 1 e K_i é um fator de interpolação que pondera as distâncias entre o vértice i em questão e os pontos P_2 e P_i. Para a região superior i=1 e para a inferior, i=3, como mostrado na figura 3.16a.

Na figura 3.17, pode-se observar o apresentador falando os fones ‘p’ e ‘a’.

3.2.3 Processamento

O usuário, já tendo escrito o *script* de entrada e inicializado o programa, comanda a leitura do mesmo. É feito o processamento do *script* e uma estrutura de dados é criada para cada fone. A estrutura de dados consiste em: fone, duração do fone e valor dos parâmetros específicos de cada movimento naquele momento.

O procedimento abaixo mostra a seqüência de passos para o processamento do *script* de entrada:

```
( 1)    WHILE existe próximo comando{
( 2)        ler próximo comando
( 3)        IF comando é texto{
( 4)           salvar fone na estrutura de dados
( 5)           salvar duração do fone na estrutura de dados
( 6)        }
( 7)        IF comando é marcação{
( 8)           ler conteúdo da marcação (volta para (2))
( 9)           calcular valor dos movimentos dada a duração
(10)           salvar valores na estrutura de dados
(11)        }
(12)    }
```

Percorre-se o arquivo de entrada (linhas 1 e 2) lendo os blocos de comandos, que podem ser seqüência de texto ou marcação. Se o comando for texto (linha 3), registram-se cada fone e sua

duração na estrutura de dados (linhas 4 e 5). Caso o comando seja marcação (linha 7), inspeciona-se seu conteúdo (linha 8), voltando para o início (linha 2). O conteúdo pode ser outra marcação e nesse caso, repete-se a leitura do seu conteúdo até que se chegue no texto a ser acompanhado pelos movimentos descritos nos marcadores.

Uma vez lido o texto marcado, tem-se a duração de cada fone, o que permite calcular a intensidade do movimento (descrito na marcação) para cada fone. Conforme mencionado previamente, todos os movimentos possuem um parâmetro “sinc”, que indica em qual fone o movimento atinge o valor máximo.

Para calcular o valor da intensidade de um movimento para cada fone marcado por ele, divide-se esse valor pela soma das durações dos fone até o fone dado por “sinc” e então multiplica-se este valor pela duração de cada fone. Para o relaxamento do movimento realiza-se os mesmos cálculos, mas para o restante dos fone marcados. As informações das intensidades dos movimentos são adicionadas à estrutura de dados (linhas 9 e 10).

Terminada a leitura do arquivo de entrada (linha 12), a estrutura de dados está completa e é possível verificar se há movimentos que devem ser inseridos automaticamente, levando em conta as regras de comportamento estudadas na seção 3.1.1. Assim, após 9 quadros sem piscar ou apertar de olhos, define-se um valor randomicamente, variando de 0 a 74, para ser o próximo quadro em que haverá piscar de olhos. Movimentos com a cabeça ao redor dos eixos x e y fazem com que os olhos sejam rotacionados com os mesmos valores, mas na direção contrária. O elevar de sobrancelhas determina randomicamente se haverá ou não um movimento de cabeça a acompanhá-lo e, ao término da apresentação, é lançado um aceno com a cabeça.

Uma vez calculados todos os movimentos, os vértices são movimentados da maneira descrita na seção 3.2.2 e são geradas as imagens correspondentes no formato bitmap, de forma que se obtenha uma taxa de 30 imagens por segundo.

3.3 Considerações finais

Neste capítulo discorreu-se sobre detalhes da implementação do sistema de animação de um apresentador de notícias virtual. A entrada é um arquivo baseado em XML que descreve a animação,

permitindo ao usuário definir qual a notícia a ser dada pelo apresentador e como o texto deve ser acompanhado por movimentos não-verbais tais como levantar de sobrancelhas e acenos com a cabeça.

Os movimentos verbais foram implementados de acordo com a abordagem descrita em [16] e representam visualmente os fone dados no arquivo de entrada. Os movimentos não-verbais que enriquecem o discurso foram implementados utilizando parametrização geométrica e são sincronizados com os segmentos sonoros que constituem a fala de acordo com as especificações do usuário e regras de comportamento.

A aparência do apresentador virtual foi inspirada na de um apresentador de notícias real, cujo comportamento foi observado de modo a tornar a animação mais convincente. Esse estudo verificou que alguns movimentos se repetem em dadas circunstâncias, como o final de uma apresentação que costuma ser marcado por um aceno com a cabeça. Ao realizar tais acenos, os olhos do apresentador se mantêm fixos em um ponto da tela. Obteve-se também a freqüência média de piscar de olhos do apresentador real para que o animador não precise se preocupar com isso. A inserção automática de movimentos facilita o trabalho do animador, simplificando o *script*.

A animação é gerada através da composição das imagens geradas pelo programa que processa o arquivo de entrada com o arquivo de áudio que apresenta a notícia. Como resultado tem-se o apresentador movendo lábios e executando movimentos não-verbais em sincronia com o som. Diversos textos podem ser apresentados e, além disto, cada texto pode ser apresentado de uma maneira diferente, dependendo das escolhas feitas pelo usuário ao escrever o *script*.

Para avaliar a contribuição dos movimentos não-verbais implementados, foram feitos testes comparativos entre diferentes atuações do apresentador virtual e o apresentador real, conforme descrito no próximo capítulo.

Capítulo 4

Testes e Avaliação

O sistema descrito neste trabalho permite a geração de animações que mostram um apresentador virtual movendo a boca em sincronia com a fala e fazendo movimentos não-verbais para complementá-la. A abordagem utilizada para a sincronização labial e representação dos movimentos articulatórios visíveis foi avaliada por [18], sendo que os resultados indicaram que a mesma contribui para a inteligibilidade da fala. Considerando esses resultados, o presente trabalho concentrou-se em verificar especificamente a contribuição dos movimentos não-verbais adicionados.

Conforme discutido anteriormente na seção 2.3, a avaliação de personagens virtuais é difícil de se fazer, por ser influenciada por uma diversidade muito grande de fatores. A interpretação dos conceitos que se deseja avaliar, por exemplo, pode variar de pessoa para pessoa e a percepção da aplicação também pode depender de sua área de especialização [61]. Os voluntários devem, portanto, ser escolhidos de modo a evitar o acréscimo de variáveis que se deve levar em conta ao analisar os resultados.

A escolha do tipo de avaliação que será usado também tem um papel importante sobre a conclusão acerca do ECA [11]. Em alguns casos, pode-se escolher fazer um estudo de caso, que envolve poucos participantes e a obtenção de muitas informações detalhadas sobre um certo fenômeno, permitindo uma avaliação aprofundada. Um levantamento, por outro lado, permite que se tenha uma boa idéia sobre a população de interesse, de um modo geral, envolvendo um número grande de participantes e métodos de obtenção de dados que não são muito trabalhosos. Há ainda a possibilidade de se fazer um experimento, que foi a estratégia usada neste trabalho.

Para realizar um experimento, é necessário que o idealizador defina uma condição de controle e condições experimentais e essas condições seja distribuídas randomicamente entre os participantes. Num experimento, pretende-se identificar uma relação causal e, no caso tratado aqui, deseja-se estabelecer uma relação entre a presença de sinais não-verbais colocados de forma a complementar a fala e a compreensão da notícia e naturalidade da apresentação.

Uma visão geral das principais etapas da avaliação pode ser observada na figura 4.1 que adota a representação de um modelo de pesquisa apresentado na literatura [11]. Tal como na referência, segue-se a sugestão de trabalhar de trás para frente, definindo primeiro o objetivo da pesquisa, em seguida a natureza da mesma e, por fim, escolhendo a referência na literatura.

O objetivo (parte A da figura 4.1) é corroborar a expectativa de que o comportamento não-verbal acompanhando o texto dado no script de forma coordenada melhora o conforto do usuário com o personagem virtual e verificar se houve aumento no entendimento de informações.

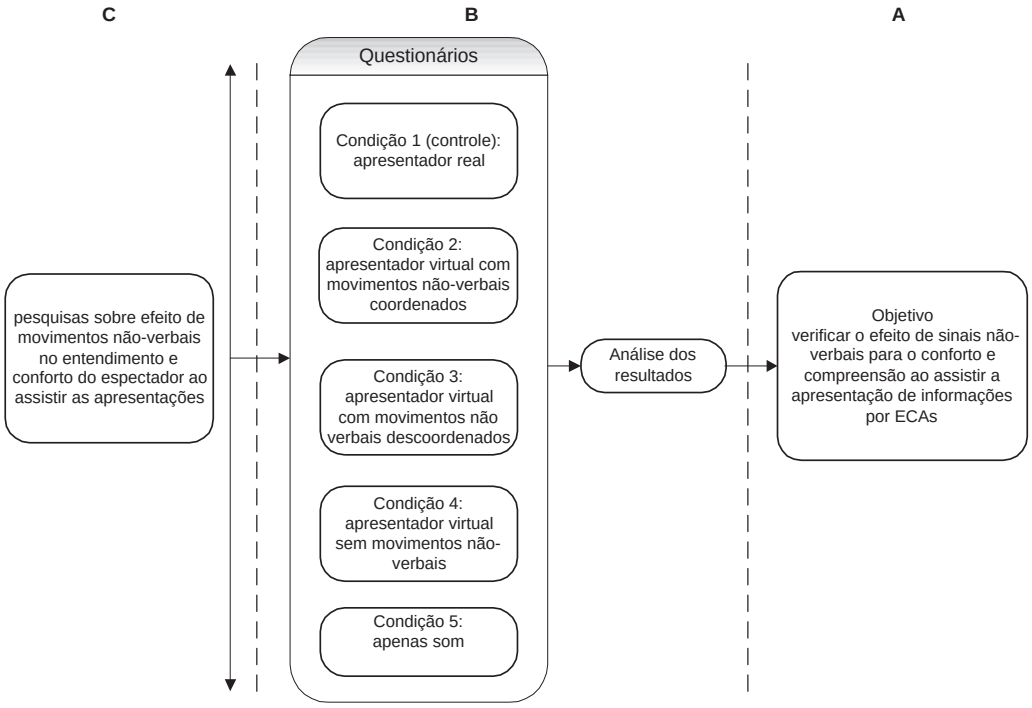


Fig. 4.1: Modelo de pesquisa

A avaliação testou os seguintes termos:

• Naturalidade - quão natural a apresentação pareceu ao voluntário: se ele viu diante de um agente robótico ou humanizado.

• Compreensão - se houve compreensão das notícias apresentadas.

Realizou-se um experimento usando questionários com perguntas abertas e fechadas. Perguntas abertas são aquelas que não possuem um limite de respostas pré-definido, enquanto que as fechadas sim [11]. Os questionários foram usados para comparar os diferentes usos das movimentações não-verbais. As situações contrastadas (parte B da fig. 4.1) podem ser detalhadas da seguinte maneira:

• apresentador real (AReal) - esta é a condição de controle. Por ter sido tomada da realidade, é considerada a referência.

• apresentador virtual com movimentos não-verbais coordenados (AVCoord) - usa movimentos não-verbais coordenados com a fala para complementar o que está sendo dito, de acordo com os estudos descritos nos capítulos 2 e 3.

• apresentador virtual com movimentos não-verbais descoordenados (AVDescoord) - usa os mesmos movimentos verbais e não-verbais que o apresentador virtual descrito acima, sem estes estarem, entretanto, coordenados com a fala. A colocação dos movimentos não-verbais ao longo da fala não tem relação com o discurso e a intensidade deles é, por vezes, exagerada, resultando em uma apresentação que se poderia considerar confusa.

• apresentador virtual sem movimentos não-verbais (AVSem) - possui os movimentos verbais idênticos aos dos outros apresentadores virtuais, mas mantém o resto da face imóvel.

• apenas o som, sem apresentador (SOM) - os mesmos arquivos de som usados para as apresentações descritas acima são usados neste caso. Esta condição permite que se verifique a contribuição da presença do apresentador.

Por fim, identificam-se as teorias usadas como ponto de partida para a avaliação (parte C da fig. 4.1). De acordo com a literatura [55], sinais não-verbais (tais como gestos com braços, expressões faciais e mudança de postura) acompanhando a fala melhoram o conforto do espectador, que passa a não se ver à frente de um agente robótico/humanóide, mas sim humanizado. Além disso, acompanhar finais de frase e pausas com gestos, bem como acompanhar palavras enfatizadas com alterações na entonação da voz e com movimentos de face, permitem que o interlocutor compreenda melhor o que está sendo dito e o que é importante no discurso. Caso a entonação da voz enfatize uma palavra e a movimentação não-verbal, outra, pode-se ter diferentes interpretações possíveis, dependendo do sinal dominante. A avaliação descrita neste capítulo visa verificar se essas teorias também se aplicam

especificamente a um apresentador virtual.

Em seguida, na seção 4.1, a descrição do protocolo de teste e na seção 4.2, a análise dos resultados.

4.1 Protocolo de teste

Foi criado um programa de avaliação, usando a linguagem de programação Java, para facilitar os testes garantindo uma padronização para todos os participantes e permitir que os dados fossem coletados automaticamente.

O experimento consiste em mostrar os apresentadores informando notícias diferentes e requisitar o preenchimento de questionários a cada apresentação. As notícias têm duração variando de 10 a 20 segundos e são trechos de um programa de notícias de um canal de informações brasileiro, gravados em diferentes dias, durante 2007.

Cada um dos 5 apresentadores informa duas notícias diferentes (Ni - sendo i o índice que indica de qual das 10 notícias se trata), somando um total de 10. A soma das durações das notícias informadas pelos apresentadores é similar para cada apresentador.

