Ambiente para interação baseada em reconhecimento de emoções por análise de expressões faciais

Caetano Mazzoni Ranieri

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Caetano Mazzoni Ranieri

Ambiente para interação baseada em reconhecimento de emoções por análise de expressões faciais

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Roseli Aparecida Francelin Romero

USP – São Carlos
Setembro de 2016

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

R197a

Ranieri, Caetano Mazzoni
Ambiente para interação baseada em reconhecimento de emoções por análise de expressões faciais / Caetano Mazzoni Ranieri; orientadora Roseli Aparecida Francelin Romero. - São Carlos - SP, 2016.
105 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2016.

1. Emoções. 2. Dispositivos móveis. 3. Interação humano-robô. 4. Robótica social. I. Romero, Roseli Aparecida Francelin, orient. II. Título.

Caetano Mazzoni Ranieri

Environment for interaction based on emotion recognition by analysis of facial expressions

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Roseli Aparecida Francelin Romero

USP – São Carlos
September 2016

AGRADECIMENTOS

Agradeço ao Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), pelo ambiente acolhedor e pela excelente infra-estrutura.

Agradecimentos especiais são direcionados à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento deste trabalho, sob processo 2014/16862-4, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelos seis meses de bolsa institucional.

Agradecimentos muito especiais são direcionados a minha orientadora, a Profa. Dra. Roseli Aparecida Francelin Romero, por ter acreditado no meu potencial e dedicado sólido suporte ao longo do desenvolvimento deste trabalho. Em medida semelhante, agradeço ao Prof. Dr. Humberto Ferasoli Filho, que, tendo me acompanhado desde os primeiros momentos da minha graduação, mostrou-me o caminho para atingir meus objetivos e teceu orientações fundamentais para que eu pudesse evoluir da melhor maneira possível. Além disso, agradeço-lhe por ter cedido o espaço para realização de experimentos na UNESP Bauru.

No mesmo contexto, agradeço ao Dr. Silas Franco dos Reis Alves, pela colaboração no decorrer dos anos, e a meus colegas do Laboratório de Aprendizado de Robôs (LAR-ICMC-USP), pela disposição em discutir ideias e compartilhar conhecimento. Também agradeço ao Prof. Dr. Giampaolo Luiz Libralon, por ter gentilmente cedido uma versão do seu sistema para reconhecimento de emoções por análise de expressões faciais, adotada neste trabalho.

Gostaria de agradecer, sobretudo, a minha família. Em especial a meus pais, Servio Tulio Vieira Ranieri e Vera Lucia Andrade Mazzoni, pelo apoio irrestrito durante toda a minha vida pessoal e acadêmica. Dedico agradecimentos muito caros a meus irmãos, Erick Andrade Busato e Jonathan Mazzoni Busato, por terem estado comigo desde sempre e me ajudado a crescer em diferentes aspectos. Agradeço, ainda, a minha namorada, Tathiana Mello, por permanecer a meu lado apesar de todas as adversidades, e a todos os amigos que se fizeram presentes nos diferentes momentos da minha história pessoal.

“Nenhuma decisão sensata pode ser tomada
sem que se leve em conta o mundo não apenas como ele é,
mas como ele virá a ser.”
(Isaac Asimov)

RESUMO

RANIERI, C. M.. **Ambiente para interação baseada em reconhecimento de emoções por análise de expressões faciais**. 2016. 105 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Nas ciências de computação, o estudo de emoções tem sido impulsionado pela construção de ambientes interativos, especialmente no contexto dos dispositivos móveis. Pesquisas envolvendo interação humano-robô têm explorado emoções para propiciar experiências naturais de interação com robôs sociais. Um dos aspectos a serem investigados é o das abordagens práticas que exploram mudanças na personalidade de um sistema artificial propiciadas por alterações em um estado emocional inferido do usuário. Neste trabalho, é proposto um ambiente para interação humano-robô baseado em emoções, reconhecidas por meio de análise de expressões faciais, para plataforma Android. Esse sistema consistiu em um agente virtual agregado a um aplicativo, o qual usou informação proveniente de um reconhecedor de emoções para adaptar sua estratégia de interação, alternando entre dois paradigmas discretos pré-definidos. Nos experimentos realizados, verificou-se que a abordagem proposta tende a produzir mais empatia do que uma condição controle, entretanto esse resultado foi observado somente em interações suficientemente longas.

Palavras-chave: Emoções, Dispositivos móveis, Interação humano-robô, Robótica social.

ABSTRACT

RANIERI, C. M.. **Ambiente para interação baseada em reconhecimento de emoções por análise de expressões faciais**. 2016. 105 f. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

In computer sciences, the development of interactive environments have motivated the study of emotions, especially on the context of mobile devices. Research in human-robot interaction have explored emotions to create natural experiences on interaction with social robots. A fertile aspect consist on practical approaches concerning changes on the personality of an artificial system caused by modifications on the user's inferred emotional state. The present project proposes to develop, for Android platform, an environment for human-robot interaction based on emotions. A dedicated module will be responsible for recognizing emotions by analyzing facial expressions. This system consisted of a virtual agent aggregated to an application, which used information of the emotion recognizer to adapt its interaction strategy, alternating between two pre-defined discrete paradigms. In the experiments performed, it was found that the proposed approach tends to produce more empathy than a control condition, however this result was observed only in sufficiently long interactions.

Key-words: Emotions, Mobile devices, Human-robot interaction, Social robotics.

LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura neurofisiológica da emoção. Adaptado de Brave e Nass (2002). . . 24
Figura 2 – Expressões faciais associadas às seis emoções em que consiste o modelo de Ekman. Na sequência: raiva, medo, aversão, surpresa, alegria e tristeza (EKMAN, 1999). 26
Figura 3 – Modelo para representação de emoções proposto por Russell, Lewicka e Niit (1989). Nesta figura, algumas emoções discretas são reconhecidas e mapeadas para o espaço apresentado. No entanto, as emoções passíveis de representação por esse modelo podem não se restringir a esse conjunto. . . . 26
Figura 4 – Etapas da classificação de emoções baseada em análise de expressões faciais. O sistema de reconhecimento pode ser dividido em três etapas: detecção de face, extração de características e classificação. 28
Figura 5 – Um robô dotado de interface social é um robô social. Adaptado de Hegel *et al.* (2009). 29
Figura 6 – Representação do *uncanny valley*. Extraído e traduzido de Mori, MacDorman e Kageki (2012). 30
Figura 7 – Regiões críticas para o teste-t pareado (a) unicaudal à direita; (b) unicaudal à esquerda; e (c) bicaudal. 34
Figura 8 – Robô *PeopleBot*. Extraído de Gácsi *et al.* (2016). 39
Figura 9 – Robô *Ugobe's Pleo*. Extraído de Putten *et al.* (2013). 40
Figura 10 – Materiais do experimento descrito em Darling, Nandy e Breazeal (2015). São eles: (a) inseto robótico; (b) espaço de confinamento para movimento; (c) mesa; (d) marreta. Extraído de Darling, Nandy e Breazeal (2015) 40
Figura 11 – Robô Kismet. Extraído de Breazeal (2003). 42
Figura 12 – Robô iCat jogando xadrez com criança. Extraído de Leite *et al.* (2012). . . . 43
Figura 13 – Agente virtual Ana (esquerda) e robô humanoide Kobiana (direita). 44
Figura 14 – Assistente virtual com personagem sintético para interação. Extraído de Yoon e Cho (2012). 45
Figura 15 – Ambiente com uso da plataforma SMILE. Extraído de Russell *et al.* (2014). 45
Figura 16 – Robô Roburguer, com *smartphone* embarcado. 46
Figura 17 – Etapas desempenhadas pelo reconhecedor de emoções adotado. 47
Figura 18 – Extração de características, com uso do *FaceTracker*. (a) imagem original; (b) conjunto de 33 pontos extraídos e relações de interesse para o reconhecedor de emoções. 48

Figura 19 – Personagem virtual: (a) paradigma *amigável*; (b) paradigma *normal*. 52
Figura 20 – Arquitetura do sistema proposto, composto por um classificador de emoções, um motor de interação e um motor de conteúdo. 52
Figura 21 – Transições entre os modos segundo os quais o ambiente pode operar. 53
Figura 22 – Ciclo de vida de uma *activity*. 55
Figura 23 – Arquitetura de software do ambiente, contemplando tanto o smartphone quanto a base computacional. A conexão está representada como uma seta bidirecional entre as interfaces de ambos os dispositivos, ignorando os dispositivos e protocolos que se possam interpor. 56
Figura 24 – Sequência de envios e recebimentos para que o aplicativo para smartphone (cliente) possa usar a base computacional (servidor) para obter a classificação da expressão facial obtida pela câmera frontal. 58
Figura 25 – *Activity* responsável pela interação entre humano e agente no aplicativo para smartphone. Para favorecer a visualização, o tamanho da legenda está superdimensionado nesta imagem. 60
Figura 26 – Diagrama de classes simplificado do sistema. Algumas classes, com relevância restrita, foram omitidas, para facilitar a compreensão da estrutura global do desenvolvimento do aplicativo. 61
Figura 27 – Exemplo do jogo *Doublets*, com as palavras *fogo* e *gelo*. 63
Figura 28 – Sequência de envios e recebimentos para que o aplicativo para *smartphone* (cliente) possa usar a base computacional (servidor) para obter notícias de um *feed* específico. 64
Figura 29 – Fluxograma mostrando a sequência de procedimentos executada para cada exibição de notícia do ambiente no modo *noticiário*. 65
Figura 30 – Histogramas representando as frequências das notas em cada amostra do experimento A. 70
Figura 31 – Histogramas representando as frequências das notas em cada amostra do experimento B. 74

LISTA DE TABELAS

Tabela 1 – Valores críticos para aplicação do teste de aderência à normalidade Kolmogorov-Smirnov. Adaptado de O'Connor e Kleyner (2011). 33
Tabela 2 – Argumentos para execução do reconhecedor de emoções. 58
Tabela 3 – Valores D_n obtidos no teste de normalidade Kolmogorov-Smirnov aplicado ao experimento A. 70
Tabela 4 – Resultados obtidos com aplicação do teste-t pareado unicaudal sobre as amostras do experimento A. As colunas *média das diferenças* e *desvio-padrão* referem-se às médias das diferenças entre as avaliações das duas condições ($\bar{x}_{\text{Claire}} - \bar{x}_{\text{Rachel}}$), tomadas par a par. 70
Tabela 5 – Questões e respostas disponibilizadas. 71
Tabela 6 – Respostas à questão Q1, por participante, para a condição Claire. 72
Tabela 7 – Respostas à questão Q1, por participante, para a condição Rachel. 73
Tabela 8 – Respostas para a questão Q2, por participante. Embora a quantidade de respostas positivas tenha sido maior para Claire do que para Rachel, a maioria dos participantes não viu diferença entre as duas abordagens. 73
Tabela 9 – Valores D_n obtidos no teste de normalidade Kolmogorov-Smirnov aplicado ao experimento B. 74
Tabela 10 – Resultados obtidos com aplicação do teste-t pareado unicaudal sobre as amostras do experimento B. As colunas *média das diferenças* e *desvio-padrão* referem-se às médias das diferenças entre as avaliações das duas condições ($\bar{x}_{\text{Claire}} - \bar{x}_{\text{Rachel}}$), tomadas par a par. 74

SUMÁRIO

1 INTRODUÇÃO 19
1.1 Objetivos 21
1.1.1 Objetivo Geral *21*
1.1.2 Objetivos Específicos *21*

2 FUNDAMENTAÇÃO TEÓRICA 23
2.1 O Estudo das Emoções 23
2.1.1 Conceitualização *23*
2.1.2 Neurofisiologia da Emoção *24*
2.1.3 Modelos para Representação de Emoções *25*
2.1.4 Reconhecimento Automático de Emoções *26*
2.1.4.1 Análise de Expressões Faciais *27*
2.2 Princípios de Robótica Social 28
2.2.1 Estatística para Experimentos com Humanos *31*
2.2.1.1 Teste Kolmogorov-Smirnov *32*
2.2.1.2 Teste-t Pareado *33*
2.3 Considerações Finais 35

3 TRABALHOS RELACIONADOS 37
3.1 Robótica Social e Emoções 37
3.1.1 Estudos Empíricos em Interação Humano-Robô *38*
3.1.2 Desdobramentos Tecnológicos *41*
3.2 Reconhecedor de Emoções Adotado neste Trabalho 46
3.3 Considerações Finais 49

4 DESENVOLVIMENTO DO PROJETO 51
4.1 Visão Global do Ambiente 51
4.2 Sistema Operacional Android 54
4.3 Arquitetura de Software 55
4.4 Adaptação do Reconhecedor de Emoções 57
4.5 Projeto e Desenvolvimento do Motor de Interação 59
4.5.1 Detalhes de Implementação *60*
4.6 Projeto e Desenvolvimento do Motor de Conteúdo 62
4.6.1 Estado Iatente *62*

4.6.2 *Jogo* *62*
4.6.3 *Noticiário* *64*
4.7 Considerações Finais 66

5 EXPERIMENTOS 67
5.1 Experimento A 68
5.1.1 *Resultados* *69*
5.2 Experimento B 71
5.2.1 *Resultados* *72*
5.3 Outros Experimentos 74
5.4 Discussão 75
5.5 Considerações finais 78

6 CONCLUSÃO 79

REFERÊNCIAS 83

APÊNDICE A TESTES DE NORMALIDADE - EXPERIMENTO A . 89

APÊNDICE B TESTES DE NORMALIDADE - EXPERIMENTO B . 95

APÊNDICE C QUESTIONÁRIO - EXPERIMENTO A 99

APÊNDICE D QUESTIONÁRIO - EXPERIMENTO B 103

CAPÍTULO

1

INTRODUÇÃO

Um aspecto importante para o sucesso da espécie humana tem sido a sua eficácia em comunicar-se. Embora a componente verbal dessa comunicação seja importante, parte da informação trocada entre seres humanos ocorre de outras maneiras. Gestos corporais podem ser uma forma sintética de apresentar informação sem necessidade de palavras, ou acrescentar significado a uma comunicação verbal. Expressões faciais podem revelar significados difíceis de serem ditos de outra forma. Diferentes entonações vocais podem dar sentidos distintos para um mesmo discurso. Parte da informação, portanto, é expressa na forma de emoções. Esse papel das emoções na comunicação entre indivíduos foi apresentado por Charles Darwin, em seu clássico *The Expression of the Emotions in Man and Animals* (DARWIN; EKMAN; PRODGER, 1998), que influenciou substancialmente a pesquisa acadêmica contemporânea em emoções (HESS; THIBAULT, 2009).

Ainda que emoções estejam presentes nas diferentes formas de comunicação entre seres humanos, e sua identificação frequentemente ajude na tomada de decisões diante de outras pessoas, a interação entre o ser humano e sistemas computacionais ainda é, essencialmente, não emocional. Contudo, não são novas as ideias de incorporação de emoções a ambientes dessa natureza. Por um lado, segundo Damasio (2005), emoções artificiais podem possibilitar que sistemas computacionais sejam mais eficazes ao simular características humanas para resolver problemas e tomar decisões, visto que emoções podem influenciar substancialmente os processos cognitivos humanos. Por outro lado, conforme sugeriu Picard (2000), inaugurando a área de computação afetiva, as habilidades de reconhecer, compreender e possuir emoções são requisitos para que computadores possam interagir de forma natural com seres humanos.

Nesse sentido, o estudo das emoções tem sido intensificado na Ciência da Computação. O reconhecimento de emoções em seres humanos tornou-se um problema exaustivamente abordado em trabalhos acadêmicos, aplicando-se diferentes técnicas para inferir emoções a partir de pistas como expressões faciais ou fala (ZENG *et al.*, 2009). Nos últimos anos, com o crescente

interesse em aprendizado profundo (comumente referido pelo termo em inglês, *deep learning*), têm ganhado destaque trabalhos buscando fazer reconhecimento de emoções multimodal, isto é, combinando duas ou mais pistas de naturezas distintas. Tais sistemas devem considerar, em conjunto, resultados de análises de expressões faciais, entonação vocal ou outros tipos de informação. Em anos recentes, sucedeu-se a criação de um desafio anual sobre o tema, o *Emotion Recognition Challenge in the Wild Challenge* (EmotiW) (DHALL *et al.*, 2015). Esse desafio consiste na construção de sistemas que infiram a emoção de indivíduos em vídeos nos quais eles proferem uma sentença enquanto manifestam uma dada emoção.

Na robótica social, reconhecer o estado afetivo de um humano pode ser de grande importância, sobretudo em aplicações que produzem relacionamentos mais íntimos entre o indivíduo e o agente robótico ou virtual, como acompanhantes para idosos ou tutores para crianças. Ambientes dessa natureza podem fazer uso de arquiteturas baseadas em emoções para facilitar e tornar crível a interação entre humanos e robôs de propósito social, gerando a ideia de um parceiro (BREAZEAL; BROOKS, 2005). Além disso, sintetizar emoções em robôs capazes de produzir estímulos emocionais para humanos pode proporcionar uma rica plataforma para a pesquisa em emoções, conforme discutido por Cañamero (2014).

Vale observar também o caso dos dispositivos móveis, especialmente os *smartphones*, que acompanham seus usuários em situações diversas do seu cotidiano. Conforme observado por Vincent (2013), o relacionamento com esses dispositivos permite ao usuário explorar diferentes emoções, de acordo com os conteúdos selecionados e a informação produzida. Tem se tornado comum o comportamento de recorrer ao *smartphone* em busca de amparo emocional, ou mesmo para compartilhar momentos regozijantes. Essa relação íntima e, em alguma medida, emocional, ocorre por se tratar de um dispositivo computacional continuamente presente e de fácil acesso. Trata-se de uma interpretação segundo a qual as emoções do usuário de um *smartphone* são levadas em conta mesmo na ausência de análise emocional pré-programada.

Alguns estudos têm investigado reações emocionais dos seres humanos diante de robôs, em diferentes situações (DARLING; NANDY; BREAZEAL, 2015). Outros têm avaliado como a seleção de comportamentos baseada na detecção de estados emocionais pode influenciar a percepção das pessoas diante de agentes robóticos (LEITE *et al.*, 2012). Estendendo estes trabalhos, a proposta desta dissertação de mestrado é projetar e desenvolver um agente personificado e dotá-lo de estratégias baseadas em reconhecimento de emoções para interação humano-robô, aproveitando as vantagens e conveniências oferecidas pelos dispositivos móveis.

O agente se concretizou na forma de uma personagem virtual, proposta nesta dissertação e agregada a um aplicativo para *smartphones* Android. A personagem foi projetada de modo a emitir um mesmo conjunto de comportamentos através de pistas verbais e não verbais distintas, adaptando seu padrão comportamental de acordo com a emoção do usuário, inferida por um classificador automático e provendo resposta afetiva pretensamente verossímil. Não se trata de emitir emoções para condensar informação, tampouco de extrair significado semântico das

emoções reconhecidas, mas sim de responder, na forma de variações de padrões comportamentais, às comunicações emocionais exibidas pelos usuários. Com isso, espera-se produzir uma sensação de identificação afetiva, ou *empatia*, contribuindo para estabelecer interações humano-robô com engajamento social progressivamente mais intenso (STEPHAN, 2015).

O reconhecedor de emoções adotado é parte da tese de doutorado apresentada em Libralon (2014) e desenvolvida no Laboratório de Aprendizado de Robôs (LAR) do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC/USP). O sistema em questão classifica imagens frontais de faces humanas, reconhecendo as seis emoções básicas definidas pelo modelo discreto de Ekman (1999) (alegria, surpresa, tristeza, medo, raiva, aversão) e a emoção neutra.

Durante toda a interação, o aplicativo obtém imagens da câmera frontal do *smartphone* e as aplica como entradas para o classificador supracitado. Como resposta às emoções reconhecidas, a personagem pode manifestar um dentre dois paradigmas de interação: *normal* e *amigável*. O paradigma *normal*, direcionado a estados emocionais negativos, exibe poucos traços de personalidade e é emocionalmente neutro. O paradigma *amigável*, por sua vez, é direcionado a estados emocionais positivos, é extrovertido e emocionalmente alegre.

Esta dissertação está organizada do seguinte modo. No Capítulo 2, será desenvolvida uma fundamentação teórica, abordando os principais conceitos relevantes à compreensão da pesquisa desenvolvida. No Capítulo 3, trabalhos relacionados serão apresentados, evidenciando o estado da arte. No Capítulo 4, será descrito como o sistema se comporta e como foram desempenhadas as diferentes etapas de seu desenvolvimento. Experimentos, resultados e discussão serão apresentados no Capítulo 5 e, finalmente, as conclusões e trabalhos futuros serão tecidos no Capítulo 6.

1.1 Objetivos

1.1.1 Objetivo Geral

Este trabalho de Mestrado tem como objetivo propor, implementar e validar um ambiente de interação humano-agente influenciado por emoções, desenvolvendo uma personagem virtual para *smartphones* e dotando-a de uma estratégia de interação adaptável às emoções inferidas por um reconhecedor automático baseado em análise de expressões faciais.