Para que as notícias não tenham influência na avaliação dos apresentadores, há 5 versões do experimento. Cada versão foi apresentada para um grupo de 8 pessoas, totalizando 5 conjuntos. Foi feito um rodízio, de forma que todas as formas de apresentação informam as 10 notícias (N1 - N10), conforme mostrado na tabela abaixo:

	Grupo 1 AReal	*Grupo 2 AVCoord*	*Grupo 3 AVDescoord*	*Grupo 4 AVSem*	*Grupo5 SOM*
Versão 1	N1,N2	N3,N4	N5,N6	N7,N8	N9,N10
Versão 2	N9,N10	N1,N2	N3,N4	N5,N6	N7,N8
Versão 3	N7,N8	N9,N10	N1,N2	N3,N4	N5,N6
Versão 4	N5,N6	N7,N8	N9,N10	N1,N2	N3,N4
Versão 5	N3,N4	N5,N6	N7,N8	N9,N10	N1,N2

Tab. 4.1: Experimentos para avaliar o sistema

Se a seqüência de apresentações sempre for a mesma para todos os participantes, é provável que isso afete os resultados. Colocar sempre o apresentador virtual com movimentos coordenados ao longo da fala antes das outras apresentações com o apresentador virtual provavelmente evidenciaria

as diferenças entre ele e o que não usa movimentos não-verbais, por exemplo. Assim, a ordem das apresentações é definida aleatoriamente no início de cada avaliação.

Os questionários envolvem avaliar conforto e compreensão ao assistir às apresentações, e envolvem estratégias distintas para cada conceito.

1. conforto: usando escalas de avaliação, o voluntário responde à questão: "quão natural você considera este apresentador?"

   As escalas vão de "muito", valendo 5, até "nada", valendo 0. O valor é inserido numa tabela que se mantém presente ao longo de todas as apresentações, assim as linhas referentes às apresentações já mostradas podem ser alteradas (veja figura 4.2). Afinal, após assistir uma notícia, o participante pode decidir que a notícia anterior merecia outra nota, já que os valores são relativos. Cada linha da tabela corresponde a uma animação, vídeo ou ao arquivo de áudio, sendo que as apresentações que apenas envolvem som não são avaliadas e as linhas correspondentes a elas estão marcadas em cinza indicando que não aceitam entrada do voluntário.

2. compreensão:

   • pede-se ao voluntário que escreva cinco palavras-chave da notícia ouvida.

   Após preencher a tabela que avalia a naturalidade da apresentação, o voluntário aciona um botão que abre uma janela pedindo cinco palavras relevantes da notícia. Elas devem ser inseridas separadas por vírgulas e, caso o voluntário não consiga definir cinco, pode completar com "não sei". Com as palavras-chave, verifica-se se o participante consegue condensar a notícia em poucas palavras.

   Para este teste, considerou-se erradas as palavras que não fossem relacionadas à notícia. Palavras que constassem na notícia e seus sinônimos foram considerados corretos.

   • faz-se três perguntas referentes a algumas palavras ditas no texto como nomes de cidades e números.

   Abaixo, um exemplo de notícia e de perguntas que poderiam ser colocadas para ela:

   Notícia: "Um médico de 68 anos foi preso em flagrante em Uberlândia no Triângulo Mineiro ao tentar vender um laudo que deveria ser fornecido de graça. A prisão aconteceu depois que a

família da paciente denunciou o caso à polícia federal.”

Questões:

- Qual a idade do personagem preso?
- Onde foi feita a prisão?
- Qual a denúncia feita?

As respostas podem ser bastante variadas. Para a segunda pergunta, por exemplo, pode-se responder tanto “Uberlândia” quanto “Triângulo Mineiro” ou mesmo “Minas Gerais” - as três respostas estariam corretas.

As questões de comprensão terminam por se misturar com um teste de memória, pois é possível que se compreenda uma informação mas não se consiga reter dados a respeito dela na memória. Considerando, entretanto, que para uma mesma pessoa o fator “memória” irá influenciar de forma semelhante todas as respostas de suas questões, o resultado não será prejudicado.

Os textos apresentados e as perguntas colocadas para eles podem ser encontrados no Apêndice A.

O teste de avaliação levou em média 15 minutos para ser realizado. A figura 4.2 mostra o ambiente de teste visto pelos voluntários.

4.1.1 Procedimento

A cada participante perguntou-se antes do início do teste se sua visão e audição não apresentavam problemas, de modo a assegurar que não haveria variáveis indesejadas influenciando o resultado. Em seguida, o voluntário recebia as instruções por escrito e, para evitar dúvidas, uma pequena demonstração da avaliação era feita.

A demonstração consistiu em mostrar apenas uma apresentação do apresentador virtual e responder todas as questões para ela. Essa apresentação não faz parte das que aparecem no teste. Além de assegurar que os voluntários compreenderam o procedimento de avaliação, a demonstração permite que o participante se familiarize com o apresentador virtual.

Fig. 4.2: Ambiente de teste

Após a demonstração, os voluntários eram deixados sozinhos em uma sala isolada. Todos fizeram a avaliação sob as mesmas condições, utilizando uma montagem que assegurou uma imagem clara e fones de ouvido com qualidade de som satisfatória.

Para validar o próprio sistema de testes, foram feitos pré-testes com um número reduzido de voluntários de forma a verificar que os questionários estavam claros. Os pré-testes são idênticos aos testes e, caso os voluntários tivesssem dúvidas em relação ao procedimento ou considerassem alguma questão ambígua, o teste seria alterado para evitar a contaminação dos resultados por fatores indesejados. Os pré-testes mostraram que as perguntas estavam claras e que com as explicações por escrito e a demonstração não havia dúvidas em relação aos procedimentos. Assim, seus resultados foram incorporados aos outros resultados.

Ao todo, participaram dos testes 40 pessoas, 8 mulheres e 32 homens, estudantes de graduação ou pós-graduação, com idade variando de 22 a 40 anos, sendo a idade média 28,3 anos. Nenhum dos participantes era especialista na área de animação facial.

4.2 Análise dos resultados

As respostas dadas pelos voluntários foram armazenadas automaticamente em arquivos de texto e processadas individualmente. Uma vez que o teste envolve modalidades de respostas diferentes, foram feitas análises distintas. Para a tabela preenchida com números, calculou-se a média deles de modo a encontrar uma nota para cada tipo de apresentação. Para as palavras-chave e as respostas às questões, calculou-se a quantidade de respostas corretas. As respostas dadas pelos participantes a todos os testes podem ser encontradas no Apêndice A.

Uma vez que o objetivo da avaliação é comparar as diferentes formas de apresentação, os resultados absolutos não têm significado quando analisados isoladamente. Foi feita portanto uma análise comparativa. É importante, entretanto, verificar a significância estatística (ou p-valor) dos resultados, que avalia a probabilidade de as diferenças entre os grupos terem ocorrido ao acaso. Aplicou-se a análise de variância (ANOVA) simples para garantir a qualidade dos resultados.

A ANOVA permite a análise de vários grupos de observação com relação a um determinado fator [60]. No caso tratado aqui, os grupos de observação são os tipos diferentes de apresentador e o fator de interesse é, para o teste da tabela, quão natural o apresentador parece e, para as questões, a facilidade em compreender a notícia.

Testam-se as hipóteses:

- H_0 (hipótese nula) - as médias são iguais
- H_1 (hipótese alternativa) - as médias não são iguais

Para tanto calcula-se a soma dos quadrados dentro de cada grupo SQ_{int} e entre os grupos, SQ_{ent},

$$SQ_{int} = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x_j})^2 \tag{4.1}$$

$$SQ_{ent} = n \sum_{j=1}^{k} (\bar{x_j} - \bar{x})^2 \tag{4.2}$$

sendo x_{ij} o valor da amostra i do grupo de observação j, $\bar{x_j}$ a média amostral do grupo j, $\bar{x}$ a média amostral global de todos os grupos e n o número de amostras de cada grupo.

A média amostral do grupo j é dada por:

$$\bar{x_j} = \sum_{i=1}^{n} x_{ij}/n \tag{4.3}$$

A média amostral global é dada por:

$$\bar{x} = \sum_{j=1}^{k} \bar{x_j}/k \tag{4.4}$$

Em seguida calculam-se os graus de liberdade,

$$gl_{ent} = (k-1) \tag{4.5}$$

$$gl_{int} = k(n-1) \tag{4.6}$$

sendo k o número de grupos existentes e n o número de amostras em cada grupo.

Com estes valores é possível calcular a média dos quadrados:

$$MQ_{ent} = \frac{SQ_{ent}}{gl_{ent}} \tag{4.7}$$

$$MQ_{int} = \frac{SQ_{int}}{gl_{int}} \tag{4.8}$$

A divisão da média dos quadrados entre grupos pela média dos quadrados intra-grupo resulta em:

$$F = \frac{MQ_{ent}}{MQ_{int}} \tag{4.9}$$

A partir de uma tabela com os valores de F críticos da distribuição de Fischer [62], obtém-se o valor de F crítico para um determinado grau de probabilidade (p-valor), dados os graus de liberdade gl_{int} e gl_{ent}. Se $F_{crit} < F$, então o F-teste é significativo para o grau de probabilidade especificado e a hipótese nula pode ser rejeitada. Foi utilizado o nível de significância de 0,05 [60], tradicionalmente aceito e empregado em análises estatísticas.

A análise de variância pode ser aplicada a uma grande quantidade de estudos. Não há limites em relação ao número de grupos ou condições que devem ser comparadas.

4.2.1 Naturalidade

Para o teste que avalia o conceito “naturalidade”, calculou-se a média das notas dadas para cada tipo de apresentação, obtendo os resultados mostrados na tabela abaixo.

Apresentador	*Nota Média*	*Variância*
AReal	4,5000	0,7342
AVCoord	3,5500	0,9089
AVSem	3,0125	1,3796
AVDescoord	2,4625	0,9606

Tab. 4.2: Média e variância das notas da avaliação de naturalidade

A figura 4.3 apresenta as notas atribuídas pelos participantes a cada tipo de apresentação.

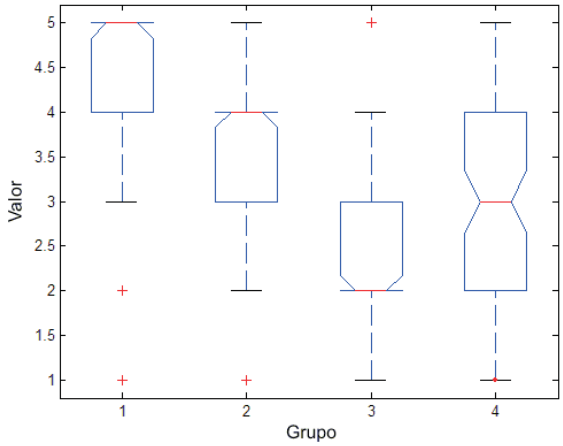


Fig. 4.3: *Box plot* das notas atribuídas à naturalidade de cada um dos quatro apresentadores

Usando a ferramenta de análise de variância (ANOVA), verificou-se que existe uma relação significativa entre o tipo de apresentador e quão realista cada um deles é considerado.

Fonte de variabilidade	*SQ*	*gl*	*MQ*	*F*	*p*
Entre Grupos	180,813	3	60,2708	60,52	<0,0001
Intra Grupos	314,675	316	0,9958		
Total	495,488	319			

Tab. 4.3: Tabela de resultados para a ANOVA

A probabilidade p indica a probabilidade de todos os apresentadores terem o mesmo valor de "realismo" e, para este caso, obteve-se p<0,0001, de forma que os resultados são considerados estatisticamente significativos e pode-se descartar a hipótese nula. Ou seja, pelo menos um grupo possui média diferente dos demais. Para verificar quais os grupos que apresentam médias diferentes entre si, realiza-se testes múltiplas comparações, como o teste de Tukey HSD (*Honestly Significant Difference*) ou o teste de Fischer [59] [41].

O teste de Tukey realizado [42] detectou que há diferenças significativas entre todos os pares comparados, ou seja:

Real *vs* Virtual Coord - p<0,01

Real *vs* Virtual Sem - p<0,01

Real *vs* Virtual Descoord - p<0,01

Virtual Coord *vs* Virtual Sem - p<0,01

Virtual Coord *vs* Virtual Descoord - p<0,01

Virtual Sem *vs* Virtual Descoord - p<0,01

4.2.2 Compreensão

Palavras-chave

A tabela 4.4 apresenta a média de palavras-chaves corretas para cada grupo. Os resultados são apresentados graficamente pela figura 4.4. O número de palavras-chave corretas foi normalizado para valores entre 0 e 1. Um valor unitário indica que o participante acertou 100% das palavras-chaves, ou seja, indicou 5 palavras-chave corretas.

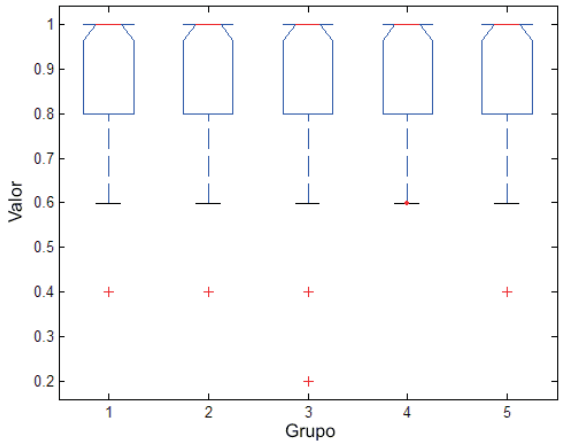


Fig. 4.4: *Box plot* do número normalizado de palavras-chaves corretas para cada tipo de apresentação

Verificou-se que, para todos os apresentadores, a grande maioria das palavras-chave inseridas condizia com a notícia informada. Houve mesmo, freqüentemente, o uso de sinônimos confirmando que as notícias foram realmente compreendidas e retidas na memória. Houve alguns erros, que se distribuíram de forma semelhante para todos os testes, não havendo concentração muito maior para um determinado apresentador.

A probabilidade de a hipótese nula estar correta é de 0,3513, que, por ser um valor alto, não permite descartá-la. Assim, não é possível afirmar que houve influência marcantemente diferenciada dos tipos de apresentação na enumeração das palavras-chave.