1.1.2 Objetivos Específicos

- Criar uma arquitetura de *software* que propicie um ambiente de interação baseado nas emoções percebidas do usuário, usando expressões faciais obtidas de imagens;

- Implementar tal arquitetura em um aplicativo para sistema operacional *Android*, fazendo uso de ferramentas contidas em dispositivos móveis, como câmera frontal e microfone;
- Fornecer arcabouço para o desenvolvimento de outros sistemas interativos baseados em emoções, os quais possam inspirar-se nas estratégias ou na arquitetura proposta;
- Contribuir para a construção de ambientes de robótica social cujo comportamento seja modificado pelo reconhecimento de emoções humanas.

CAPÍTULO

2

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, será apresentado o referencial teórico aplicado no desenvolvimento desta dissertação. Serão discutidos fundamentos sobre diversos aspectos das emoções humanas e da robótica social. Quanto às emoções, uma abordagem multidisciplinar é apropriada, levando em conta mecanismos responsáveis por produzi-las, modelos para sua representação e técnicas para seu reconhecimento automático em humanos. Quanto aos robôs sociais, uma discussão introdutória faz-se relevante, em especial sobre algumas definições. Além disso, é importante apresentar a estatística aplicada na condução dos experimentos nessa área.

2.1 O Estudo das Emoções

Nesta seção, será apresentada uma discussão introdutória sobre emoções, elucidando os pontos mais relevantes para este trabalho. As emoções são definidas e situadas enquanto objeto de estudo, e serão contemplados aspectos biológicos das emoções, em especial sua manifestação no cérebro humano. Modelos para representação de emoções serão apresentados, além de métodos para seu reconhecimento automático.

2.1.1 Conceitualização

Não se deve confundir *emoção* com *humor*. Uma emoção é uma reação, tipicamente breve e intensa, relacionada a um evento determinado. Diferentemente, humores são estados afetivos globais e difusos, de duração indeterminada (PERGHER *et al.*, 2006). Por exemplo, uma pessoa pode ficar triste diante de alguma situação (emoção), ou ficar genericamente deprimida (humor). De uma perspectiva funcional, a emoção pode surgir de uma situação específica, a fim de preparar o organismo para reagir adequadamente. O humor, por outro lado, tende a enviesar estratégias cognitivas de longo prazo, servindo como uma lente através da qual os eventos são observados. Humores tendem a determinar quais emoções estão mais propensas a serem

experimentadas, atuando como determinantes de limiares de ativação diferentes para cada tipo de emoção. Da mesma maneira, emoções também podem influenciar humores.

Examinando as emoções de um ponto de vista evolutivo, Charles Darwin estabeleceu três princípios para explicar por que as emoções são expressas da forma como são (DARWIN; EKMAN; PRODGER, 1998). O primeiro é o princípio das ações aproveitáveis, segundo o qual determinadas respostas a estímulos representaram vantagem evolutiva em algum ponto da filogenia de uma dada espécie. O segundo princípio, importante para as habilidades de comunicação, é o princípio da antítese: quando um indivíduo manifesta um estado mental diametralmente oposto a um estado derivado do princípio das ações aproveitáveis, a expressão resultante também tende a ser oposta, como ao encolher os ombros em sinal de impotência ou arrependimento. Finalmente, o terceiro princípio é o da ação direta de excitação do sistema nervoso no corpo, segundo o qual algumas expressões, como o riso e o tremor, decorrem de descarga de excitação do sistema nervoso. Embora esses três princípios tenham sido pouco examinados em pesquisas contemporâneas em emoções, segundo Hess e Thibault (2009), as ideias que levaram a sua elaboração foram e são bastante influentes no desenvolvimento da área.

Com o desenvolvimento da neurociência e da psicologia cognitiva, os processos através dos quais as emoções são produzidas tornaram-se mais conhecidos. Na próxima subseção, esses aspectos serão apresentados.

2.1.2 Neurofisiologia da Emoção

Na Figura 1, adaptada de Brave e Nass (2002), é mostrada a estrutura neurofisiológica da emoção. *Informação sensorial* refere-se aos eventos responsáveis por gerar uma emoção, e as respostas fisiológicas são a forma como as emoções manifestam-se no indivíduo. Nas elipses da figura, são representadas as três regiões do cérebro humano envolvidas no processo de produzir emoções. São elas o *tálamo*, o *sistema límbico* e o *córtex*. As emoções se manifestam na forma de *respostas fisiológicas*, traduzindo-se em reações autonômicas, como alterações na frequência cardíaca e na pressão sanguínea.

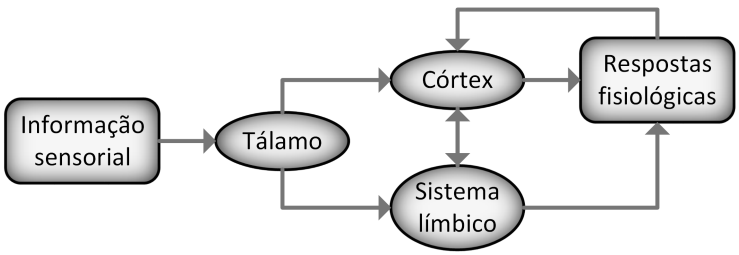


Figura 1 – Estrutura neurofisiológica da emoção. Adaptado de Brave e Nass (2002).

O tálamo recebe toda a informação sensorial, funcionando como processador de sinais básico e comunicando-se, simultaneamente, com o córtex e com o sistema límbico. O sistema límbico avalia as suas entradas constantemente e, no caso de se determinar relevância, envia

sinais para o corpo, na forma de respostas fisiológicas, e para o córtex, para o caso de haver necessidade de processamento cognitivo.

O caminho direto entre tálamo e sistema límbico (mecanismo tálamo-límbico) é responsável por emoções ditas primárias, referentes a reações primitivas, como aversões e atrações inatas. Nos casos em que o córtex aciona o sistema límbico após processamento cognitivo (mecanismo córtico-límbico), têm-se emoções secundárias, como frustração e orgulho. Há, ainda, emoções mistas, produtos desses dois mecanismos atuando em conjunto, como eventos que, após causarem uma reação inicial de medo, são reconhecidos como inofensivos após avaliação racional.

Em meio a uma extensa discussão em que se procura determinar se as emoções são inatas ou aprendidas, alguns pesquisadores defendem que algumas das emoções humanas podem ser categorizadas em conjuntos básicos independentes de traços culturais (EKMAN; FRIESEN, 2003), independentemente de quais sejam essas emoções, enquanto as outras advêm de combinações entre múltiplas emoções básicas ou são aprendidas no convívio social (BOWER, 1992). Dessa maneira, o sistema límbico poderia ser pré-programado para reconhecer as emoções básicas, enquanto as demais seriam resultantes do mecanismo tálamo-cortical.

2.1.3 Modelos para Representação de Emoções

Para representar emoções computacionalmente, seja em simulação de processos cognitivos, seja em reconhecimento de suas expressões em seres humanos, é importante ter um modelo que possa representa-las. Esses modelos podem ser divididos em *discretos* ou *contínuos* (GIL; BERLANGA; MOLINA., 2012).

Modelos discretos, especialmente aplicáveis para denotar emoções primárias, desenvolvem-se sobre a premissa de que emoções podem ser categorizadas dentro de um conjunto finito de categorias. No modelo de Ekman (1999), amplamente utilizado, são definidas seis emoções básicas: alegria, tristeza, raiva, surpresa, medo e aversão. Na Figura 2 são mostradas expressões faciais associadas a cada uma dessas emoções, conforme a pesquisa de Ekman. Os trabalhos conduzidos pelo autor concluíram que se trata de expressões universais, isto é, inatas e independentes de aspectos culturais.

Por outro lado, modelos contínuos não fixam um conjunto finito de emoções. Em vez disso, descrevem-nas por meio de variáveis contínuas em espaços, em geral, multidimensionais. Como esses modelos não nomeiam categoricamente as emoções, por meio deles podem ser inferidas não somente emoções primárias, mas também emoções secundárias. Um modelo contínuo bastante relevante é o de Russell, Lewicka e Niit (1989), que considera as emoções dentro de um espaço bidimensional, em que uma dimensão é a valência (i.e., o quão agradável é a emoção) e a outra é a excitação (i.e., a propensão do indivíduo de agir em função da emoção experimentada). Na Figura 3, é ilustrado esse modelo.

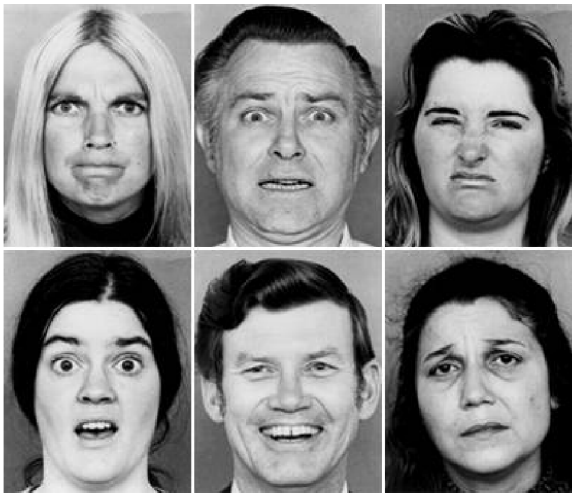


Figura 2 – Expressões faciais associadas às seis emoções em que consiste o modelo de Ekman. Na sequência: raiva, medo, aversão, surpresa, alegria e tristeza (EKMAN, 1999).

Figura 3 – Modelo para representação de emoções proposto por Russell, Lewicka e Niit (1989). Nesta figura, algumas emoções discretas são reconhecidas e mapeadas para o espaço apresentado. No entanto, as emoções passíveis de representação por esse modelo podem não se restringir a esse conjunto.

2.1.4 Reconhecimento Automático de Emoções

Muitas técnicas para reconhecimento automático de emoções podem ser encontradas na literatura. Não somente os algoritmos de aprendizado de máquina variam, como também as informações sensoriais usadas como entrada para o reconhecedor. A seguir, serão citados alguns exemplos, mencionados em Brave e Nass (2002):

- **Respostas neurológicas:** é possível associar informação proveniente de eletroencefalograma (EEG) a emoções, considerando que o cérebro é a fonte mais fundamental das

emoções.

- **Atividade autonômica:** componentes do sistema nervoso autonômico (SNA), como frequência cardíaca, pressão sanguínea, respiração, dilatação de pupila, condutividade da pele e tensão muscular, podem ser medidos com relativa facilidade.

- **Expressões faciais:** uma forma não invasiva e confiável de se reconhecer emoções. Essa abordagem, usada no reconhecedor a ser aplicado no presente trabalho, será detalhada adiante.

- **Voz:** entonação e velocidade da fala, por exemplo, podem ser interpretados como indicadores de características emocionais.

Alguns pesquisadores argumentam que um único método para classificação de emoções é insuficiente para lidar com determinadas situações, como o caso em que um usuário diz "olá" com um sorriso. Nessa situação, um reconhecedor de emoções baseado somente em expressões faciais pode falhar, visto que o ato de falar tende a gerar uma expressão facial diferente daquela que se refere a um sorriso característico (SEBE *et al.*, 2006). Nesse sentido, muitos trabalhos propõem sistemas multimodais para reconhecer emoções com maior precisão, abordagem que tem ganhado destaque nos últimos anos (DHALL *et al.*, 2015).