Apresentador	*média*	*variância*
Real	0,8805	0,0319
Virtual Coord	0,9065	0,0259
Virtual Sem	0,9143	0,0205
Virtual Descoord	0,8805	0,0319
Som	0,9221	0,0191

Tab. 4.4: Média e variância do número de acertos de palavras-chave normalizado entre [0, 1]

Fonte de variabilidade	*SQ*	*gl*	*MQ*	*F*	*p*
Entre Grupos	0,1147	4	0,02868	1,11	0,3513
Intra Grupos	9,81506	380	0,02583		
Total	9,92977	384			

Tab. 4.5: Tabela de resultados para a ANOVA (palavras-chave corretas)

Questões específicas

A figura 4.5 e a tabela 4.6 apresentam o número de respostas corretas às questões específicas para cada grupo. O número de respostas corretas foi normalizado para valores entre 0 e 1. Um valor unitário indica que o participante acertou 100% das perguntas.

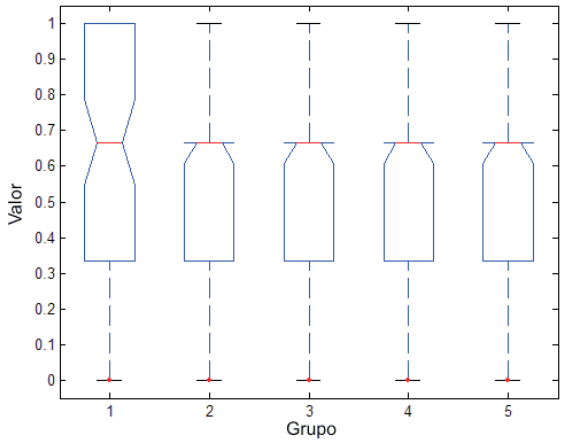


Fig. 4.5: *Box plot* do número normalizado de respostas corretas às questões específicas

A ANOVA para as questões específicas de cada notícia mostrou que a probabilidade de os participantes terem a mesma capacidade de acertar um número parecido de respostas para todos os tipos de

Apresentador	*média*	*variância*
Real	0,5702	0,1239
Virtual Coord	0,5526	0,0905
Virtual Sem	0,5614	0,0954
Virtual Descoord	0,5132	0,0872
Som	0,5219	0,1047

Tab. 4.6: Média e variância do número de respostas corretas normalizado entre [0, 1]

Fonte de variabilidade	*SQ*	*gl*	*MQ*	*F*	*p*
Entre Grupos	0,1901	4	0,04751	0,47	0,7552
Intra Grupos	37,6345	375	0,10036		
Total	37,8246	379			

Tab. 4.7: Tabela de resultados para a ANOVA (acertos nas questões específicas)

apresentação é alta, de 0,7552. Assim, não é possível concluir que os diferentes usos de movimentos não-verbais entre os apresentadores alteraram a capacidade de retenção de informações, apesar de o apresentador com movimentos não-verbais colocados de forma confusa ao longo da fala ter obtido um valor de acertos ligeiramente menor que os outros.

4.3 Considerações Finais

A realização desta avaliação para o apresentador comprovou o relato de que validar características de um ECA é uma atividade que envolve desafios de naturezas distintas. Vale discutir brevemente sobre dois deles, o grupo de teste e a análise de dados.

O grupo de teste foi composto por pessoas com alto grau de escolaridade, o que pode ter tido influência nos resultados. Caso se desejasse introduzir um apresentador virtual no mercado por exemplo, seria necessário fazer toda uma pesquisa definindo o público-alvo e realizar as experiências com pessoas pertencentes a ele. No caso relatado neste capítulo, entretanto, pretendia-se apenas conferir a contribuição dos movimentos não-verbais especificamente para o apresentador de notícias desenvolvido neste trabalho. Assim, uma comparação entre diversos usos de tais movimentos já permite algumas conclusões, mesmo que os voluntários pertençam a um grupo específico.

A análise de dados envolveu a realização de uma média simples para os números do teste de naturalidade e uma verificação cuidadosa para as palavras do teste de compreensão. Possivelmente haveria pequenas discrepâncias entre duas pessoas que definissem quais palavras estão corretas e quais não. O teste em que os participantes deram uma nota para a naturalidade da apresentação, por envolver valores numéricos permitiu uma análise objetiva. Os resultados para este caso confirmaram que os movimentos não-verbais coordenados tornaram a apresentação mais natural, enquanto que os movimentos descoordenados foram considerados mais artificiais do que o apresentador sem movimentos acompanhando a fala. A análise de variância confirmou que o resultado é estatisticamente significativo.

Vale comentar que uma mudança que possivelmente contribuiria para facilitar a análise de dados e torná-la mais objetiva seria utilizar apenas questões fechadas, ou seja, oferecer um número limitado de respostas para as questões entre as quais o participante selecionaria uma, ou oferecer um conjunto de palavras para a definição de palavras-chave.

Havia a expectativa de que o uso de movimentos não-verbais complementando a fala iriam tornar o texto mais claro, facilitando a compreensão do que foi dito, o que não se concretizou.

A curiosidade que as pessoas têm de dar mais atenção aos detalhes da animação do que à apresentação da notícia de um modo geral pode ter interferido com os resultados. Assim, poderia haver mais algumas notícias dadas pelo apresentador virtual, bem no início do teste, cujos resultados não seriam levados em conta, para os participantes se "acostumarem" com ele. Afinal, apesar de ter sido mostrada uma apresentação com o apresentador virtual na fase de demonstração, ela era curta (por volta de 15 segundos).

Entretanto, os resultados para o apresentador real e para o som, que deveriam ser as condições "de controle" também foram muito parecidos entre si e com todos os outros. Talvez, se todas as notícias fossem mais compridas, houvesse alguma diferença mais acentuada. Para o teste feito, no entanto, a maior diferença para respostas corretas é de 5% do apresentador virtual sem movimentos em relação ao apresentador virtual com movimentos confusos.

Outra possibilidade para o teste de compreensão ter sido inconclusivo é que, por se tratar de notícias neutras e com bastante informação dada pela entonação na fala, ao contrário de uma fala em que se usa expressão de emoção e movimentos dícticos que podem alterar completamente o que se

entende, os movimentos não-verbais acabam não influenciando tanto na compreensão e no que estava sendo dito. Assim, os movimentos confusos podem causar estranhamento, mas não o suficiente para comprometer a compreensão. Caso os textos fossem de natureza emocional e complementados por expressões de emoção, é bem possível que os movimentos não-verbais contribuissem de maneira mais evidente para a interpretação da fala e do estado emocional do falante.

Apesar de apenas um dos testes ter se mostrado conclusivo, a avaliação permite que se tenha uma expectativa otimista em relação ao apresentador relatado neste trabalho. Afinal, os resultados inconclusivos mostram que o apresentador virtual alcança resultados semelhantes aos do apresentador real, ou seja, ele não incorpora estímulos negativos à apresentação. O resultado do teste de conforto (naturalidade) mostra que existe uma clara preferência pelo apresentador com movimentos coordenados. Verifica-se deste teste também que os movimentos estão naturais o suficiente para fazerem uma diferença - seja ela positiva ou negativa - em relação ao apresentador sem movimentos não-verbais.

Capítulo 5

Conclusão

O desenvolvimento de ambientes virtuais vem facilitando diversas atividades como aulas à distância, aulas imersivas em que há interação com equipamentos específicos e elaboração de filmes e jogos, entre outros. Criar os elementos que permeiam tais ambientes, entretanto, é objeto de estudos que possuem muitas frentes complementares. O desenvolvimento de personagens virtuais por si só já envolve uma ampla gama de detalhes.

Agentes conversacionais personificados (ECAs - *Embodied Conversational Agents*) são personagens virtuais através dos quais se busca reproduzir os mecanismos utilizados na comunicação face-a-face humana: gestos com mãos e braços, expressões faciais e mudanças na postura, além da fala.

Criar um ECA com aspecto realista exige atenção não apenas à modelagem de sua aparência mas também ao seu comportamento. Deve haver uma preocupação com a maneira como o personagem anda, fala, olha, move os braços, enfim, suas peculiaridades.

No presente trabalho, relatou-se a criação de um ECA com função de apresentador de notícias que envolve a exposição de informações, feita de uma forma isenta. Não há expressão de emoções, mas há movimentos que acompanham a fala, de modo a torná-la mais clara, chamados movimentos não-verbais.

Os movimentos não-verbais implementados que auxiliam a estruturar o texto são os pontuadores, que evidenciam pausas, e ilustradores, que enfatizam determinadas palavras. Para aumentar a sensação de realismo também foram inseridos manipuladores, que têm função biológica, no caso, o piscar de olhos para umedecê-los. Foram implementados elevação de sobrancelhas, piscar e semicerrar de

olhos, rotação dos globos oculares e do pescoço além dos movimentos verbais.

Este trabalho se insere em um projeto maior de desenvolvimento de um ECA completo, equipado com as possíveis formas de comunicação usadas pelos humanos, com cabeça e corpo articulados. O projeto foi iniciado com o desenvolvimento de um algoritmo para a fala [16] que propõe um modelo para movimentação dos lábios e maxilar aplicando o conceito de contexto fonético. O termo contexto fonético se refere ao fato de um segmento sonoro ter sua representação visual (chamada visema) alterada por segmentos vizinhos. Assim, quando se diz “ra” e “ro”, apesar de o fone “r” ser usado em ambas as sílabas, ele é afetado pela vogal seguinte e é representado por visemas distintos.

O algoritmo de sincronismo labial foi aplicado a um modelo que difere do usado em [16] por possuir uma malha mais refinada e ser texturizado. Os movimentos não-verbais são realizados utilizando parametrização geométrica, técnica de animação que envolve a movimentação de vértices de acordo com funções dedicadas.

Apesar da parte móvel do modelo virtual se resumir à sua cabeça, para caracterizar sua função como apresentador de notícias, sua aparência foi complementada por um corpo texturizado com paletó e gravata e um modelo de cabelo. O fundo mostra um ambiente de telejornal. Ainda para manter o apresentador virtual parecido com um real, foram gravadas apresentações de notícias de um canal de televisão brasileiro para permitir uma observação cuidadosa da atuação do apresentador durante a informação de notícias.

A entrada do sistema é dada pelo usuário através de um arquivo baseado em XML, que consiste no texto da notícia a ser apresentada oralmente e nos comandos dos movimentos que o acompanham. Uma pessoa sem treinamento pode facilmente operar o sistema através da adição de marcação ao texto. A marcação consiste em envolver o texto que deve ser afetado por um movimento com comandos que descrevem tal movimento. Assim, o usuário entra com o texto a ser apresentado oralmente e os movimentos que irão enfatizar e pontuar a fala. No sistema, alguns movimentos são inseridos automaticamente, de acordo com estudos feitos envolvendo o comportamento de um apresentador de notícias real.

A análise do comportamento de um apresentador real auxiliou a coordenar os movimentos com a fala de forma mais realista usando o arquivo de entrada e permitiu derivar regras que descrevem movimentos que se repetem em determinadas circunstâncias. Usando essas regras, o sistema insere

movimentos automaticamente. O piscar de olhos, por exemplo, é inserido sem o comando do usuário, de acordo com a média e desvio padrão do piscar do apresentador real. Também o final da apresentação é marcado com um aceno de cabeça.

Os estudos do apresentador real de uma maneira geral, corroboram os da literatura [55], apesar de haver divergências em alguns detalhes, como a freqüência do piscar e também a movimentação dos olhos, que estão sempre fixos num ponto da tela, enquanto que nos estudos relatados nas referências, os olhos estão em constante movimento. Essas diferenças se devem ao fato de o personagem estar numa situação específica, em que faz parte do papel do apresentador tomar a atitude de olhar para a câmera e talvez isso altere também seu padrão de piscar.

Foi feito um experimento para avaliar a contribuição dos sinais não-verbais implementados no conforto do espectador e para verificar sua contribuição para uma melhor compreensão da fala. Essa validação experimental contou com a participação de 40 pessoas. Foram mostradas apresentações de um apresentador real, do apresentador virtual com movimentos colocados coordenadamente ao longo da fala, com movimentos colocados de forma descoordenada e sem movimentos, além de uma apresentação que envolveu apenas o áudio. Após cada apresentação solicitou-se ao participante que respondesse perguntas a respeito da notícia e desse uma nota à apresentação de acordo com quão natural ela lhe parecia.

Usou-se como ferramenta de validação estatística dos resultados a ANOVA (análise de variância). A análise indicou que o uso de movimentos não-verbais pode fazer uma diferença positiva para a avaliação da naturalidade da apresentação quando colocados coordenadamente ao longo da fala ou uma diferença negativa, no caso contrário. Não foi possível verificar diferenças para a compreensão da notícia apresentada, pois os resultados foram muito semelhantes. Apesar de a literatura [55] afirmar que os movimentos não-verbais de ECAs facilitam o entendimento do que é dito, este experimento não comprovou isso, talvez por ter todas as informações presentes no arquivo de áudio.

O aprofundamento de estudos sobre comportamentos de apresentadores de notícias para amadurecer o iniciado aqui é deixado como trabalho futuro. Tomar diversos apresentadores que assumam posturas isentas durante suas apresentações auxiliaria a deixar o modelo de comportamento mais robusto. Outros detalhes, como a verificação de diferenças entre apresentadores homens e mulheres, o tempo de sustentação de um movimento e o uso de expressões emocionais contribuiriam com o

estudo. Isso permitiria o mapeamento de diferentes padrões de comportamento. Vale ressaltar que o conceito de modelo de comportamento pode ser aplicado para personagens com outras funções como um professor, vendedor ou guia.

O exemplo da avaliação da contribuição de movimento não-verbal para diferentes línguas, dado na seção 2.3.1, indica que o comportamento do apresentador pode depender da cultura e da língua usadas para informar as notícias. A mudança da entonação ao falar diferentes línguas afeta a movimentação não-verbal. O alcance da influência para diferentes línguas, portanto, também exige um aprofundamento das pesquisas a respeito do comportamento.

Foi desenvolvido, como parte do projeto maior de ECA, um modelo biomecânico, que poderia substituir o método de animação utilizado neste trabalho, de parametrização geométrica, por ser mais flexível [14]. Também a velocidade dos movimentos poderia ser alterada para permitir uma maior diversidade na realização dos movimentos; no momento, a movimentação ocorre de forma linear.