Neste trabalho, conforme já mencionado, foi usado um sistema que analisa somente de expressões faciais, em parte por ser o único disponível quando da concepção e da realização da maior parte deste projeto. Para se introduzir outras técnicas, seria necessário projetar e desenvolver um sistema dedicado a essa finalidade, o que poderia ser proibitivamente dispendioso, ainda que se tomasse como base o reconhecedor de emoções já disponível. Além disso, obter e submeter à análise somente um tipo de entrada sensorial ajudou a manter o foco do trabalho nas atividades mais relacionadas com seus objetivos específicos. Na subseção seguinte, serão apresentados mais detalhes acerca de reconhecimento de emoções por análise de expressões faciais.

2.1.4.1 Análise de Expressões Faciais

Tradicionalmente, como a maioria dos problemas de reconhecimento de padrões em imagens, a análise automática de expressões faciais pode ser dividida em três etapas, conforme são mostrados na Figura 4: detecção de face, extração de características e classificação. A entrada do sistema é uma imagem, na qual se espera que a face a ser analisada esteja contida. Gera-se, como saída, a emoção obtida, com uso de uma técnica de aprendizado de máquina.

Quando se pretende classificar imagens em tempo real, a detecção pode ser realizada a cada quadro (ou *frame*) ou apenas no primeiro, usando técnicas de rastreamento para os demais. Para os sistemas de classificação, de um ponto de vista estritamente técnico, é preferível que as

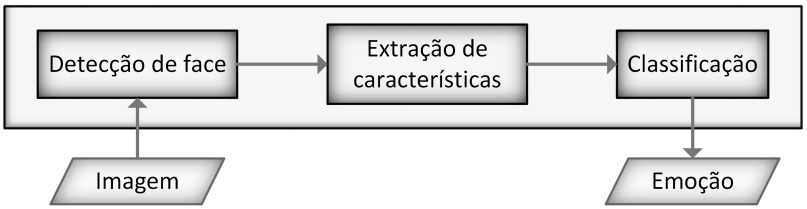


Figura 4 – Etapas da classificação de emoções baseada em análise de expressões faciais. O sistema de reconhecimento pode ser dividido em três etapas: detecção de face, extração de características e classificação.

faces sejam obtidas com vista frontal. Ruídos, movimento ou oclusão são alguns dos fenômenos que podem dificultar o procedimento (VALSTAR, 2008).

Em seguida o processo de extração de características, que cria representações dos fatores associados a diferentes emoções. Nesse estágio, são analisadas pistas que levam à identificação de um estado emocional. Como exemplos, podem-se citar deslocamentos de padrões faciais, como sobrancelhas erguidas, ou alterações na superfície da pele, como testa franzida. Para análise da expressão facial, dois tipos diferentes de métodos podem ser utilizados: métodos baseados em padrões geométricos, os quais identificam pontos de interesse nos componentes da face, como boca e sobrancelhas, e métodos baseados em aparência, os quais abordam mudanças na textura da face, como rugas e sulcos. Em ambos os casos, é gerado um vetor de características, para posterior análise.

Por fim, a classificação da expressão facial é realizada. Quando as imagens são extraídas de vídeos, esse problema pode ser tratado de duas formas. A primeira abordagem é aplicar um método de classificação a cada quadro, individualmente, podendo-se usar uma imagem de referência, relativa a uma face em estado emocional neutro. A segunda é usar sequências de quadros para realizar a classificação. Como classificadores, podem-se usar redes neurais artificiais, máquinas de vetores de suporte (SVMs – Support Vector Machines), modelos ocultos de Markov, aprendizado baseado em instâncias, modelos de regressão linear múltipla, entre outros (TIAN; KANADE; COHN, 2005).

2.2 Princípios de Robótica Social

O termo *robot* foi introduzido em 1921, por Karel Capek, na peça de teatro R.U.R. (*Russum's Universal Robots*). Deriva da palavra tcheca *robota*, que pode ser traduzida como trabalho compulsório. Uma coleção de histórias publicada pelo escritor de ficção científica Isaac Asimov, entre 1938 e 1942, pode ser considerada a responsável pela popularização dos robôs na ficção (HOCKSTEIN *et al.*, 2007). Especula-se que o termo *robotics*, ou *robótica*, tenha sido criado por Asimov, como derivação da palavra *robot*.

Pode-se definir um robô como "um sistema autônomo existente no mundo físico que pode sentir o ambiente e agir sobre ele para atingir objetivos" (MATARIC, 2007, tradução nossa). O Robot Institute of America (RIA), define um robô como "um manipulador reprogramável e

multifuncional projetado para mover materiais, partes, ferramentas ou dispositivos especializados através de diversas funções programadas para a realização de uma variedade de tarefas". (HEGEL *et al.*, 2009, tradução nossa).

Breazeal (2004) definiu um robô *sociável* como uma entidade que apresenta as seguintes características: corpo físico ou virtual, qualidades que simulem seres vivos, possibilitando aos humanos atribuir-lhe intencionalidade, capacidade de reconhecer pessoas e suas atividades, habilidade de aprender situações sociais por meio de imitação e meios para expressar-se e ser compreendido. Por enquanto, é difícil atender a todos esses requisitos simultaneamente, porém essa concepção serve como inspiração para a construção de sistemas que resolvam um subconjunto dessas características.

Simplificadamente, um robô social é um robô provido de interface social, combinando aspectos técnicos com aspectos sociais, de modo que os aspectos sociais sejam seu propósito fundamental (HEGEL *et al.*, 2009). Isso implica que o robô social deve comportar-se socialmente, em um dado contexto, e ter uma aparência entendida como social por qualquer usuário. A interface social, portanto, encerra as características por meio das quais o usuário percebe o robô como um ente com qualidades sociais. A Figura 5 ilustra essa ideia.

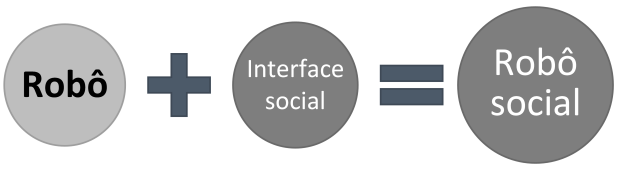


Figura 5 – Um robô dotado de interface social é um robô social. Adaptado de Hegel *et al.* (2009).

A construção de robôs sociais está intimamente ligada à área de interação humano-robô (HRI – *Human-Robot Interaction*), a que é atribuída a tarefa de criar robôs que possam interagir de forma natural com seres humanos. O domínio de problemas de HRI inclui busca e resgate, robótica educacional e assistiva, entretenimento, aplicações militares e policiais, entre outras aplicações que podem requerer, em algum grau, habilidades sociais (GOODRICH; SCHULTZ, 2007).

Em Damiano, Dumouchel e Lehmann (2015), é exposta uma visão geral da pesquisa envolvendo robôs, emoções e empatia, área que o autor classifica como uma das mais ativas da ciência contemporânea. O artigo classifica as pesquisas na área dentro de duas dimensões distintas, chamadas *externa* e *interna*. A dimensão externa corresponde aos aspectos sociais das emoções, enquanto a dimensão interna lida com seus aspectos individuais. A ideia de um *loop* afetivo, argumenta o autor, tem emergido entre essas duas dimensões, caracterizando-se por robôs que apresentam tanto a dimensão interna quanto a externa.

A dimensão externa trata de estudos sobre a expressividade que movimentos, gestos e posturas podem conferir ao robô. Pretende-se dotar os robôs de competências que lhes permitam interagir de forma eficiente e natural com os seres humanos, melhorando a sua aceitação. Essa

dimensão tem se mostrado uma plataforma fértil para investigar características de antropomorfismo em objetos, sendo a conjectura do *uncanny valley* (MORI; MACDORMAN; KAGEKI, 2012) um dos seus princípios-chave. Segundo essa conjectura, o robô produz uma sensação de familiaridade mais alta quanto mais semelhante é de um ser humano, até se atingir um determinado nível de semelhança. Entretanto, quando essa semelhança se torna muito grande, o robô passa a ser visto como perturbador, ou mesmo monstruoso.

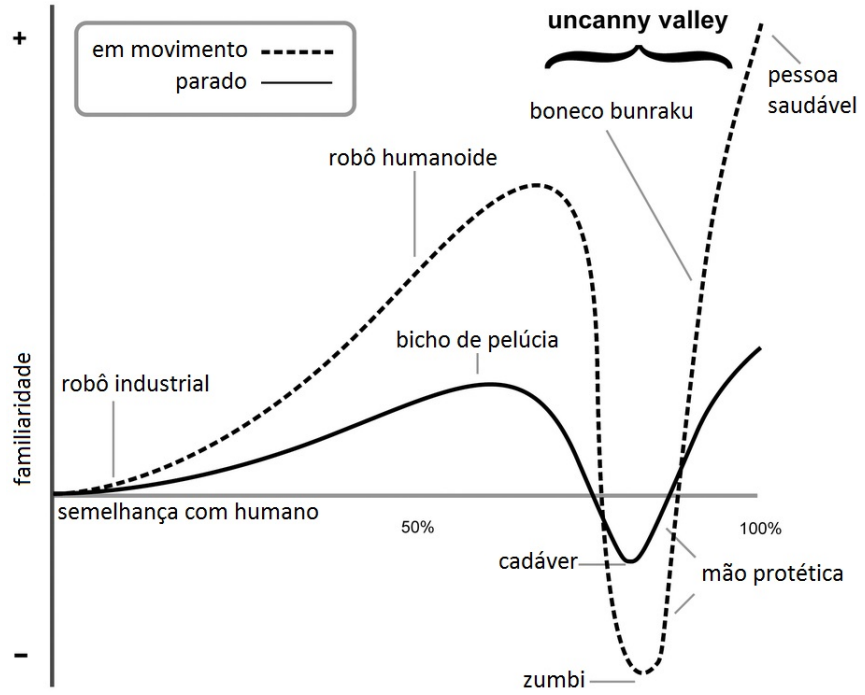


Figura 6 – Representação do *uncanny valley*. Extraído e traduzido de Mori, MacDorman e Kageki (2012).

A noção de *empatia* é central para o desenvolvimento de estudos na dimensão externa dessa pesquisa. Decety e Cowell (2014) conceituaram essa noção de acordo com três componentes: emocional, motivacional e cognitiva. A componente emocional é a capacidade de responder afetivamente às emoções de outras pessoas; a motivacional é o ímpeto de buscar o bem-estar de outras pessoas; e a cognitiva é a habilidade de tomada de perspectiva.