O sistema atual pode ser facilmente estendido pela adição de novos movimentos faciais ao repertório, permitindo expressão emocional. Assim, o arquivo de entrada pode passar a incluir esses novos gestos e eles podem ser usados em um novo modelo de comportamento. Outra contribuição de grande valia é a adição de movimentos da língua, pois manter a língua apenas acompanhando os movimentos do maxilar reduz a credibilidade do apresentador, afinal, há segmentos sonoros como ’l’ e ’t’ que dependem da movimentação dela. Pode-se também incrementar o potencial de comunicação do ECA com o uso de um corpo articulado e o acréscimo de gestos com os braços e mudanças de postura.

O uso de textura permite uma representação mais realista de personagens virtuais e, para melhorar esta impressão, o uso de rugas é de grande auxílio. Já há um trabalho em andamento que permitirá adicionar rugas ao sistema aumentando a expressividade dos movimentos [56].

Personagens virtuais realistas têm muitas aplicações em potencial. Comprovar isso exige a criação de um ECA munido das habilidades multimodais presentes na comunicação entre humanos e uma série de testes de validação. Tanto o modelamento dos movimentos e a inteligência por trás deles quanto o planejamento dos testes não são triviais, mas o uso de diversas abordagens permite, aos poucos, uma convergência em direção às soluções mais apropriadas.

Referências Bibliográficas

[1] Albrecht, I.; Haber, J.; Kähler, K.; Schröder, M.; Seidel, H.-P. May I talk to you? - Facial Animation From Text. In *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*, 2002.

[2] Albrecht, I.; Haber, J.; Seidel, H.-P. Automatic Generation of Non-verbal Facial Expressions from Speech. In *IEEE Proceedings Computer Graphics INT*, pages 283–293, 2002.

[3] Arafa, Y.; Mamdani, A. Scripting Embodied Agents Behaviour with CML: Character Markup Language. *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 313–316, 2003.

[4] Barbosa, P. A.; Violaro, F.; Albano, E. C.; Simões, F.; Aquino, P.; Madureira, S.; Françoso, E. Aiuruetê: A high-quality concatenative texto-to-speech system for Brazilian Portuguese with demisyllabic analysis-based units and a hierarchical model of rhythm production. In *Proceedings of the 6th European Conference on Speech Communication and Technology - Eurospeech'99*, pages 2059–2062. ISCA, Setembro 1999.

[5] Blascovich, J.; Loomis, J.; Beall, A. C.; Swinth, K. R.; Hoyt, C. L.; Bailenson, J. N. Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry*, 13:103–124, 2002.

[6] Buisine, S.; Abrilian, S.; Martin, J.-C. *From Brows to Trust*, chapter Evaluation of Multimodal Behaviour of Embodied Agents, pages 217–238. Springer, 2005.

[7] Cassell, J.; Bickmore, T.; Campbell, L.; Vilhjálmsson, H.; Yan, H. *Embodied Conversational Agents*, chapter Human Conversation as a System Framework: Designing Embodied Conversational Agents, pages 29–63. The MIT Press, 2000.

[8] Cassell, J., Sullivan J.; Prevost, S; Churchill, E. F. *Embodied Conversational Agents*. MIT Press, 2000.

[9] Cassell, J.; Vilhjálmsson, H. H.; Bickmore, T. Beat The behavior expression animation toolkit. In *Proceedings of SIGGRAPH 2001*, pages 477–486, 2001.

[10] Chau, M.; Betke, M. Real Time Eye Tracking and Blink Detection with USB Cameras. *Boston University Computer Science Technical Report*, 2005.

[11] Christoph, N. *From Brows to Trust*, chapter Empirical Evaluation Methodology for Embodied Conversational Agents, pages 67–90. Springer, 2005.

[12] Collier, G. *Emotional Expression*. Lawrence Erlbaum Associates, 1985.

[13] Coombs J. H.; Renear A. H.; DeRose S. J. Markup systems and the future of scholarly text processing. *Communications of the Association for Computing Machinery - CACM*, 1987.

[14] Corrêa, R. *Animação facial por Computador baseada em modelagem biomecânica*. Tese de mestrado, Faculdade de Engenharia Elétrica e de Computação, UNICAMP, 2007.

[15] Curtin University of Technology. VHML Homepage. Página na internet, 2008. http://www.interface.computing.edu.au/documents/VHML/ (último acesso em outubro/2008).

[16] De Martino, J. M. *Animação facial sincronizada com a fala: visemas dependents do contexto fonético para o português do Brasil*. Tese de doutorado, Faculdade de Engenharia Elétrica e de Computação, UNICAMP, Julho 2005.

[17] De Martino, J. M.; Magalhães, L. P.; Violaro, F. Facial Animation Based on Context-Dependent Visemes. *Computers and Graphics*, 30(6):971–980, Dezembro 2006.

[18] De Martino, J. M.; Violaro, F. Benchmarking speech synchronized facial animation based on context-dependent visemes. In *WSCG '07: Full Papers Proceedings of The 15th International*

Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2007, volume 1, pages 105–112, University of West Bohemia, Pilsen, República Tcheca, 29 - 01 February 2007.

[19] Deng, Z.; Noh, J. *Computer Facial Animation: A Survey*, chapter Data-Driven 3D Facial Animation, pages 1–28. Springer London, 2007.

[20] Duncan, S. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2):283–292, 1972.

[21] Ekman, P. *Human ethology*, chapter About brows: emotional and conversational signals, pages 169–248. CRC Press, 1979.

[22] Ekman, P.; Friesen, W. V. Facial Action Coding System. *Consulting Psychologist Press, Palo Alto*, 1978.

[23] Fuchs, T.; Haber, J.; Seidel, H.-P. MIMIC A Language for Specifying Facial Animations. In *WSCG 2004*, 2004.

[24] Garchery, S.; Giacomo, T. Di; Magnenat-Thalmann, N. *Real-Time Adaptive Facial Animation* , chapter Data-Driven 3D Facial Animation, pages 217–247. Springer London, 2007.

[25] Geroch, M. S. Motion capture for the rest of us. *Journal of Computing Sciences in Colleges*, 19(3):157–164, 2004.

[26] Granström, B.; House, D. Measuring and modeling audiovisual prosody for animated agents. In *Proceedings of Speech Prosody 2006 Dresden*, 2006.

[27] Huang, Z.; Eliens, A.; Visser, C. STEP A Scripting Language for Embodied Agents. In *Proceedings of the Workshop on Lifelike Animated Agents 2002*, Tokyo, 2002.

[28] Ishizuka Lab. MPML Homepage. Página na internet, 2008. http://www.miv.t.u-tokyo.ac.jp/MPML/en/ (último acesso em outubro/2008).

[29] Kalra, P.; A. Mangili; N. Magnenat-Thalmann; D. Thalmann. Simulation of Facial Muscle Actions Based on Rational Free Form Deformations. In *Computer Graphics Forum (EUROGRAPHICS '92 Proceedings)*, pages 59–69, 1992.

[30] Kalra, P. K. *An Interactive Multimodal Facial Animation System*. Tese de doutorado, Département d'informatique, Institut des systèmes informatiques et multimédias ISIM - Ecole polytechnique fédérale de Lausanne EPFL, 1991.

[31] Kähler, K.; Haber, J.; Seidel, H.-P. Geometry-based Muscle Modeling For Facial Animation. In *Proc. Graphics Interface'01*, pages 37–46, 2001.

[32] Kopp, S.; Krenn, B; Marsella, S.; Marshall, A. N.; Pelachaud, C.; Pirker, H.; Thórisson, K. R.; Vilhjálmsson, H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *IVA*, pages 205–217, 2006.

[33] Krahmer, E.; Swerts, M. *From Brows to Trust*, chapter More About Brows, pages 191–216. Springer, 2005.

[34] Krenn, B. The NECA Project: Net Environments for Embodied Emotional Conversational Agents. In *Proc. of Workshop on emotionally rich virtual worlds with emotion synthesis at the 8th International Conference on 3D Web Technology*, Março 2003.

[35] Kshirsagar S. et al. Avatar Markup Language. *Proc. of 8th Eurographics Workshop on Virtual Environments*, pages 169–177, 2002.

[36] Kunc, L.; Kleindienst, J. *Text, Speech and Dialogue*, chapter ECAF: Authoring Language for Embodied Conversational Agents, pages 206–213. Springer, 2007.

[37] Lee, S.-H; Terzopoulos, D. Head Up! Biomechanical Modeling and Neuromuscular Control of the Neck. In *Proceedings of SIGGRAPH 2006*, pages 1188–1198, 2006.

[38] Lee, Y.; Terzopoulos, D.; Waters, K. Realistic Modeling for Facial Animation. In *Computer Graphics Proceedings, Annual Conference Series*, pages 55–62, 1995.

[39] Lewis, M.; Jacobson, J. Introduction: Game Engines in Scientific Research. *Comm. ACM*, 45(1):27–31, 2002.

[40] Linden Research. Secondlife Homepage. Página na internet, 2008. http://secondlife.com (último acesso em outubro/2008).

[41] Lowry, R. Independent Samples - One-way ANOVA. Página na internet, 2008. http://faculty.vassar.edu/lowry/ch14pt2.html (último acesso em outubro/2008).

[42] Lowry, R. One-way ANOVA. Página na internet, 2008. http://faculty.vassar.edu/lowry//anova1u.html (último acesso em outubro/2008).

[43] Magnenat-Thalmann, N.; Primeau, E.; Thalmann, D. Abstract muscle action procedures for human face animation. *The Visual Computer*, 3(5):290–297, Março 1988.

[44] Massaro, D. W.; Cohen, M. M.; Beskow, J.; Daniel, S.; Cole, R. A. *Embodied Conversational Agents*, chapter Developing and Evaluating Conversational Agents, pages 287–318. The MIT Press, 2000.

[45] Mesquita, R. M. Comunicação não-verbal: Relevância na atuação Profissional. In *Revista Paulista de Educação Física*, pages 155–163, 1998.

[46] Nowak, K. J.; Biocca, F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, 2003.

[47] Parke, F. I. *Computer generated animation of faces*. Tese de mestrado, University of Utah, Salt Lake City, Junho 1972.

[48] Parke, F. I. *A Parametric Model for Human Faces*. Tese de doutorado, Dezembro 1974.

[49] Parke, F. I. Parametrized Models for Facial Animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, novembro 1982.

[50] Parke, F. I.; Waters, K. *Computer Facial Animation*. AK Peters, Ltd., 1996.

[51] Pelachaud, C.; Badler, N. I.; Steedman, M. Generating Facial Expressions for Speech. *Cognitive Science, A Multidisciplinary Journal of Artificial Intelligence, Linguistics, Neuroscience, Philosophy*, pages 1–47, 1996.

[52] Pirker, H.; Krenn, B. Assessment of Markup Languages for Avatars, Multimedia and Multimodal Systes. Technical Report D9c, Austrian Research Institute for Artificial Intelligence, 2002. http://www.oefai.at/NECA/ (último acesso em outubro/2008).

[53] Piwek P.et al. RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA. In *Proceceedings of the AAMAS workshop on "Embodied conversational agents - let's specify and evaluate them!"*, 2002.

[54] Platt, S. M. *A system for computer simulation of the human face*. Tese de mestrado, University of Pennsylvania, 1980.

[55] Poggy, I; Pelachaud, C. *Embodied Conversational Agents*, chapter Performative Facial Expressions in Animated Faces, pages 155–188. The MIT Press, 2000.

[56] Reis, C. D. G.; De Martino, J. M.; Batagelo, H. Real-time simulation of wrinkles. In *WSCG '08: Communicatin Papers Proceedings of The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2007*, pages 109–115, University of West Bohemia, Pilsen, República Tcheca, 04 - 07 February 2008.

[57] Rist, T.; Andre, E.; Baldes, S.; Gebhard, P.; Klesen, M.; Kipp, M.; Rist, P.; Schmitt, M. *Life-like Characters: Tools, Affective Functions, and Applications*, chapter A Review on the Development of Embodied Presentation Agents and their Application Fields, pages 377–404. CRC Press, 2003.

[58] Rosenbloom, A. Introduction: A Game Experience in Every Application. *Comm. ACM*, 46(7):28–31, 2003.

[59] Rumsey, D. *Intermediate Statistics for Dummies*. For Dummies, 2007.

[60] Rutherford, A. *Introducing Anova and Ancova: a GLM approach*. SAGE, 2001.

[61] Ruttkay, Zs.; Dormann, C.; Noot, H. Evaluating ECAs - What and How . In *Proc. of the AAMAS02 Workshop on Embodied Conversational Agents*, 2002.

[62] Soper, D. S. Table of Critical Values of the Fischer F-Distribution. Página na internet, The Statistics Knowledge Base Website, 2008. http://www.danielsoper.com/statkb/topic1001.aspx (último acesso em outubro/2008).

[63] Terzopoulos, D., Waters, K. Physically-based Facial Modelling, Analysis and Animation. *The Journal of Visualization and Computer Animation*, 1:73–80, 1990.

[64] Tschirren, B. Realism and believability in MPEG-4 Facial Models. Página na internet, 2000. http://www.vhml.org/theses/tschirre/html/ (último acesso em outubro/2008).

[65] Ultimate Disney. Pixar Shorts. Página na internet, 2008. http://www.ultimatedisney.com/pixarshorts-vol1.html (último acesso em outubro/2008).

[66] Universo HQ. Final Fantasy. Página na internet, 2008. http://www.universohq.com/cinema/rc14082001_01.cfm (último acesso em outubro/2008).

[67] UOL. Cidade do Conhecimento da USP ganha versão virtual no Second Life. Página na internet, PCWorld, 2008. http://pcworld.uol.com.br/noticias/2007/08/21/idgnoticia.2007-08-21.9565630076/ (último acesso em outubro/2008).

[68] Waters, K. A Muscle Model for Animating Three Dimensional Facial Expression. *Computer Graphics*, 21(4):17–24, julho 1987.

[69] Welbergen, H. van; Nijholt, A.; Reidsma, D.; Zwiers, J. Presenting in Virtual Worlds: An Architecture for a 3D Anthropomorphic Presenter. *IEEE Intelligent Systems*, 21(5):47–53, 2006.

[70] Wikipedia. Final Fantasy: The Spirits Within. Página na internet, 2008. http://pt.wikipedia.org/wiki/Final_Fantasy:_The_Spirits_Within (último acesso em outubro/2008).

[71] Wikipedia. Framework. Página na internet, 2008. http://pt.wikipedia.org/wiki/Framework (último acesso em outubro/2008).

[72] Wikipedia. Institutions and Organizations in SecondLife. Página na internet, 2008. http://simteach.com/wiki/index.php?title=Institutions_and_Organizations_in_SL (último acesso em outubro/2008).

[73] Zong Y.; Dohi, H.; Ishizuka, M. Multimodal Presentation Markup Language MPML with Emotion Expression Functions Attached. *Proc. 2000 Int'l Symp. on Multimedia Software Engineering (IEEE Computer Soc.)*, pages 359–365, 2000.

[74] Zong, Y.; Dohi, H.; Prendinger, H.; Ishizuka, M. Emotion Expression Functions in Multimodal Presentation. In *Proceedings of ICMI 2000*, pages 57–64, 2000.

[75] Zyda, M. Introduction: Creating a Science of Games. *Comm. ACM*, 50(7):26–29, 2007.

Apêndice A

Informações complementares sobre a avaliação

A.1 Notícias apresentadas e questões correspondentes

Em seguida, as notícias colocadas nas experiências de avaliação descritas no capítulo 4 e as questões colocadas para elas.

Fragmento 1: No Maranhão a audácia da empresa Gautama é vista nas dezenas de obras inacabadas no estado. Na BR quatro zero dois, leste do estado, quatro pontes ligando nada a lugar algum.

Perguntas:

1. Sobre qual estado trata a notícia?
2. De que empresa trata a notícia?
3. Em que região do estado estão as pontes mencionadas?

Fragmento 2: O presidente dos Estados Unidos Jorge Bush conseguiu uma importante vitória política. Seguindo os passos da câmara dos representantes, o senado aprovou um projeto de lei que concede cem bilhões de dólares para financiar as guerras no Iraque e no Afeganistão sem um calendário para retirada das tropas americanas.

Perguntas:

1. De quantos dólares se fala na notícia?

2. Quem aprovou o uso de tantos dólares para guerras?

3. Em que países estão as guerras mencionadas?

Fragmento 3: Terminou sem acordo uma reunião entre representantes do governo de São Paulo e estudantes que ocupam a reitoria da USP desde o dia três de maio. Uma nova reunião foi marcada para segunda-feira.

Perguntas:

1. Desde quando a reitoria está ocupada?

2. De qual universidade é a reitoria?

3. Quando será a próxima reunião?

Fragmento 4: Mais de noventa e seis mil inscritos no Bolsa Família no Brasil inteiro estão com o benefício bloqueado ou suspenso. O principal motivo é a falta de comprovação da freqüência escolar dos filhos. Em Irará, na Bahia, comunidades rurais estão preocupadas pelo benefício ser a única renda da família.

Perguntas:

1. São quantos os inscritos no Bolsa família com irregularidade no benefício?

2. Qual o motivo de haver a irregularidade mencionada na notícia?

3. Onde estão as comunidades mencionadas na notícia?

Fragmento 5: Preso em Pernambuco mais um suspeito de fazer parte de um grupo de extermínio que seria responsável por mais de mil assassinatos no nordeste nos últimos cinco anos.

Perguntas:

1. Onde foi feita a prisão noticiada?

2. O preso é suspeito de quantos assassinatos?

3. Durante quantos anos tais assassinatos foram cometidos?

Fragmento 6: O alto número de mortes de bebês recém-nascidos nas maternidades de Alagoas assusta autoridades e mães. As unidades estão superlotadas e a secretaria de saúde do estado reclama que não há verba suficiente para aumentar o número de leitos.

Perguntas:

1. Onde está havendo alta taxa de mortalidade de recém-nascidos?

2. Qual o problema das maternidades?

3. Qual a reclamação da secretaria de saúde do estado?

Fragmento 7 Um médico de sessenta e oito anos foi preso em flagrante em Uberlândia no Triângulo Mineiro ao tentar vender um laudo que deveria ser fornecido de graça. A prisão aconteceu depois que a família da paciente denunciou o caso à polícia federal.

Perguntas:

1. Qual a idade do personagem preso?

2. Onde foi feita a prisão?

3. Qual a denúncia feita?

Fragmento 8: Só este ano já foram registrados quatro mil casos de dengue no Piauí. O estado sofre com uma epidemia, mas a secretaria de saúde alerta que o número pode ser bem maior.

Perguntas:

1. Quantos casos de dengue foram registrados?

2. Onde foi feito esse registro?

3. Quem alertou que o número de casos de dengue pode ser maior?

Fragmento 9: Policiais federais retornam ao trabalho hoje depois de uma paralisação de setenta e duas horas. O governo fechou um acordo para o reajuste salarial da categoria. E os servidores aceitaram a proposta para dividir o reajuste de trinta por cento em três parcelas.

Perguntas:

1. Qual a duração da paralisação?

2. De que categoria se trata na notícia?

3. Qual o valor do reajuste?

Fragmento 10: Há catorze anos não fazia tanto frio em São Joaquim na serra de Santa Catarina. Hoje de manhã, os termômetros marcaram seis graus abaixo de zero.

Perguntas:

1. Onde fez frio?

2. Quando se verificou os termômetros?

3. Há quanto tempo fez um frio igual?

A.2 Resultados dos testes de avaliação

A seguir, as respostas dadas pelos participantes da avaliação para todos os testes. Em A.2.1, as notas dadas para os apresentadores para o questito "naturalidade". A seguir, em A.2.2, as palavras-chave dadas para cada fragmento de notícia e em A.2.3, as respostas às questões específicas para cada notícias. Usa-se as seguintes abreviações, como feito anteriormente no capítulo 4:

- AReal - apresentador real
- AVCoord - apresentador virtual com movimentos não-verbais coordenados
- AVDescoord - apresentador virtual com movimentos não-verbais descoordenados
- AVSem - apresentador virtual sem movimentos não-verbais
- Som - apenas o áudio, sem apresentador

A.2.1 Resultados da questão de naturalidade

Na tabela A.1, as notas dadas pelos participantes aos diferentes apresentadores no quesito "naturalidade".

AReal	*AVCoord*	*AVDescoord*	*AVSem*
5	4	2	4
5	4	3	4
5	4	2	3
5	4	2	3
4	4	1	4
4	3	3	4
5	4	2	3
4	5	4	2
4	3	4	4
5	4	3	5
3	4	1	3
5	2	4	2
5	2	2	3
4	4	1	5
3	3	2	1
5	3	2	2
5	5	2	3
5	3	2	3
5	5	2	5
5	3	2	3
5	4	2	4
5	5	3	3
4	3	3	3
4	3	3	4
5	3	5	5
5	5	3	5
5	3	2	1
5	4	2	1
4	4	3	5
5	4	3	4
5	3	3	2
5	1	4	2
5	4	1	3
5	3	3	3
5	4	3	3
5	4	4	5
5	3	3	2
4	4	3	4
5	4	3	2
5	3	2	1

AReal	*AVCoord*	*AVDescoord*	*AVSem*
5	5	2	5
5	5	2	3
4	4	4	1
2	2	2	3
5	3	4	4
3	3	3	3
5	4	1	3
5	5	3	1
3	2	4	4
2	4	1	5
5	4	1	3
4	2	2	2
3	5	3	4
5	3	2	4
5	2	1	1
5	2	1	2
5	5	3	3
5	3	4	1
4	4	2	2
3	3	2	3
4	4	3	3
4	4	3	4
5	3	3	3
5	3	3	2
5	4	3	2
5	4	4	4
5	4	2	1
5	2	1	3
5	5	1	3
5	4	1	4
4	5	2	2
5	4	2	1
5	4	1	3
5	3	2	2
5	2	1	3
3	3	4	4
5	2	2	2
1	2	3	3
5	4	3	3
4	3	3	1

Tab. A.1: Notas dadas à naturalidade dos apresentadores

A.2.2 Resultados da questão de palavras-chave

Abaixo, as palavras-chave dadas para as notícias expostas na seção A.1 e a indicação de quantas são as corretas entre parênteses, ao final de cada linha.

Fragmento 1

AReal

- Maranhão, pontes, empreiteira, Estado, não sei (4)
- 4, pontes, nada, lugar, algum (5)
- Gautama, audácia, ponte, não sei, não sei (3)
- Maranhao, Gautama, estradas, irregularidades, pontes (5)
- pontes, estado, lugar, nada, algum (5)
- maranhão, gautama, obras inacabadas, pontes, não sei (4)

- fraude, gautama, pontes, irregularidade, BRXXX (4)
- quatro ponte, inacabadas, galtama, não sei, não sei (3)

AVSem

- gautama, audácia, estado, pontes, nada (5)
- Gautama, Paraná, irregularidades, pontes, não sei (3)
- pontes, estado, nada, lugar, algum (5)
- Maranhão, Gautama, ponte, audacia, lugar nenhum (5)
- nada, astúcia, ponte, oeste, não sei (3)
- Gautama, ponte, leste, não sei, não sei (3)
- galtama, nada, lugar algum, estrada, estado (5)
- Maranhao, construtura, ponte, galtama, estado (5)

AVDescoord

- gautama, pontes inacabadas, ousadia, nao sei, nao sei (3)
- pontes, não sei, não sei, não sei, não sei (1)
- pontes, nada, lugar, algum, ligam (5)
- maranhão, ponte, leste do estado, sem ligação, não sei (4)
- 4pontes, nada, lugar algum, leste do estado, não sei (4)
- nada, lugar nenhum, maranhao, ponte, não sei (4)
- ponte, galtama, inacabada, não sei, não sei (3)
- maranhao, estado, ponte, noroeste, liga (4)

AVCoord

- maranhão, ponte, nada, lugar, algum (5)
- nada, lugar algum, br, ponte, não sei (4)
- gautama, maranhao, leste, ousadia, pontes (5)
- Gautama, BR 342, leste do Estado, ligar coisa a lugar nenhum, não sei (4)
- br102, pontes, nada, lugar algum, não sei (4)
- galtama, pontes, estradas, obras, irregularidades (5)

- pontes, ligam, nada, lugar, algum (5)
- Gautama, nordeste, obras inacabadas, nada, lugar algum (5)

Som

- maranhão, galtama, pontes, nada, br402 (5)
- nada, lugar, algum, não sei, não sei (5)
- Maranhão, Gautama, Pontes, Inacabada, Construtora (5)
- estrada, quatro, pontes, ligam, nada (5)
- Ponte, inacabada, gautama, maranhão, leste do estado (5)
- audacia, Gautama, pontes, BR 402, lugar nenhum (5)
- Maranhão, Galtama, obras, pontes, nada (5)
- gautama, audácia, maranhão, pontes, nada (5)

Fragmento 2

AReal

- bush, iraque, eua, 4 bilhoes, tropas americanas (5)
- bush, guerra, iraque, afeganistão, eua (5)
- presidente, americano, tropas, dolares, retirar (5)
- tropas americanas, iraque, afeganistão, dolares, retirada (5)
- George Bush, tropas americanas, Iraque, não sei, não sei (3)
- presidente, iraque, afeganistao, tropa, guerra (5)
- Bush, senado, aprovação, 100 bilhões, guerra (5)
- estados unidos, iraque, afeganistao, tropas americanas, bush (5)

AVSem

- Bush, senado, tropas americanas, Iraque, Afeganistão (5)
- tropas, afeganistão, calendário, 100, americanas (5)
- Tropas, Iraque, calendário, Estados Unidos, Presidente (5)
- iraque, tropas americanas, senado, retirada, aprovacao (5)
- tropas, Bush, Iraque, calendário, americanas (5)

- governo americano, iraque, afeganistao, verbas, guerra (5)
- Bush, financiar, guerra, Iraque, Afeganistão (5)
- tropas, verba, calendário, retirada, bush (5)

AVDescoord

- George Bush, EUA, Afeganistão, tropas, americanas (5)
- tropas, americanas, senado, Bush, não sei (4)
- 100 bilhões, guerras, Iraque, Afeganistão, calendário (5)
- Iraque, EUA, senado, tropas, Bush (5)
- George Bush, Senado Americano, Retirada, Tropas, Calendário (5)
- guerra, calendario, vitoria politica, Bush, Iraque (5)
- guerra, EUA, George Bush, tropas, retirada (5)
- bush, vitória, bilhões, guerra, retirada (5)

AVCoord

- senado, bush, tropas, retirada, aprovação (5)
- gaza, tropas, presidente, senado, Bush (5)
- Bush, Iraque, Afeganistão, guerra, aprovação (5)
- retirada, tropas, calendário, Senado, Bush (5)
- Bush, tropas, USA, 100 milhoes, retirada (5)
- bush, congresso, verba, iraque, tropas (5)
- bush, senado, iraque, afeganistao, tropas (5)

Som

- eua, gerge bush, bilhões, guerra, afeganistão (5)
- bush, afeganistao, bilhoes, guerra, presidente (5)
- George Bush, tropas, senado, Afeganistão, não sei (4)
- tropas americanas, iraque, bush, retirada, não sei (4)
- Bush, guerra, iraque, senado, lei (5)
- Iraque, Afeganistao, EstadosUnidos, tropas, Americanas (5)

- EUA, presidente, guera do Iraque e Afeganistão, tropas americanas, congresso (5)

Fragmento 3

AReal

- Estudantes, USP, Reitoria, não sei, não sei (3)
- reitoria, usp, reunião, acordo, segunda-feira (5)
- usp, ribeirao, greve, segunda-feira, reuniao (4)
- Reitoria, USP, ocupação, estudantes, reunião (5)
- reunião, estudantes, USP, reitoria, invasão (5)
- greve, USP, reitoria, reuniáo, acordo (5)
- acordo, estudantes, usp, governo, rebelião (4)
- usp, invasáo, reitoria, reuniáo, segunda-feira (5)