A dimensão interna estuda o papel dos processos emocionais no comportamento dos indivíduos. Os objetivos são reproduzir artificialmente os processos emocionais, a fim de investigar o papel das emoções na organização do comportamento, e aplicar esse conhecimento, a fim de construir robôs com maior autonomia. Portanto, esses estudos são direcionados, à modelagem das emoções dos seres vivos em agentes artificiais.

Há estudos relevantes envolvendo a interações com agentes virtuais (STRATOU *et al.*, 2015), e muito do trabalho desenvolvido encontra intersecção com as pesquisas com robôs sociais, ainda que a ausência de presença física implique na ausência de efeitos relevantes produzidos por robôs, como o da proxêmica (BAINBRIDGE *et al.*, 2011). De todo modo, com exceção de aprimoramentos em sistemas de percepção ou de avanços em arquiteturas cognitivas

para agentes sociais, a maioria dos estudos nessas áreas é avaliada a partir de estudos com seres humanos.

Muitos trabalhos em interação humano-robô, especialmente com robôs sociais, contam com experimentos seguindo abordagens do tipo *Wizard of Oz*. Esses experimentos são caracterizados por "um indivíduo (geralmente o experimentador ou um parceiro) operando remotamente um robô, controlando diversas entidades, como movimento, navegação, fala, gestos etc."(RIEK, 2012, tradução nossa). Nesse cenário, os participantes do experimento não são informados de que o robô é tele-operado, atuando como se estivessem lidando com um agente autônomo. O objetivo é dar flexibilidade aos estudos empíricos, reduzindo a necessidade de se obter ou produzir aparato tecnológico específico para as aplicações experimentadas.

2.2.1 Estatística para Experimentos com Humanos

Experimentos com humanos são conduzidos de modo a avaliar um conjunto de *fatores*. Trata-se dos aspectos da abordagem em análise passíveis de produzir alterações nos resultados. Para avaliar cada fator, é considerado um conjunto de *condições*, que são as variações imputadas aos fatores analisados. Por exemplo, para avaliar de que modo uma determinada estratégia influencia o engajamento dos usuários diante de um agente, pode-se comparar uma condição implementando a estratégia proposta a uma condição controle, desprovida da referida estratégia.

A fim de examinar a influência das condições sobre os fatores, os experimentos conduzidos podem ser *intra-sujeito*, quando todos os participantes são submetidos a todas as condições, ou *entre sujeitos*, quando grupos distintos de participantes são atribuídos a cada uma das condições (NAVARRO, 2014). Nos dois casos, podem ser avaliadas métricas objetivas, como tempo de hesitação dos participantes para executar uma ação (DARLING; NANDY; BREAZEAL, 2015), ou subjetivas, como avaliações dos participantes em escalas discretas de valores numéricos (GÁCSI *et al.*, 2016). Testes de significância estatística podem ser aplicados sobre os resultados, permitindo reconhecer quais aspectos, nas diferentes condições, exercem influência sobre os fatores analisados.

Uma apresentação teórica detalhada sobre os testes de significância é apresentada em Spiegel, Schiller e Srinivasan (2012). Esses testes, também conhecidos como testes de hipóteses, aplicam preceitos estatísticos de modo a provar que uma hipótese nula H_0 é falsa. Por exemplo, para verificar se uma moeda está viciada, formula-se a hipótese nula de que a moeda está equilibrada. Assim, tem-se $H_0 : P = 0,5$, onde P é a probabilidade do resultado *cara*. Em seguida, aplica-se um teste de significância para provar que a hipótese nula enunciada é falsa, portanto a hipótese alternativa $H_1 : P \neq 0,5$ é verdadeira.

Dois tipos de erros de decisão podem suceder a aplicação de um teste de significância. São eles o erro Tipo I, quando se rejeita uma hipótese verdadeira, e o erro Tipo II, quando se aceita uma hipótese falsa. Em geral, esforços para reduzir a incidência de erros de um desses

tipos tendem a aumentar as chances de se cometer erros do outro tipo. A única forma de reduzir a incidência de ambos os tipos de erro simultaneamente é aumentar o tamanho da amostra.

O nível de significância, denotado por α, é definido como a probabilidade máxima aceitável de se cometer um erro Tipo I. Por exemplo, um nível de significância $\alpha = 0,05$ implica em uma probabilidade de 5% de se rejeitar incorretamente a hipótese nula e 95% de confiança de que a decisão tomada está correta.

Muitos testes de significância, denominados testes paramétricos, assumem que as amostras em análise foram extraídas de uma distribuição de probabilidade específica, geralmente uma distribuição normal. Portanto, a aplicação desses testes deve ser precedida de algum teste para validar a qualidade do ajuste das amostras obtidas à distribuição assumida, como o teste Chi-Square ou o teste Kolmogorov-Smirnov. Existem, também, os testes não-paramétricos, que não fazem hipóteses acerca da distribuição de probabilidade da qual as amostras foram extraídas. Esses testes, entretanto, não foram aplicados neste trabalho.

Em testes paramétricos envolvendo a distribuição normal, o *valor-p* é o menor nível de significância com que não se rejeita a hipótese nula. Dito de outra maneira, a hipótese nula é falsa se $p \leq \alpha$, para um dado nível de significância α.

Esses testes podem ser unicaudais ou bicaudais. Quando se está verificando a hipótese de que uma variável tem média superior em uma condição do que em outra, existe interesse em somente uma cauda (i.e., região crítica) da distribuição, sendo adequada a aplicação de um teste unicaudal. Por outro lado, quando se está verificando somente se as médias são diferentes nas duas condições, não havendo direção específica para essa diferença, é adequada a aplicação de um teste bicaudal. Informações mais específicas a esse respeito serão apresentadas na subseção 2.2.1.2.

A seguir, serão apresentados os testes aplicados para análise dos resultados obtidos neste trabalho, apresentados no Capítulo 5. São eles o teste de aderência à normalidade Kolmogorov-Smirnov e o teste-t pareado.

2.2.1.1 Teste Kolmogorov-Smirnov

Sejam $X = \{x_1, x_2, \ldots, x_N\}$ a amostra em análise, $F(x)$ a função de distribuição acumulada assumida para os dados (e.g., a distribuição normal acumulada) e $F_N(x)$ a distribuição acumulada empírica dos dados. A estatística D_n usada para o teste é representada pela Equação 2.1:

$$D_n = \sup_x |F(x) - F_n(x)|, \quad x \in X \tag{2.1}$$

Por meio da formulação discutida em O'Connor e Kleyner (2011), mostra-se que o ajuste é bom se $D_n \leq D_{n,\alpha}$, onde α é o nível de significância, normalmente tomado como $\alpha = 0,05$. Embora os valores críticos possam ser calculados, são comumente disponibilizados na forma de

tabela, para o caso de $F(x)$ ser uma distribuição normal. A Tabela 1 mostra os valores críticos para diferentes tamanhos de amostra e níveis de significância.

Tabela 1 – Valores críticos para aplicação do teste de aderência à normalidade Kolmogorov-Smirnov. Adaptado de O'Connor e Kleyner (2011).

$n \setminus \alpha$	0.10	0.05	0.02	0.01
1	0,95000	0,97500	0,99000	0,99500
2	0,77639	0,84189	0,90000	0,92929
3	0,63604	0,70760	0,78456	0,82900
4	0,56522	0,62394	0,68887	0,73424
5	0,50945	0,56328	0,62718	0,66853
6	0,46799	0,51926	0,57741	0,61661
7	0,43607	0,48342	0,53844	0,57581
8	0,40962	0,45427	0,50654	0,54179
9	0,38746	0,43001	0,47960	0,51332
10	0,36866	0,40925	0,45662	0,48893
11	0,35242	0,39122	0,43670	0,46770
12	0,33815	0,37543	0,41918	0,44905
13	0,32549	0,36143	0,40362	0,43247
14	0,31417	0,34890	0,38970	0,41762
15	0,30397	0,33760	0,37713	0,40420
16	0,29472	0,32733	0,36571	0,39201
17	0,28627	0,31796	0,35528	0,38086
18	0,27851	0,30936	0,34569	0,37062
19	0,27136	0,30143	0,33685	0,36117
20	0,26473	0,29408	0,32866	0,35241

2.2.1.2 Teste-t Pareado

Uma apresentação sobre o teste-t e suas implicações práticas pode ser encontrada em Navarro (2014). Quando se pretende comparar duas amostras usando o teste-t, deve-se observar se elas são independentes ou dependentes. Para amostras dependentes, deve ser aplicado o teste-t pareado (em contraste com o teste-t independente).

Se $X = \{x_1, \ldots, x_N\}$ e $Y = \{y_1, \ldots, y_N\}$ são duas amostras dependentes, considerar observações pareadas significa tratar esses dados como uma única amostra, formada pelos pares $(x_1, y_1), \ldots, (x_N, y_N)$. Note que essa abordagem reflete o que ocorre nos experimentos *intra-sujeito*, como os realizados neste trabalho.

O teste-t pareado toma as diferenças dos valores relativos a cada par, de modo que $D = X - Y$. Considere $D_i \sim N(\mu_D, \sigma_D^2)$, onde μ_D é a média amostral das diferenças, estimada por $\bar{D}$, e σ_D^2 é a variância amostral das diferenças. Note que, se duas amostras correlacionadas apresentam distribuição normal, a diferença entre elas também segue a distribuição normal (HUNTER; HUNTER; GEORGE, 1978).

As hipóteses devem ser estabelecidas conforme segue, considerando, respectivamente, os casos unicaudal à esquerda, unicaudal à direita e bicaudal.

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D < 0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 \end{cases}$$

O teste é realizado pela Equação 2.2, onde S_D é o desvio padrão amostral. Sob H_0, T segue uma distribuição t de Student com $n-1$ graus de liberdade.

$$T = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \tag{2.2}$$

Dependendo do tipo de teste aplicado, consideram-se regiões críticas diferentes, conforme ilustra a Figura 7. Para o teste unicaudal à esquerda, o ponto crítico é $-t_\alpha$, e H_0 é rejeitada se $T < -t_\alpha$. Para o teste unicaudal à direita, o ponto crítico é t_α, e H_0 é rejeitada se $T > t_\alpha$. Para o teste bicaudal, os pontos críticos são $-t_\alpha/2$ e $t_\alpha/2$, e H_0 é rejeitada se $T > t_{\alpha/2}$ ou $T < -t_{\alpha/2}$.

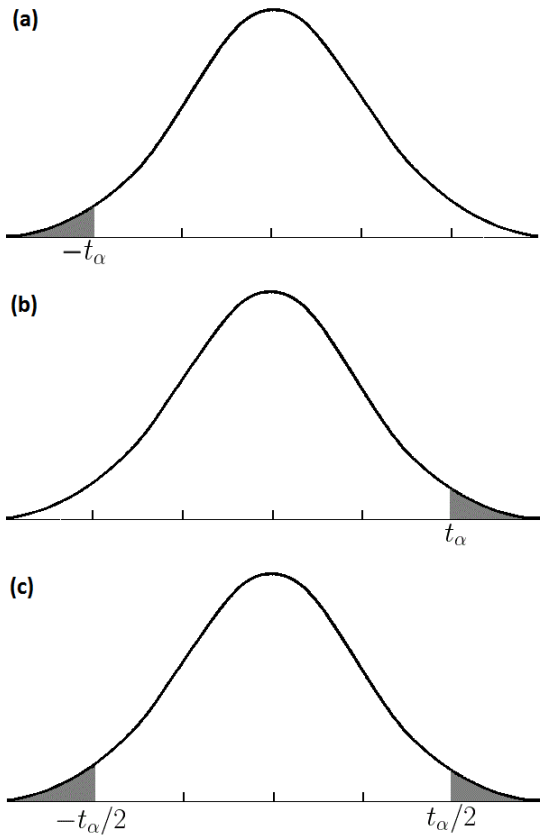


Figura 7 – Regiões críticas para o teste-t pareado (a) unicaudal à direita; (b) unicaudal à esquerda; e (c) bicaudal.

O valor-p, já definido, é mais inteligível para decidir sobre significância. Seu cômputo é realizado conforme segue, onde $P(x)$ representa probabilidade:

$$\begin{cases} p = P(t < T|H_0), \text{ para teste unicaudal à esquerda} \\ p = P(t > T|H_0), \text{ para teste unicaudal à direita} \\ p = P(|t| > |T||H_0) = 2P(t > |T||H_0), \text{ para teste bicaudal} \end{cases}$$

2.3 Considerações Finais

Foram apresentados os principais conceitos teóricos sobre emoções, incluindo sua definição, suas bases cognitivas e os mecanismos neurofisiológicos subjacentes. Também foram discutidos modelos computacionais de emoções e técnicas para reconhecimento automático. Em seguida, foram discutidos conceitos de interação humano-robô, com enfoque em robótica social. Por fim, foram introduzidos princípios de estatística para experimentos com humanos, seguidos das técnicas aplicadas neste trabalho.

Esses fundamentos serão usados como referencial no decorrer de toda a dissertação. No próximo capítulo, serão discutidos alguns trabalhos relacionados ao tema deste trabalho.

CAPÍTULO

3

TRABALHOS RELACIONADOS

Neste capítulo, serão apresentados trabalhos que se relacionam, de diversas formas, com a proposta apresentada nesta dissertação. Primeiro, serão apresentadas aplicações de emoções em diferentes tipos de robôs sociais e agentes virtuais personificados. Em seguida, será descrito o sistema para reconhecimento de emoções adotado neste trabalho.

3.1 Robótica Social e Emoções

O aperfeiçoamento de técnicas para reconhecimento de emoções é importante não somente para a robótica social, mas para qualquer aplicação computacional que demande *feedback* afetivo por parte do usuário. Por esse motivo, muitas pesquisas recentes concentram-se na concepção de sistemas para reconhecimento de emoções.

Um reconhecedor de emoções para robôs foi apresentado em Rabiei e Gasparetto (2014), identificando as seis emoções básicas de Ekman (1999) e a emoção neutra. Para isso, foi realizada análise propriedades fonéticas e acústicas no sinal da fala. Entre essas propriedades, estão a agudeza, a intensidade, a velocidade e formantes. Um grupo de adultos europeus realizou gravações de frases pré-determinadas, dando a entonação adequada, e essas gravações foram usadas para compor uma base de dados. A classificação foi feita por meio de regras, obtidas de observações na base criada. Por considerar os resultados pouco acurados para uma aplicação concreta de interação humano-robô, os autores propuseram, como trabalho futuro, combinar esse classificador com classificadores baseados em análise de expressões faciais.

Como já mencionado na subseção 2.1.4, uma forma de aprimorar o reconhecimento de emoções é desenvolver sistemas multimodais, os quais tomam mais de um tipo de informação como entrada. Em Alonso-Martin *et al.* (2013), por exemplo, foi apresentado um sistema multimodal para detecção da emoção do usuário por robôs sociais. São combinadas análises de características da fala, com o componente Gender and Emotion Voice Analysis (GEVA), e

de expressões faciais, com o componente Gender and Emotion Facial Analysis (GEFA). Esse módulo foi incorporado ao sistema de interação apresentado, denominado Robotics Dialog System (RDS). Ambos os módulos foram avaliados com várias diferentes técnicas para extração de características e classificação. Uma regra de decisão é aplicada para combinar os resultados do GEVA e do GEFA, integrando a emoção reconhecida ao sistema de diálogo.

Recentes avanços em reconhecimento multimodal de emoções incentivaram o surgimento do *Emotion Recognition Challenge in the Wild*, ou *EmotiW* (DHALL *et al.*, 2015). Os dados do *EmotiW* são utilizados para produzir pesquisa na área de reconhecimento multimodal de emoções. Kächele *et al.* (2016), por exemplo, apresentaram uma técnica para fusão, baseada em computação evolutiva, para dados provenientes de vídeo e áudio, avaliando e comparando essa técnica usando esses dados. No artigo, uma discussão sobre os trabalhos submetidos às três edições já realizadas desse desafio, iniciado em 2013, é apresentada em conjunto aos resultados.

Uma abordagem que levou a bons resultados foi apresentada em Kahou *et al.* (2013). Trata-se da submissão da equipe da Universidade de Montreal para o *EmotiW 2013*. Foram utilizadas *Convolutional Neural Networks* (CNN) para classificação de emoções nos quadros da imagem. O áudio, por sua vez, foi classificado com uso de *Deep Belief Networks* (DBN), outro tipo de rede neural profunda, sobre 29 características extraídas dos arquivos de áudio. Outras informações também eram consideradas, sendo, por fim, combinadas no resultado final.

Embora o reconhecimento de emoções seja uma tarefa muito importante para o desenvolvimento da robótica social, trata-se da etapa inicial de um processo que deve culminar em mudanças nas estratégias de interação. Em outras palavras, obtido o resultado do procedimento de classificação, é preciso saber o que fazer com essa informação.

Nessa direção, a pesquisa em robótica social pode ser dividida, para fins de organização, em dois tipos de abordagem: estudos empíricos ou desdobramentos tecnológicos. Estudos empíricos são as pesquisas realizadas para entender as reações humanas diante de robôs, em diferentes situações, e procurar estabelecer alguns padrões. Desdobramentos tecnológicos são os desenvolvimentos de robôs sociais, agentes virtuais e estratégias de interação, isto é, de novas tecnologias que possam aprimorar as interações entre humano e robô. As subseções seguintes consideram essas abordagens separadamente.

3.1.1 Estudos Empíricos em Interação Humano-Robô

Em um estudo envolvendo 81 voluntários com idade entre 18 e 26 anos, Gácsi *et al.* (2016) investigaram se humanos podem atribuir emoções a robôs que manifestam padrões comportamentais baseados em cães. Para isso, foram gravados vídeos em que um cão realizava movimentos associados a diferentes emoções. Esses movimentos eram uma combinação entre emoções reais, induzidas pela comunicação com o dono, e instruções pré-determinadas por adestramento. Posteriormente, foram gravados vídeos do robô móvel *PeopleBot*, mostrado na

Figura 8, programado para imitar esses movimentos. Foi necessário adaptar alguns comportamentos, levando em conta as limitações impostas pelos atuadores disponíveis em sua arquitetura de *hardware*. O robô também produziu ruídos correspondentes às vocalizações do cão. Ao todo, foram gravados cinco vídeos do cão e cinco vídeos do robô, todos com duração de 8 a 14 segundos.

Figura 8 – Robô *PeopleBot*. Extraído de Gácsi *et al.* (2016).

Em uma abordagem *intra-sujeito*, todos os participantes assistiram a todos os vídeos e o fizeram em seis grupos de 11 a 17 pessoas. Para contrabalancear o experimento, metade deles assistiu primeiro aos cinco vídeos do cão, depois aos cinco vídeos do robô, enquanto a outra metade fez o contrário. Houve duas exibições dos vídeos, na mesma sequência, para cada grupo. Durante a primeira exibição, os participantes foram solicitados a relatar, na visão deles, o que ocorreu no vídeo. A partir dessas respostas, os autores do trabalho atribuíram, a partir de alguns critérios objetivos, uma nota de 1 a 4 para quantificar o quanto os participantes reconheceram a emoção emitida pelo robô ou cão. Já na segunda exibição, foram aplicadas questões de múltipla escolha, questionando quais estados emocionais os participantes observaram e com qual nível de certeza, em uma escala de 0 a 4. Os participantes atribuíram emoções tanto ao robô quanto ao cão, ainda que o tenham feito com maior intensidade em relação ao cão.

Um passo importante para o desenvolvimento de interações sociais entre humanos e robôs pode ser verificar a capacidade de se produzir empatia nessa relação. Essa avaliação foi realizada em Putten *et al.* (2013). O experimento verificou as reações emocionais de humanos diante do dinossauro robótico *Ugobe's Pleo* (Figura 9) mostrado em diferentes situações. Estabeleceu-se uma configuração experimental 2x2, com um fator *entre sujeitos* e um fator *intra-sujeito*. O fator *entre sujeitos* é o histórico de interação com o robô, dividindo os 41 voluntários entre os que nunca haviam visto o robô e os que tiveram 10 minutos de interação prévia com ele. O fator *intra-sujeito* referia-se ao tipo de vídeo, que poderia conter uma interação amigável ou uma forma de tortura sobre o robô.

O estado emocional dos voluntários foi verificado analisando medições fisiológicas e emoções auto-reportadas. Os resultados das duas medições apontaram que a existência de interação prévia com o robô não influenciou os resultados, porém o tipo de vídeo o fez. Os

Figura 9 – Robô *Ugobe's Pleo*. Extraído de Putten *et al.* (2013).

voluntários que assistiram ao vídeo de interação amigável experimentaram emoções amenas e positivas, enquanto os voluntários que assistiram ao vídeo de tortura experimentaram emoções mais intensas e negativas, relatando terem sentido empatia.

Esse trabalho inspirou a pesquisa apresentada em Darling, Nandy e Breazeal (2015). Na ocasião, foi realizado um experimento para verificar o quanto alguns fatores poderiam influenciar o tempo de hesitação dos participantes humanos para cumprir a tarefa de destruir um inseto robótico com uso de uma marreta. Os materiais são mostrados na Figura 10. O primeiro fator consistia em mostrar uma história fictícia sobre o robô antes de apresenta-lo, enquanto o segundo fator consistia em faze-lo executar um curto movimento.