AVSem

- governo, são paulo, usp, acordo, estudantes (5)
- usp, reuniao, segunda-feira, não sei, não sei (3)
- reuniao, governo, alunos, maio, acordo (5)
- reitoria, ocupação, estudantes, USP, governo de São Paulo (5)
- tres de maio, usp, segunda-feira, reuniao, não sei (4)
- USP, reunião, estudantes, reitoria, não sei (4)
- reitoria, usp, reuniao, acabou, segundafeira (5)
- usp, reitoria, ocupação, estudantes, universidade (5)

AVDescoord

- estudantes, reitoria, reunião, ocupam, terminou (5)
- usp, reitoria, estudantes, terminou, 2a. feira (5)
- Ocupação, reunião, reitoria, USP, segunda-feira (5)
- usp, reitoria, governo, ocupacao, acordo (5)
- ocupação, reitoria, estudantes, maio, três (5)
- estudantes, reitoria, usp, reunião, ocupação (5)

- Reunião, USP, reitoria, estudantes, segunda-feira (5)
- usp, ocupação, reitoria, segunda-feira, reunião (5)

AVCoord

- terminou, acordo, USP, reitoria, nova (5)
- usp, reunião, estudantes, segunda, feira (5)
- estudantes, acordo, reitoria, USP, maio (5)
- reitoria, USP, alunos, reuniao, marcada (5)
- USP, reitoria, reunião, estudantes, não sei (4)
- reuniáo, USP, estudantes, paralizacao, segunda-feira (4)
- Reunião, USP, ocupação, reitoria, alunos (5)
- acordo, reitoria, usp, reunião, maio (5)

Som

- reitoria usp, governo do estado, sem acordo, nova reuniao, ocupacao da reitoria (5)
- usp, estudantes, reitoria, ocupação, reunião (5)
- reitoria, usp, reuniao, ocupação, maio (5)
- são paulo, USP, reunião, segunda-feira, estudantes (5)
- reuniao, usp, estudandes, reitoria, ocupam (5)
- segunda feira, reuniao, sem acordo, usp, 13 maio (5)
- reunião, estudantes, acordo, terminou, maio (5)
- sem acordo, reuniao, reitoria, estudantes, ocupacao (5)

Fragmento 4

AReal

- Bolsa família, comprovação escolar, não sei, não sei, não sei (2)
- irará, bolsa-família, renda, benefício, preocupação (5)
- irara, bahia, bolsa-familia, frequencia, falta (5)
- Bolsa-família, Bahia, comprovação, freqüência,escolar (5)
- renda, bolsa, escola, não sei, não sei (3)

- bolsa familia, irara, não sei, não sei, não sei (2)
- cancelamento, bolsa-família, falta de comprovante escolar, irará, única renda (5)
- irara, bahia, bolsa ,familia, suspenso (5)

AVSem

- bolsa familia, beneficio, comprovação, escola, renda (5)
- irará, bolsa familia, 3000, familias, renda (5)
- irara, bolsa-familia, beneficio, interior, suspensa (5)
- 96 mil famílias, bolsa família, bloqueado, Irará, Bahia (5)
- bolsa família,bahia,irará,suspensas,96mil (5)
- Bolsa-Família, Irará, Bahia, renda, família (5)
- bolsa familia, bahia, preocupado, governo, frequencia escolar (5)
- bolsa família, benefício, renda familiar, freqëncia, escola (5)

AVDescoord

- bolsa-família, benefício, família, frequência escolar, renda (5)
- irará, bahia, bolsa-familia, renda, beneficio (5)
- Bahia, Bolsa-Família, suspenso, rendimento, renda (4)
- bolsa familia, bahia, perda do beneficio, frequencia escolar, preocupação (5)
- bolsa, família, preocupadas, renda, suspensão (5)
- bolsa-familia, corte, irara, renda familiar, não sei (4)
- Bolsa, família, Irará, filhos, renda (5)
- Bolsa-família, suspensa, Irará, comprovar matrícula, Bahia (5)

AVCoord

- bolsa-familia, Irara, beneficio, renda, unico (5)
- bolsa, família, irauá, falta de frequencia, renda (5)
- Bolsa-Família, Irará, Bahia, freqüência, renda (5)
- bolsa família, renda, bahia, escola, creanças (5)
- Irará, Bahia, Bolsa Família, Benefício, preocupados (5)

- bolsa familia, Bahia, renda da familia, 9600, suspensos (5)
- Bolsa-família, benefício, Irará, Bahia, escola (5)
- bolsa família, bloqueada, frequencia, renda, irará (5)

Som

- bolsa familia, irara, bahia, 96 mil incsritos, nao sei (4)
- bolsa família, recurso bloqueado, falta de comprovação, renda, presença na escola (5)
- bolsa, familia, suspenso, frequencia, filhos (5)
- beneficio, bolsa familia, irará, inscritos, não sei (4)
- bolsa familia, irara, comunidade rural, renda, ausencia as aulas (5)
- irara, bolsa familia, preocupados, beneficio, não sei (4)
- bolsa família, comprovação, freqüência, irara, preocupadas (5)
- bolsa familia, renda, frequencia escolar, beneficio,não sei (4)

Fragmento 5

AReal

- nordeste, assassinato, pernambuco, nao sei, nao sei (3)
- grupo extermínio, nordeste, assassinatos, não sei, não sei (3)
- nordeste, mortos, anos, não sei, não sei (3)
- grupo de esterminio, pernambuco, mais de 5 anos, presos, não sei (4)
- nordeste, grupo de exterminio, 1000 assassinatos, 5 anos, presos (5)
- assassinato, nordeste, preso, grupo, não sei (4)
- grupo de extermínio, preso, nordeste, são paulo, não sei (3)
- prisao, grupo exterminio, assassinato, nordeste, 5 anos (5)

AVSem

- Grupo, extermínio, assassinatos, nordeste, preso (5)
- pernambuco, preso, nordeste, assassinato, 5 anos (5)
- grupo, extermínio, assassinato, Nordeste, presos (5)
- Prisao, Pernambuco, assassinato, grupo de exterminio, nordeste (5)

- preso, grupo, assassinatos, 5, anos (5)
- prisão, pernambuco, nordeste, assassinatos, grupo exterminio (5)
- Preso, grupo, exteminio, nordeste, assassinatos (5)
- nordeste, cinco anos, mil assasinatos, presos, não sei (4)

AVDescoord

- nordeste, assassinato, anos, suspeitos, últimos (5)
- extermínio, 5 anos, nordeste, preso, não sei (4)
- grupo, extermínio, assassinatos, nordeste, cinco anos (5)
- assassinato, nordeste, pernambuco, cinco, anos (5)
- Assassinato, Nordeste, Pernambuco, grupo de extermínio, Prisão (5)
- Pernambuco, grupo de exterminio, 5 anos, 1000 assaninatos, não sei (4)
- preso, assassinato, nordeste, Pernambuco, 5 anos (5)
- preso, pernambuco, extermínio, nordeste, assassinatos (5)

AVCoord

- Pernambuco, assassinato, não sei, não sei, não sei (2)
- grupo de extermínio, preso, nordeste, cinco anos, não sei (4)
- pernanbuco, grupo, exterminio, 5 anos, nordeste (5)
- Pernambuco, assassinato, preso, nordeste, mortes (5)
- extermínio, Nordeste, suspeito, preso, não sei (4)
- assassinato, nordeste, pernambuco, 5 anos, não sei (4)
- preso, grupo, extermínio, assassinato, nordeste (5)
- grupo, exterminio, pernambuco, assassinatos, ano (5)

Som

- preso, pernambuco, grupo de esterminio, norderde, suspeito (5)
- nordeste, assassinato, grupo esterminio, não sei, não sei (3)
- grupo, exterminio, comeco do ano, nordeste, naosei (3)
- grupo de extermínio, preso, Nordeste, mil assassinatos, 5 anos (5)

- nordeste, assassinato, não sei, não sei, não sei (2)
- Pernambuco, extermínio, assassino, nordeste, não sei (4)
- pernambuco, assassinatos, nordeste, pernambuco, extermínio (5)
- grupo de esterminio, nordeste, Pernanbuco, prisão, assassinatos (5)

Fragmento 6

AReal

- nascimento, bebes, mães, leitos, verba (5)
- leitos, maes, verbas, aumentar, não sei (4)
- marternidade, numero de leitos, reclamacao, bebes, falta (5)
- leitos, falta, taxa, mortalidade, maternidade (5)
- leitos, verba, prefeitura, aumentar, numero (4)
- leitos, verba, mortes, recém-nascidos, não sei (4)
- Alagoas, maes, leitos, mortes, estado (5)
- maternidade, bebes, mortes, mães, Alagoas (5)

AVSem

- maternidade, Alagos, leitos, autoridades, não sei (4)
- bebês, secretaria, segurança, mortos, não sei (3)
- leitos, Alagoas, bebês, saúde, não sei (4)
- bebês, alagoas, verba, leitos, maes (5)
- Leitos, Mortalidade Infantil, Alagoas, Saúde, não sei (4)
- morte, numero de leitos, verbas, não sei, não sei (3)
- Alagoas, recém-nascidos, mortos, maternidades, leitos (5)
- alagoas, maternidade, mães, verba, leitos (5)

AVDescoord

- falta de leitos, não sei, não sei, não sei, não sei (1)
- secretaria, saúde, mães, leitos, não sei (4)
- secretaria, leitos, estado, aumentar, número (5)

- leitos, verba, secretaria de saude, maternidade, pacientes (5)
- leitos, maternidade, Alagoas, mães, secretaria de Saúde (5)
- alagoas, recem-nascido, leito, superlotacao, não sei (4)
- verba, enfermaria, interditada, reclama, comunidade (3)
- recem-nascidos, leitos, mortalidade, alagoas, autoridades (5)

AVCoord

- numero de leitos, nao ha verba, nao sei, nao sei, nao sei (2)
- bebês, maternidade, leito, verba, mortes (5)
- maternidades, alagoas, leitos, verba, não sei (4)
- numeros de leitos, reclamação de mães, alagoas, não sei, não sei (2)
- leitos, verbas, maes, assustam, não sei (4)
- bebes, Alagoas, verba, morte, autoridades (5)
- alagoas, maternindade, mortes, autoridades, susto (5)
- maternidade, mortes, recem-nascidos, verbas, alagoas (5)

Som

- Alagoas, mortes, recém-nascidos, leitos, secretaria de saúde (5)
- leitos, mães, verba, auotridades, não sei (4)
- leitos, maternidade, recém-nascidos, Alagoas, morte (5)
- Maternidade, superlotacao, saude, maes, falta de verba (5)
- leitos, verbas, secretaria, número, não sei (4)
- superlotação, maternidades, alagoas, leitos, não sei (4)
- mortes, bebes, aumento, verba, leitos (5)
- recém-nascidos, morte, leitos, belém, verba (5)

Fragmento 7

AReal

- médico, Uberlândia, paciente, laudo, anos (5)
- médico, laudos, paciente, denuncia, polícia (5)

- laudo, graça, paciente, polícia federal, vender (5)
- policia federal, medico, laudo, paciente, denuciou (5)
- Médico, Uberlândia, preso, laudo, paciente (5)
- medico, laudo, policia federal, Uberlandia, 68 anos (5)
- laudo, venda, minas gerais, polícia, medico (5)

AVSem

- venda de laudo, 68 anos, denunciado pela pf, uberlandia, triangulo mineiro (4)
- polícia federal, denúncia, caso, não sei, não sei (3)
- laudo, medico, policia, paciente, denuncia (5)
- medico, uberlândia, fraude, laudo, prisão (5)
- policia federal, medico, laudo, paciente, denuncia (5)
- triangulo mineiro, medico, dennunciado, familia, não sei (4)
- laudo, médico, preso, paciente, cobrar (5)
- medico, 61anos,paciente, laudo, prisão (5)

AVDescoord

- policia federal, denuncia, caso, mãe, não sei (3)
- 65, laudo medico, policia federal, venda, não sei (4)
- laudo, medico, 60, gratis, preso (5)
- médico, laudo, Uberlândia, Polícia Federal, paciente (5)
- médico, polícia federal, caso, não sei, não sei (3)
- laudo, médico, uberlândia, prisão, polícia federal (5)
- anos, policial, federal, vendeu, laudo (5)
- médico, venda, laudo, polícia federal, denúncia (5)

AVCoord

- laudo, médico, preso, família, polícia federal (5)
- laudo, de graça, preso, flagrante, familia (5)
- Laudo, venda, médico, denúncia, Polícia Federal (5)

- medico, policia federal, uberlandia, cobrança irregular, denuncia (5)
- 68, polícia, federal, denunciou, caso (5)
- medico, uberlancia, venda de laudos, policia federal, prisáo (5)
- preso, paciente, policia, federal, medico (5)
- médico, laudo, de graça, vender, aposentado (4)

Som

- Uberlandia, medico, laudo, não sei, não sei (3)
- médico, laudo, venda indevida, polícia federal, uberlândia (5)
- medico, laudo, graça, vender, 68 anos (5)
- Uberlandia, medico, laudo, falso, preso (4)
- laudo, familia, denunciou, Uberlandia, não sei (4)
- falso medico, policia federal, familia, não sei, não sei (3)
- médico, uberlândia, vender, laudo, ilegal (5)
- medico, preso, laudo, vitima, policia federal (5)

Fragmento 8

AReal

- epidemia, dengue, número, maior, estado (5)
- dengue, piaui, maior, secretaria, alerta (5)
- dengue, Piauí, casos, número, maior (5)
- dengue, piauí, epidemia, secretaria, saúde (5)
- Epidemia, Piauí, saúde pública, não sei, não sei (3)
- epidemia, dengue, Piauí, 4000, numero maior (5)
- Piauí, dengue, casos, vigilância sanitária, 4000 (4)
- dengue, casos, piauí, epidemia, secretaria de saúde (5)

AVSem

- dengue, epidemia, piaui, nao sei, nao sei (3)
- piaui, epidemia, maior, secretaria, não sei (4)

- dengue, piaui, epidemia,alerta,numero (5)
- piaui, dengue, estado, casos, não sei (4)
- piaui, dengue, epidemia, secretaria de saude, casos (5)
- piaui, casos de dengue, expectativa de ser maior, não sei, não sei (3)
- epidemia, piaui, secretaria, saude, maior (5)
- dengue, piauí, epidemia, saúde, não sei (4)

AVDescoord

- dengue, piauí, número maior, não sei, não sei (3)
- casos, dengue, piaui, maior, secretaria (5)
- Dengue, epidemia, Piaui, secretaria da saúde, não sei (4)
- piaui, dengue, casos, maior, numero (5)
- Piaui, dengue, pior, não sei, não sei (3)
- piaui, casos, 4000, numero, pior (5)
- dengue, epidemia, pior, rio de janeiro, não sei (4)
- piaui, dengue, secretaria de saude, epidemia, risco (4)

AVCoord

- dengue, Piauí, epidemia, casos, não sei (4)
- dengue, casos, maior, Piauí, 4000 (5)
- Piauí, dengue, 4000, não sei, não sei (3)
- dengue, epidemia, piaui, saude, não sei (4)
- piauí, número, epidemia, casos, não sei (4)
- epidemia, dengue, piaui, não sei, não sei (3)
- Epidemia, dengue, Piauí, casos, maior (5)
- dengue, epidemia, piauí, 4 casos, bem maior (5)

Som

- Piaui, dengue, não sei, não sei, não sei (2)
- secretaria, dengue, piauí, casos, registrados (5)

- piaui, dengue, epidemia, maior, numero (5)
- dengue, epidemia, Piauí, casos, maior (5)
- Piauí, dengue, número, maior, alerta (5)
- piaui, 4 mil, dengue, não sei, não sei (3)
- dengue, piauí, secretaria, número, maior (5)
- dengue, piaui, epidemia, estado, maior (5)

Fragmento 9

AReal

- Policiais Federais, reajuste, acordo, governo, salário (5)
- policiais, reajuste, governo, parcela, paralizacao (5)
- reajuste, trabalhadores, parcelas, não sei, não sei (3)
- tres parcelas, ajuste, salarial, não sei, não sei (3)
- reajuste, servidores, parcelas, governo, aprovou (5)
- greve, polícia federal, reajuste salarial, 5 parcelas, não sei (4)
- parcelas, reajuste, greve, não sei, não sei (3)
- reajuste, paralização, policiais, governo, parcelas (5)

AVSem

- federal, reajuste, parcelas, não sei, não sei (3)
- policiais, federais, paralização, reajuste, parcelas (5)
- Policiais, federais, reajuste, parcelas, categoria (5)
- policiais, federais, reajuste, governo, parcelas (5)
- Servidores, reajuste, greve, não sei, não sei (3)
- reajuste, greve, 3 parcelas, policia ferderal, náo sei (4)
- Polícia Federal, greve, reajuste, parcela, acordo (5)
- policiais, greve, reajuste, governo, parcelas (5)

AVDescoord

- Policia Federal, aumento salarial, não sei, não sei, não sei (2)

- reajuste, parcelas, governo, categoria, não sei (4)
- reajuste, policia, parcelas, hoje, terminou (5)
- reajuste, paralisação, federal, proposta, greve (5)
- parcelas, reajuste, policiais, não sei, não sei (3)
- Reajuste, 3 parcelas, não sei, não sei, não sei (2)
- reajuste, policia federal, 32, parcelas, paralizacao (5)
- policial, federal, reajuste, acordo, governo (5)

AVCoord

- reajuste, professores, parcelas, trinta %, não sei (4)
- servidores, 30%, parcelas, reajuste, aceitaram (5)
- reajuste, 3 parcelas, governo, não sei, não sei (3)
- servidores, reajuste, parcelas, 30%, policia federal (5)
- policiais, gorveno, federal, reajuste, parcelas (5)
- policiais federais, paralisacao, reajuste salarial, proposta, parcelas (5)
- reajuste, servidores, categoria, não sei, não sei (3)
- reajuste, aceitaram o acordo, 3 parcelas, paralizacao, nao sei (4)

Som

- polícia federal, paralização, reajuste, categoria, parcelas (5)
- reajuste, 72 horas, 3 parcelas, 30%, aceitaram (5)
- paralizacao,policiais, governo, reajuste, acordo (5)
- Polícia, Governo, greve, reajuste, parcelas (5)
- parcelas, reajuste, não sei, salário, não sei (3)
- reajuste salarial, acordo, paralização, servidores federais, não sei (4)
- reajuste, acordo, aceitaram, dividir, três parcelas (5)
- reajuste, servidores, três parcelas, governo, não sei (4)

Fragmento 10

AReal

- Serra, Santa Catarina, frio, termômetros, zero (5)
- santa catarina, são joaquim, 6 graus, zero, abaixo (5)
- frio, santa catarina, termometro, não sei, não sei (3)
- Santa Catarina, frio, termõmetro, graus, zero (5)
- serra, seis, graus, frio, abaixo (5)
- sao joaquim, santa catarina, inverno, temperaturas baixas, não sei (4)
- frio, Santa Catarina, serra, 6 graus, abaixo de zero (5)
- abaixo de zero, São Joaquim, não sei, não sei, não sei (2)

AVSem

- Santa Catarina, Temperatura baixa, não sei, não sei, não sei (2)
- temperatura, são joaquim, termômetros, frio, abaixo de zero (5)
- sao joaquim, santa catarina, frio, 6 graus, zero (5)
- frio, São Joaquim, Santa Catarina, temperatura, negativo (5)
- serra, frio, Santa Catarina, zero, termômetros (5)
- santa cararina, frio, serra, 6 graus, não sei (4)
- sao joaquim, frio, 14 anos, 6 graus abaixo de zero, esta manha (5)
- frio, são joaquim, termômetros, serra, 14 anos (5)

AVDescoord

- frio, santa catarina, tempo, meteorologia, recorde (5)
- santa, catarina, termometros, baixo, zero (5)
- grau, temperatura, são joaquim, seis, abaixo de zero (5)
- 14anos, santa catarina, abaixo de zero, termometros,não sei (4)
- termometros, santa caratina, 6 graus , abaixo, termometro (5)
- frio,termometros,6 graus,14 anos, abaixo de zero (5)
- frio, abaixo de zero, santa catarina, não sei, não sei (3)
- frio, santa catarina, sao joaquim, inverno, nao sei (4)

AVCoord

- Frio, Santa Catarina, temperatura, São Joaquim, termometros (5)
- anos, sao joaquim, termometro, zero, graus (5)
- são joaquim, santa catarina, frio, termometro, abaixo de zero (5)
- temperatura, santa catarina, seis graus abaixo de zero, não sei, não sei (3)
- frio, termometro, abaixo, saojoaquim, tempo (5)
- 14 anos, frio, São Joaquim, abaixo de zero, não sei (4)
- santa catarina, zero, termometros, seis graus, 14anos (5)
- frio, são joaquim, serra, santa catarina, graus (5)

Som

- santa, catarina, frio, graus, zero (5)
- frio, serra, SC, São Joaquim, não sei (4)
- frio, São Joaquim, Santa Catarina, Serra, termômetros (5)
- termometro, frio, santa catarina, seis, graus (5)
- São Joaquim, Serra, Santa Catarina, 14 anos, 6 graus abaixo de zero (5)
- frio, Santa Catarina, São Joaquim, abaixo de zero, termometro (5)
- frio, Santa Catarina, termômetro, temperatura, abaixo de zero (5)
- frio, santa catarina,termometro, manhã, -6 graus (5)

A.2.3 Resultados das questões específicas

Nas tabelas A.2 a A.11, as respostas dadas pelos participantes dos testes às questões específicas colocadas para as notícias, conforme exposto na seção A. As respostas corretas estão marcadas por um $\surd$.

Tab. A.2: Respostas às questões para o fragmento 1

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	Maranhão √	Gualtânama √	não sei
	não sei	não sei	não sei
	não sei	Gautama √	não sei
	Maranhão √	Gautama √	Leste do estado √
	não sei	Gautama √	não sei
	Maranhão √	Gautama √	Leste √
	não sei	Gautama √	não sei
	não sei	Gautama √	não sei
AVSem	Bahia	Gautama √	Norte
	Paraná	Gautama √	norte
	não sei	não sei	sudeste
	Maranhão √	Gautama √	não sei
	Paraná	não sei	oeste
	Roraima	Gautama √	Leste √
	não sei	galtama √	nordeste
	Maranhão √	galtama √	leste √
AVDescoord	não sei	Gautama √	norte
	não sei	não sei	leste √
	não sei	não sei	não sei
	Maranhão √	não sei	leste √
	maranhão √	não sei	nada a lugar nenhum
	não sei	galtama √	não sei
	Maranhão √	não sei	maranhão
AVCoord	maranhão √	não sei	não sei
	maranhão √	não sei	não sei
	fraudes da gautama no nordeste	gautama √	leste √
	não sei	Gautama √	leste √
	não sei	de contrução de pontes	não sei
	maranhão √	galtama √	noroeste
	maranhão √	não sei	não sei
	obras inacabadas	Gautama √	não sei
Som	maranhão √	Galtama √	não sei
	belem	não sei	não sei
	Maranhão √	gautama √	norte
	Pará	não sei	norte
	Maranhão √	Construtora Gautama √	leste √
	maranhão √	Gautama √	não sei
	maranhão √	Gautama √	norte do estado
	Maranhão √	Gautama √	sul

Tab. A.3: Respostas às questões para o fragmento 2

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	4 bilhoes	senado √	iraque
	100 milhões	senado americano √	iraque e afeganistão √
	não sei	não sei	não sei
	100 Bilhões de dólares √	Bush	iraque e afeganistão √
	não sei	senado √	iraque
	não sei	senado √	afeganistao, iraque √
	100 bilhões √	senado √	iraque e afeganistão √
	mais de 100 bilhoes √	senado √	iraque, afeganistao √
AVSem	100 milhões	o senado americano √	Iraque e Afeganistão √
	100 bilhões √	congresso americano	não sei
	não sei	não sei	Iraque e Afeganistão √
	não sei	Senado √	Iraque
	100 bilhões √	não sei	no Iraque e no Afeganistão √
	alguns bilhoes	governo americano	iraque e Afeganistão √
	100 bilhões √	não sei	Iraque e Afeganistão √
	dois bilhões	congresso	Iraque e Afeganistão √
AVDescoord	100 milhões	George Bush	Afeganistão e não sei
	não sei	senado √	não sei
	10 bilhões	Câmara dos Deputados	Iraque e Afeganistão √
	não sei	senado √	Iraque
	100 bilhões √	Senado americano √	Iraque e Afeganistão √
	100 bilhóes √	senado √	Iraque e Afeganistáo √
	não sei	não sei	Afeganistào
	100 bilhoes √	a camara dos deputados	iraque e afeganistão √
AVCoord	100 bilhões √	senado dos EUA √	iraque e afeganistão √
	não sei	Bush	israel
	não sei	Congresso	Iraque e Afeganistão √
	100 bilhões √	Senado √	Iraque e Afeganistão √
	100 milhoes	Senado √	Afeganistao e Iraque √
	100 bilhões √	o congresso americano	Iraque
	100 bilhoes √	senado √	iraque e afeganistao √
Som	100 bilhões √	senado √	eua
	100	senado √	afeganistao
	não sei	senado americano √	Afeganistão
	100mil	bush	iraque e eua
	não sei	Senado dos Estados Unidos √	Iraque
	1bilhao	não sei	Iraque, Afeganistao √
	100 bilhões √	Congresso americano	Iraque e Afeganistao √

Tab. A.4: Respostas às questões para o fragmento 3

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	0	não sei	não sei
	não sei	usp √	terça-feira
	não sei	usp √	segunda-feira √
	segunda-feira	USP √	não sei
	03 de Maio √	USP √	segunda-feira √
	3 de maio √	USP √	segunda-feira √
	13 de maio	usp √	nesta quinta-feira
	1 semana	usp √	segunda-feira √
AVSem	não sei	usp √	segunda feira √
	não sei	usp √	segunda feira √
	não sei	usp √	3 de maio
	não sei	usp √	próxima segunda feira √
	3 de maio √	usp √	próxima segunda feira √
	não sei	usp √	segunda feira √
	não sei	usp √	segunda feira √
	03 de maio √	usp √	próxima segunda feira √
AVDescoord	3 de maio √	usp √	segunda-feira √
	não sei	usp √	segunda-feira √
	não sei	usp √	segunda-feira √
	não sei	usp √	segunda-feira √
	desde 3 de maio √	usp √	na próxima segunda-feira √
	não sei	usp √	segunda-feira √
	não sei	usp √	segunda-feira √
	não sei	usp √	segunda-feira √
AVCoord	03 de maio √	usp √	segunda feira √
	não sei	usp √	segunda feira √
	maio	usp √	próxima segunda feira √
	3 de maio de ano corrente √	usp √	próxima segunda feira √
	03 de maio √	usp √	segunda feira √
	3 de Maio √	usp √	segunda feira √
	não sei	usp √	segunda feira √
	não sei	usp √	3 de maio
Som	não sei	usp √	proxima segunda √
	não sei	usp √	proxima segunda-feira √
	3 de maio √	usp √	terca feira
	não sei	usp √	segunda-feira √
	não sei	usp √	próxima segunda-feira √
	13 maio	usp √	segunda-feira √
	maio	usp √	não sei
	não sei	usp √	segunda-feira √