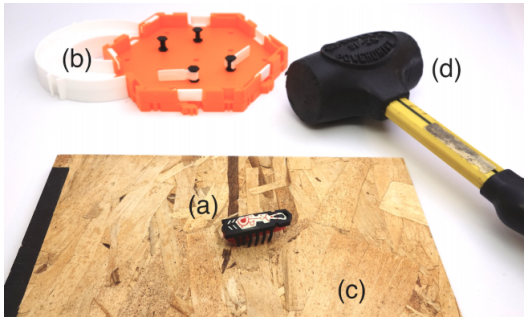


Figura 10 – Materiais do experimento descrito em Darling, Nandy e Breazeal (2015). São eles: (a) inseto robótico; (b) espaço de confinamento para movimento; (c) mesa; (d) marreta. Extraído de Darling, Nandy e Breazeal (2015)

Um conjunto de 101 participantes foi dividido em seis grupos, em uma configuração experimental 3x2, com todos os fatores *entre sujeitos*. Como primeiro fator, foram consideradas três condições: sem história, uma história com um elemento personificação (por exemplo, um nome foi atribuído ao robô) ou uma história sem o elemento personificação. No segundo fator, duas condições possíveis foram consideradas, correspondendo a se o robô se moveu ou não no início da sessão. Os autores mediram o intervalo entre o pedido para destruir o robô e a iniciativa do participante em executar a ação, interpretado como *hesitação*. Os resultados mostraram que ambas as condições do fator *história* aumentaram significativamente a hesitação, em comparação com a condição controle (sem história). O fator *movimento* não produziu alterações significativas.

Outro resultado correlato é descrito em Jo *et al.* (2013). Foi conduzido um estudo para

examinar os aspectos sociais da interação humano-robô e verificar se é criada empatia nesse contexto. Um total de 37 estudantes de graduação foram tomados como voluntários, tendo sido divididos em dois grupos. Atribuiu-se a um grupo uma companhia humana e, ao outro, a de um robô *Nao*. Os participantes de ambos os grupos, cada um individualmente, assistiram a dois vídeos de humor. No primeiro vídeo, foram introduzidos risos gravados nas partes supostamente engraçadas, tratamento que não foi dado ao segundo vídeo. Na seção em que o primeiro vídeo foi exibido, os voluntários permaneceram sozinhos na sala, enquanto, na exibição do segundo vídeo, foi introduzida a companhia do humano ou do robô (dependendo do grupo a que o voluntário pertencia), que ria nos momentos pretensamente engraçados.

Por meio de questionários, aplicados após cada seção, verificou-se que os voluntários se divertiram mais quando havia companhia. Entretanto, o fato dessa companhia ser humana ou do robô não exerceu influência significativa no resultado. Os autores apontaram esse resultado como indício de que os voluntários entenderam a presença do robô como uma companhia real com a qual haviam assistindo ao vídeo. Isso pode significar que houve interação empática entre os participantes do estudo e o robô, diante da experiência de emoções positivas.

É importante reconhecer que robôs com presença física tendem a impor uma presença social mais proeminente, como foi verificado em Wainer *et al.* (2006). Conduziu-se um estudo com uma amostra de 11 participantes, que foram solicitados a jogar o clássico enigma *torre de Hanói*. As hastes e discos do desafio foram posicionados entre o usuário e o mediador, que poderia ser um robô ou um agente virtual. O mediador teria o papel de dar instruções aos participantes e acompanhar o processo, detectando jogadas não permitidas. Três condições foram consideradas, relacionadas aos diferentes tipos de mediador: um robô móvel *ActivMedia Pioneer 2 DX* em presença física, o mesmo robô em presença remota (vídeo-conferência) e uma representação virtual do robô, modelada por computação gráfica. Tratou-se de um experimento *intra-sujeito*, isto é, cada um dos participantes experimentou as três condições, atribuídas em ordem aleatória. Através de questionários, verificou-se que o mediador robótico, com presença física ou remota, foi entendido de forma mais agradável do que o agente virtual.

Ainda assim, agentes virtuais são introduzidos, com sucesso, em muitas pesquisas na área, como as recepcionistas virtuais que serão apresentadas na próxima subseção (KIRBY; FORLIZZI; SIMMONS, 2010). Particularmente, os *smartphones* são dotados de conjuntos de sensores e características úteis para testar estrategias sem a necessidade de robôs fisicamente presentes. Com mudanças na aparência de um agente para *smartphones*, ainda, é possível usar robôs apropriados e, com algumas adaptações, portar o ambiente para um robô móvel, como pode ser encontrado em Ranieri *et al.* (2014).

3.1.2 Desdobramentos Tecnológicos

Na subseção anterior, foram descritos estudos empíricos, que trazem importantes resultados para o desenvolvimento deste trabalho. A partir de agora, serão apresentadas pesquisas

voltadas a desdobramentos tecnológicos em robótica social.

Uma das primeiras experiências bem-sucedidas na criação de robôs com características sociais foi o robô Kismet, mostrado na Figura 11. Trata-se de um humanoide com muitos graus de liberdade na face, que o tornaram capaz de reproduzir uma grande variedade de expressões faciais humanas. Em Breazeal (2003), foi detalhada uma arquitetura que visava selecionar as emoções a serem produzidas pelo robô Kismet, em dadas circunstâncias, bem como as formas de expressa-las.

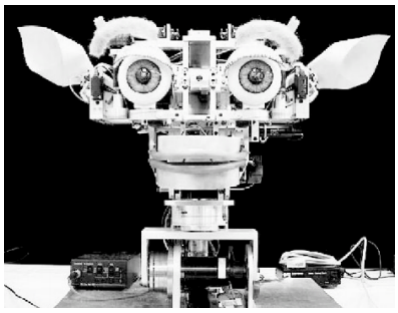


Figura 11 – Robô Kismet. Extraído de Breazeal (2003).

Outros robôs capazes de exibir emoções humanas foram desenvolvidos, e uma aplicação recente pode ser encontrada em Arnold (2016). Trata-se do robô *Emobie*, que visaria auxiliar o tratamento de crianças que sofrem de ansiedade. Emoções seriam manifestadas através de expressões nos olhos, ruídos, movimentos dos braços e coloração do estômago. O robô se comportaria de modo a refletir as emoções das crianças. No artigo, não foram fornecidas informações sobre experimentos com usuários, que seriam conduzidos em uma etapa posterior.

São de interesse a esta dissertação, por outro lado, aplicações que produzam reações a estados emocionais inferidos de usuários humanos. Nesse sentido, Leite *et al.* (2012) criaram um ambiente em que o robô *iCat* jogaria xadrez contra crianças com idade entre 8 e 10 anos. A configuração do experimento consta na Figura 12. O robô não movimentava as peças no tabuleiro, cabendo à criança posicionar as peças de acordo com o movimento enunciado pelo robô, por meio de síntese de fala. Uma *webcam* era usada para capturar expressões faciais da criança, permitindo inferir, com base na probabilidade de a criança ter sorrido nos últimos seis segundos, na direção do olhar da criança e no contexto da partida, a valência de seu estado emocional. As classificações possíveis eram *positivo*, *neutro* ou *negativo*.

Essas inferências eram usadas para que, quando se detectasse que a criança não estava experimentando boas emoções, o robô empregasse uma estratégia de empatia, que poderia ser uma dentre as quatro seguintes: fazer comentários encorajadores, comentar a última jogada do usuário ou permitir-lhe refazê-la, oferecer ajuda ou realizar uma má jogada deliberadamente. Foi implementado um algoritmo de aprendizado por reforço para selecionar a estratégia a ser adotada, uma vez que cada uma delas poderia surtir efeitos diferentes em usuários diferentes.

Além de robôs, a pesquisa em robótica social também pode ser conduzida com agentes

Figura 12 – Robô iCat jogando xadrez com criança. Extraído de Leite *et al.* (2012).

virtuais personificados. Em Gockley *et al.* (2005), a recepcionista virtual Valerie atuou na entrada de um prédio da Universidade Carnegie Mellon. Ao interagir com os frequentadores, a personagem lhes contava, gradualmente, histórias pessoais a ela atribuídas. Decorrido certo período, muitos frequentadores continuavam a interagir diariamente, contudo dificilmente o faziam por mais de 30 segundos, o e indicou um decaimento no engajamento da interação. Kirby, Forlizzi e Simmons (2010) estenderam essa pesquisa, imputando a Valerie a capacidade de expressar três diferentes humores (positivo, neutro ou negativo), condizentes com a história que contava. Através de análise de comportamento e aplicação de questionários, observou-se que o padrão de interação varia de acordo com o humor exibido, e a forma como esses padrões se modificam dependia do nível de familiaridade do frequentador com a recepcionista.

Nesse mesmo contexto, Trovato *et al.* (2015) desenvolveram uma recepcionista virtual, a que chamaram Ana. Foram analisadas informações demográficas do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Ministério do Trabalho, a fim de criar uma personagem com traços físicos característicos de recepcionistas brasileiras. No intuito de comparar dois tipos de personificação, o experimento realizado introduziu, também, o robô humanoide Kobiana. Ana e Kobiana estão representados na Figura 13. Foram recrutados 60 participantes, os quais foram divididos em dois grupos: o grupo EV, compreendido por funcionários e visitantes do centro de pesquisa em que o estudo foi conduzido, e o grupo CL, compreendido por pessoas de pouca instrução, que faziam parte da equipe de limpeza da mesma instituição.

Todos os participantes interagiram primeiro com Ana, depois com Kobiana. No caso de Kobiana, foi adotada uma abordagem *Wizard of Oz*, em que o robô Kobiana emitia um pequeno monólogo durante uma comunicação por videoconferência simulada. Após cada abordagem, em condições controladas, os participantes responderam a questionários sobre sua experiência. Na avaliação dos resultados, observou-se que os dois grupos de usuários mostraram melhor aceitação a Ana do que a Kobiana. Conforme observado pelos autores, esse resultado pode ter sido influenciado pelo fato de ambas as recepcionistas terem sido mostradas em uma tela, eliminando o efeito da proxêmica. O grupo CL percebeu ambas as recepcionistas de forma mais antropomórfica.

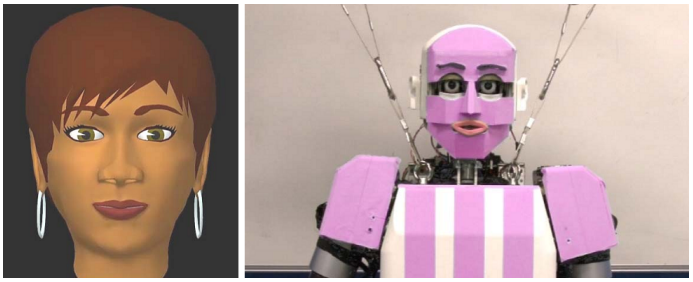


Figura 13 – Agente virtual Ana (esquerda) e robô humanoide Kobiana (direita).