Tab. A.5: Respostas às questões para o fragmento 4

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	não sei	não sei	Bahia √
	não sei	não sei	irará-bahia √
	não sei	falta de frequencia escolar √	Irará, Bahia √
	96 mil √	Falta da comprovação de freqüência escolar √	Irará, Bahia √
	96000 √	não sei	Irará, Bahia √
	6 mil	não sei	irará √
	96000 √	Falta de comprovação de freqüência escolar √	Irará, Bahia √
	90000	falta de comprovacao de assiduidade escolar √	irara, Bahia √
AVSem	não sei	falta de comprovação de frequencia dos filhos na escola √	irarara √
	3000	comprovante de renda	irara √
	não sei	falta de comprovação de presença escolar √	irara, BA √
	96 mil √	falta de comprovação que os filhos estão indo na escola √	Irará, Bahia √
	96mil √	desvio de bolsa	bahia √
	96 mil √	não sei	Bahia √
	96000 √	frequencia escolar √	Bahiia (Arira) √
	68000	Problemas na freqüencia escolar √	Na Bahia √
AVDescoord	não sei	Falta de comprovação escolar	Irará na Bahia √
	mais de 1000	crianças não estão na escola √	Irará, Bahia √
	não sei	Baixo rendimento escolar	Irará, BA √
	76mil	Falta de comprovacao da frequencia escolar √	Irará, Bahia √
	não sei	não sei	em Iralá, na Bahia √
	2000	falta de comprovação de freqüência escolar √	irará, bahia √
	não sei	não sei	Irará √
	96 mil √	não comprovar matrícula dos filhos	Irará, BA √
AVCoord	não sei	não sei	Irará √
	não sei	falta de frequencia √	Bahia √
	não sei	Não há comprovação de freqüência dos filhos √	Zona Rural da Bahia
	não sei	não sei	bahia √
	não sei	não sei	Irará, BA √
	96000 √	falta de frequencia escolar √	Irará Bahia √
	96 mil √	Nào comprovação da frequencia escolar √	Irará, Bahia √
	96 mil √	frequencia escolar √	Irará √
Som	96 mil √	falta de comprovacao de renda	irara, Bahia √
	96 mil √	falta de comprovação de frequência escolar √	irauá √
	96 mil √	não comprovação da frequencia dos filhos √	iralá BA √
	96 mil √	Beneficios estão bloqueados	Irará/Bahia √
	não sei	ausencia as aulas √	irara no interior da Bahia √
	não sei	não sei	irara
	96000 √	falta de comprovação da freqüência escolar √	nordeste
	não sei	falta de frequencia escolar √	não sei

Tab. A.6: Respostas às questões para o fragmento 5

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	pernambuco √	mais de 1000 √	não sei
	nordeste √	1000 assassinatos √	5 anos √
	nordeste √	não sei	não sei
	pernambuco √	1000 assassinatos √	mais de 5 anos √
	nordeste √	1000 √	5 anos √
	nordeste √	1000 √	25
	são paulo	1000 √	10 anos
	nordeste √	mais de 1000 √	mais de 5 anos √
AVSem	No nordeste √	mais de 1000 √	5 anos √
	Pernambuco √	não sei	nos últimos 5 anos √
	Não sei	5	não sei
	Pernambuco √	1000 √	não sei
	no nordeste √	não sei	5 √
	pernambuco √	não sei	20 anos
	Pernambuco √	não sei	não sei
	Pernambuco √	mil √	não sei
AVDescoord	Nordeste √	não sei	5 anos √
	nordeste √	não sei	5 anos √
	não sei	não sei	5 anos √
	pernambuco √	4	5 √
	Pernambuco √	1000 √	não sei
	Pernambuco √	1000 √	5 anos √
	Pernambuco √	não sei	5 anos √
	pernambuco √	1000 √	5 anos √
AVCoord	Pernambuco √	1000 √	não sei
	nordeste √	não sei	5 anos √
	pernambuco √	não sei	5 anos √
	Pernambuco √	não sei	5 anos √
	Nordeste √	1000 √	não sei
	Pernambuco √	não sei	5 anos √
	Pernambuco √	não sei	5 anos √
	pernambuco √	1000 √	não sei
Som	pernambuco √	não sei	não sei
	pernambuco √	não sei	não sei
	nordeste √	não sei	5 anos
	Nordeste √	mil √	não sei
	nordeste √	2	não sei
	Pernambuco √	um mil √	5 anos √
	Pernambuco √ não sei	não sei	
	Pernambuco √	1000 √	5 anos √

Tab. A.7: Respostas às questões para o fragmento 6

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	não sei	falta de leitos √	falta de verba para aumentar nro de leitos √
	não sei	falta de leitos √	falta de verba √
	não sei	falta de leitos para recém-nascidos √	falta de leitos para recém-nascidos
	não sei	falta de leito √	falta de verba √
	não sei	falta verba, baixo nro de leitos √	falta de verba √
	não sei	Falta de leitos, morte de recém nascidos √	falta de verba √
	Alagoas √	falta de leitos √	falta verba √
	Alagoas √	falta de leitos √	Falta de recursos √
AVSem	Alagoas √	Superlotação √	não sei
	alagoas √	não sei	não sei
	Alagoas √	Super lotação √	não há verbas √
	alagoas √	é necessario aumentar o numero de leitos √	não há verbas suficientes √
	Alagoas √	Falta de leitos √	Não há verbas √
	não sei	não sei	falta de verba √
	Alagoas √	Super-lotação √	falta de verba √
	alagoas √	Superlotação √	pouca verba √
AVDescoord	não sei	falta de leitos √	não sei
	alagoas √	falta de leitos √	não há verba √
	alagoas √	falta de leitos √	falta de leitos
	não sei	falta de leitos √	falta de verbas √
	Alagoas √	insuficiência de leitos √	não há verba para aumentar o número de leitos √
	Alagoas √	super-lotacao √	nao ha leitos disponiveis
	não sei	falta de leitos √	não há repasse de verbas √
	alagoas √	falta de leitos √	nao ha vagas
AVCoord	alagoas √	super-lotacao √	nao ha como aumentar o numero de leitos
	maternidade	falta de leitos √	falta de verba √
	alagoas √	falta de leitos √	falta de verbas para ampliar as maternidades √
	alagoas √	falta de leitos √	não sei
	não sei	falta de leitos √	falta de verbas √
	Alagoas √	morte de bebes	falta de verbas √
	alagoas √	não sei	não há verbas √
	alagoas √	falta de leitos √	nao ha verbas para aumentar o numero de leitos √
Som	Alagoas √	falta de leitos √	Que não há verba para o aumento de leitos √
	não sei	falta de leitos √	não há verba para aumentar o número de leitos √
	Alagoas √	Alta taxa de mortalidade de recém-nascidos	Não pode aumentar o nro de leitos por causa da alta taxa de natalidade
	Alagoas √	Superlotação √	Falta de verbas √
	em algum hospital	há poucos leitos √	poucos leitos
	alagos √	superlotação e falta de leitos √	não há verbas √
	não sei	falta de leitos √	Não há verba √
	Belém	poucos leitos √	falta verba √

Tab. A.8: Respostas às questões para o fragmento 7

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	68 √	Uberlândia √	não sei
	68 anos √	minas √	vender laudos que deveriam ser gratuitos √
	não sei	não sei	Venda de laudo que deveria ser fornecido de graça √
	68 anos √	não sei	tentativa de venda de um laudo √
	68 √	Uberlândia √	Venda de laudo √
	68 anos √	Uberlândia √	Venda de laudo √
	não sei	não sei	não sei
AVSem	68 √	uberlândia √	venda de laudo ilegal √
	não sei	não sei	não sei
	não sei	uberlândia √	venda de laudo medico √
	68 anos √	uberlândia/MG √	cobrança de um laudo que não deveria ser cobrado √
	64	não sei	o medico estava cobrando um laudo que era para ser gratuito √
	68 √	triangulo mineiro √	não sei
	não sei	não sei	cobrar um laudo que deveria ser gratuito √
	61 anos	uberlândia √	venda de laudo medico √
AVDescoord	não sei	não sei	não sei
	65 anos	paraná	venda de laudo medico √
	60	não sei	cobrança indevida √
	não sei	Uberlândia √	venda de laudo √
	não sei	não sei	não sei
	não sei	Uberlândia √	Venda de laudos médicos √
	66 anos	não sei	Venda de laudo √
	68 √	ribeirão preto	Venda de laudo médico √
AVCoord	não sei	Uberlândia √	Vende de laudo √
	65 anos	interior de são paulo	venda de laudo que deveria ser fornecido de graça √
	65	Belo Horizonte	Venda de laudo gratuito por lei √
	não sei	Uberlandia √	Cobranca irregular √
	68 anos √	não sei	não sei
	68	uberlandia √	venda de laudos √
	63	Uberlandia √	não sei
	68 anos √	não sei	venda de laudo gratuito √
Som	68 √	uberlândia √	venda de laudos √
	68 anos √	uberlândia √	venda de laudos √
	68 anos √	não sei	venda de laudo médico √
	Não sei	Uberlandia √	Falsificação de laudos
	62 anos	Uberlandia √	venda de laudo √
	65	não sei	não sei
	68 anos √	Uberlândia √	Venda de laudo que deveria ser de graça √
	68 √	não sei	venda de laudo que era gratuito √

Tab. A.9: Respostas às questões para o fragmento 8

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	não sei	Piauí √	não sei
	não sei	Piauí	secretaria
	4000 √	Piauí √	Secretaria de saúde √
	4000 √	Piauí √	Secretaria de saúde √
	4000 √	Piauí √	Secretaria de saúde √
	4000 √	Piauí √	Secretaria de saúde √
	4000 √	Piauí √	vigilância sanitária √
	4000 √	Piauí √	Secretaria de saúde √
AVSem	1000	Piauí √	secretaria da saúde √
	não sei	Piauí √	a secretaria √
	não sei	Piauí √	secretaria √
	não sei	Piauí √	governo
	4	Piauí √	secretaria da saúde √
	não sei	Piauí √	secretaria de saúde de estado √
	não sei	Piauí √	secretaria da saúde √
	não sei	Piauí √	autoridades
AVDescoord	6 mil casos	Piauí √	não sei
	4000 √	Piauí √	a secretaria da saude √
	mais de 4000 √	Piauí √	Secretaria da saude √
	4mil √	Piauí √	prefeitura
	não sei	Piauí √	Secretaria da saude √
	4000 √	Piauí √	ministério da saude
	não sei	Rio de Janeiro	secretaria da saude √
	não sei	Piauí √	secretaria da saude √
AVCoord	mais de 4000 √	Piauí √	A secretaria de saúde √
	mais de 4000 √	Piauí √	Governo do Estado
	Mais de 4000 √	não sei	Secretaria de Saúde √
	3000	Piauí √	ministerio da saúde
	não sei	Piauí √	não sei
	não sei	Piauí √	especialistas
	não sei	Piauí √	não sei
	4	Piauí √	não sei
Som	não sei	Piauí √	não sei
	4000 √	Piauí √	secretaria de saúde √
	não sei	Piauí √	não sei
	3000	Piauí √	não sei
	4000 √	Piauí √	vigilância sanitária
	4 mil √	Piauí √	não sei
	4000 √	Piauí √	Secretaria de saúde √
	não sei	Piauí √	estado

Tab. A.10: Respostas às questões para o fragmento 9

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	30 dias	Policiais federais √	30% √
	não sei	policiais √	não sei
	não sei	trabalhadores	não sei
	3 semanas	sindical	não sei
	3 meses	servidores √	30% √
	não sei	Policial federal √	não sei
	não sei	Polícia civil	30% √
	72 horas √	greve	não sei
AVSem	não sei	salário	30% √
	72 Hs √	Policiais Federais √	30% √
	não sei	Policiais Federais √	não sei
	72 horas √	social política	30% √
	72 horas √	Servidores públicos √	30% √
	4 dias	Policia Federal √	30% √
	não sei	Policia Federal √	30% √
	72 horas √	Policia Federal √	30% √
AVDescoord	não sei	policia federal √	não sei
	30 dias	não sei	30% √
	não sei	paralisacao da policia √	não sei
	não sei	servidores federais √	não sei
	72 horas √	não sei	30% √
	não sei	policia √	não sei
	32	policia federal √	32
	não sei	policiais federais √	30% √
AVCoord	não sei	professores	30% √
	não sei	servidores publicos √	30% √
	não sei	não sei	não sei
	não sei	policia federal √	30% √
	não sei	policia federal √	30% √
	3 semanas	policiais federais √	não sei
	não sei	servidores √	30% √
	nao sei	não sei	30% √
Som	não sei	Polícia Federal √	30% √
	72 horas √	agentes federais	30% √
	72 horas √	Policia √	30% √
	Não sei	Policiais Federais √	não sei
	não sei	não sei	30% √
	não sei	não sei	não sei
	não sei	não sei	não sei
	não sei	servidores √	não sei

Tab. A.11: Respostas às questões para o fragmento 10

Apresentação	*Questão 1*	*Questão 2*	*Questão 3*
AReal	-5 graus celsius	- 5 graus celsius	não sei
	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	agora	-6	14 anos √
	Hoje de manhã	5 graus abaixo de zero	não sei
	não sei	seis graus abaixo de zero	há bastante tempo
	inverno	6 graus abaixo de zero	alguns anos
	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	não sei	abaixo de zero	não sei
AVSem	não sei	10 graus abaixo de zero	não sei
	não sei	não sei	não sei
	6 graus abaixo de zero	6 graus abaixo de zero	6 anos
	manhã	-6	não sei
	não sei	6 graus abaixo de zero	14 anos √
	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	esta manha	menos 6	14 √
	hoje	6 graus abaixo de zero	14 anos √
AVDescoord	6 graus abaixo de 0	6 graus abaixo de 0	não sei
	madrugada	abaixo de zero	não sei
	nesta semana	6 graus abaixo de zero	não sei
	não sei	abaixo de zero	14 anos √
	-6	-6	não sei
	6 graus abaixo de zero	6 graus abaixo de zero	14 anos √
	esta manhã	abaixo de zero	não sei
	hoje	6 abaixo de zero	14 anos √
AVCoord	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	hj	4 abaixo de zero	14 anos √
	esta manhã	6 graus abaixo de zero	há muito tempo
	6 graus abaixo de zero	6 graus abaixo de zero	há muito tempo
	-6	-6	3 anos
	abaixo de zero	não sei	14 anos √
	14 de maio	-6 graus	14 anos √
	6 graus negativos	6 graus negativos	não sei
Som	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	14 graus abaixo de zero	-14	50 anos
	manhã	10 graus abaixo de zero	15 anos
	6 graus abaixo de zero	6 graus abaixo de zero	não sei
	não sei	6 graus abaixo de zero	14 anos √
	-6	-6	10 anos
	-6	-6	14 anos √
	-6 graus	-6 graus	muitos anos