Dispositivos móveis podem ser uma plataforma conveniente para agentes virtuais personificados, considerando a disponibilidade de sensores e atuadores voltados para interação social (e.g., câmera frontal, microfone, alto-falantes) (ALVES; FERASOLI, 2016) e a presença constante desses dispositivos na vida das pessoas (VINCENT, 2013).

Um agente virtual para *smartphones*, dotado de emoções, foi apresentado em Yoon e Cho (2012). Trata-se de um assistente pessoal, ao qual foi incorporada a representação de um cão virtual, como mostra a Figura 14. O sistema foi dotado de cinco componentes: sistema de percepção, sistema de reconhecimento de contexto, sistema de emoções, sistema de motivação e sistema de geração de comportamentos. O sistema de percepção é responsável por obter informações do usuário a partir de dados disponíveis no *smartphone*, como lista de chamadas, agenda e local. O sistema de reconhecimento de contexto usa informações provenientes do sistema de percepção para inferir a emoção do usuário, em um modelo dimensional valência-excitação, e o seu *business state* (i.e., o quanto o usuário se encontra ocupado). O sistema de emoções é responsável pela síntese de emoções do sistema, representadas por um modelo discreto com 14 categorias. O sistema de motivação controla os objetivos a serem atingidos pelo sistema. Finalmente, o sistema de geração de comportamentos é incumbido de selecionar o comportamento mais adequado para a presente situação.

A adequação dos comportamentos gerados foi avaliada por meio de um teste de usabilidade, em uma abordagem *intra-sujeito*. Oito voluntários experimentaram o sistema em dez cenários. Em cada cenário, os participantes observaram uma sequência de cinco comportamentos gerados aleatoriamente e uma outra sequência de cinco comportamentos gerados com o método proposto, cabendo a eles avaliar se o comportamento lhes pareceu adequado para a situação. Através da análise dos resultados, foi comprovado que a abordagem apresentada gera comportamentos mais apropriados do que a condição controle.

Uma forma de integrar agentes virtuais personificados a robôs sociais é aplicar dispositivos móveis como parte de sistemas robóticos. O aplicativo *Smartphone Intuitive Likeness and Expression* (SMILE), apresentado em Russell *et al.* (2014), era uma interface usada para produzir emoções animadas, sintetizar fala e aprender vocabulário. O *smartphone* foi acoplado ao robô humanoide UM-L8, voltado ao público infantil. No aplicativo, ilustrado na Figura 15, era mostrado um par de olhos, que piscavam de forma intermitente e sofriam modulações para

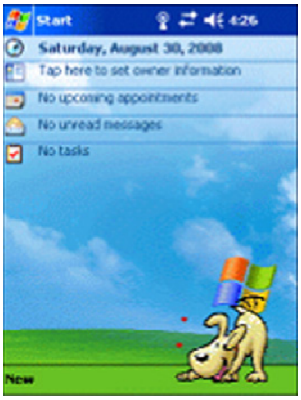


Figura 14 – Assistente virtual com personagem sintético para interação. Extraído de Yoon e Cho (2012).

corresponder a diferentes emoções. Ocorriam síntese de emoções no aplicativo, sem haver reconhecimento das emoções do usuário. As emoções representadas eram alegria, surpresa, tristeza, raiva, confusão e neutralidade.

O aplicativo reconhecia 30 palavras-chave, associando cada uma delas a uma dentre cinco emoções. A ocorrência de alguma dessas palavras na frase dita pelo usuário e inferida pelo reconhecedor de fala propulsionava a mudança da expressão mostrada na tela. Algumas palavras-chave produziam respostas pré-definidas, por fala sintetizada. Novas emoções poderiam ser ensinadas para o aplicativo em tempo de execução, com uso dos comandos apropriados. Os experimentos realizados foram exploratórios, apontando as direções para aprimoramentos.

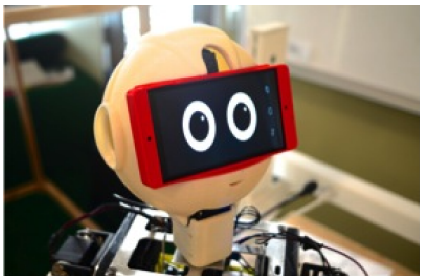


Figura 15 – Ambiente com uso da plataforma SMILE. Extraído de Russell *et al.* (2014).

Essa associação entre dispositivos móveis e sistemas robóticos pode ser estendida para outras aplicações. Adalgeirsson e Breazeal (2010) apresentaram uma plataforma robótica, denominada MeBot, para presença remota imersiva, em que um *smartphone*, acoplado a um pequeno robô móvel, executava uma aplicação para teleconferências. O robô produzia alguns movimentos condizentes com os gestos do interlocutor remoto, representando uma espécie de presença física. Com isso, esperava-se produzir interações mais agradáveis e cativantes entre as partes.

Nos experimentos realizados, um total de 42 participantes avaliaram duas condições, em uma abordagem *entre sujeitos*. Na primeira condição, o corpo do robô expressava movimentos relacionados às ações do interlocutor. Na segunda condição (i.e., a condição controle), o robô permanecia estático durante toda a interação. Os resultados mostraram que as pessoas se sentiram

mais envolvidas quando o robô atuou de maneira expressiva. Portanto, tecnologias de telepresença poderiam obter benefícios ao permitir que os usuários expressem comportamentos não-verbais, indo além do áudio e do vídeo.

Alguns robôs com suporte para *smartphones* foram desenvolvidos pelo Grupo de de Integração de Sistemas e Dispositivos Inteligentes (GISDI), da Universidade Estadual Paulista (UNESP), do qual fiz parte durante a graduação. Em um trabalho de iniciação científica, detalhado em Ranieri *et al.* (2014), desenvolvi o *software* de um ambiente interativo de robótica, em que o robô Roburguer, mostrado na Figura 16, era controlado por um *smartphone* embarcado, com o qual se comunicava por meio de um enlace Bluetooth. Informações provenientes da câmera e do microfone, bem como de alguns sensores de proximidade embarcados no robô, eram usadas como informação sensorial de uma arquitetura de controle baseada em comportamentos, construída após o desenvolvimento de uma biblioteca orientada a objetos específica para essa finalidade.

Figura 16 – Robô Roburguer, com *smartphone* embarcado.

Os desdobramentos que se seguiram resultaram na construção de um módulo para representação de emoções. Alves e Ferasoli (2016) apresentaram um ambiente em que o robô Pomodoro, uma versão mais nova e robusta do Roburguer, foi provido de uma arquitetura de controle inteligente, baseada em sistemas cognitivos. O modelo de emoções de Russell *et al.* (2014) foi usado para representação interna do estado emocional do robô. As variáveis correspondentes seriam, no decorrer das interações, modificadas gradualmente pelo contexto da interação, e as emoções resultantes, exprimidas por meio de padrões comportamentais e expressões faciais na tela do *smartphone*. O sistema foi desenvolvido para aplicações terapêuticas com crianças autistas. Os experimentos conduzidos foram de natureza técnica, visando validar os diferentes aspectos da arquitetura de controle desenvolvida.

3.2 Reconhecedor de Emoções Adotado neste Trabalho

Apresentados os trabalhos mais relevantes para esta dissertação, resta descrever o sistema para reconhecimento de emoções por análise de expressões faciais a ser utilizado. Como já

mencionado, trata-se de um sistema completamente desenvolvido como parte de um projeto anterior (LIBRALON, 2014), por um ex-membro do Laboratório de Aprendizado de Robôs (LAR) do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC/USP).

O sistema cedido para este trabalho foi apresentado em Libralon e Romero (2013). A aplicação estava agregada a uma interface gráfica para ambiente Windows. Assim, para que esse sistema pudesse executar em tempo real, seria necessário remover essa interface e fazer as adaptações necessárias, Além disso, embora o reconhecedor de emoções reconheça sete emoções, o sistema desenvolvido como parte desta dissertação faz um mapeamento para se trabalhar com somente três estados emocionais: positivo, negativo ou neutro. O modo como essas adaptações foram realizadas não é descrita nesta seção, que lida com o reconhecedor de emoções tal como foi concebido. O processamento da saída do reconhecedor consta na seção 4.4.

A biblioteca FaceTracker (SARAGIH; LUCEY; COHN, 2011), criada para rastreamento de faces, foi incorporada ao sistema. Essa biblioteca, de código-fonte aberto em linguagem `C++`, está disponível para uso não comercial. Também o simulador Weka, que encerra diversas técnicas de aprendizado de máquina e roda em uma máquina virtual `Java`, foi aplicado pelo sistema. As bases de dados utilizadas para treinamento e teste das técnicas de classificação foram a *Radboud Faces Database* (RaFD) (LANGNER *et al.*, 2010) e a *Extended Cohn-Kanade* (CK+) (LUCEY *et al.*, 2010). Ambas as bases eram de acesso livre, contendo expressões faciais relativas a diferentes emoções, para gêneros masculino e feminino e para pessoas de diferentes etnias.

A Figura 17 ilustra os procedimentos desempenhados pelo reconhecedor de emoções, bem como as bibliotecas e serviços externos a ele agregados em cada etapa. O sistema tomava como entrada uma imagem, presumivelmente contendo uma face frontal, e retornava, como saída, uma das seis emoções de Ekman (1999) ou a emoção neutra. Todo o processo descrito na figura, e por conseguinte em toda esta seção, foi implementado como parte do trabalho apresentado em Libralon e Romero (2013).

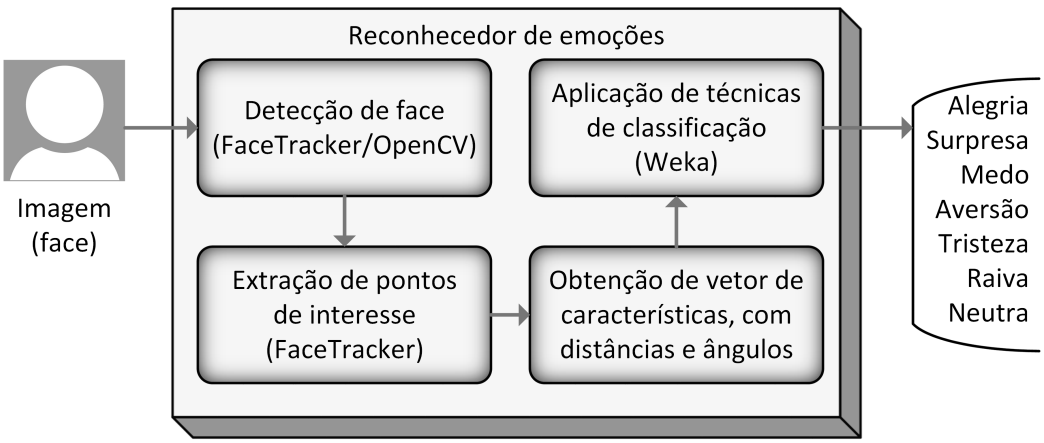
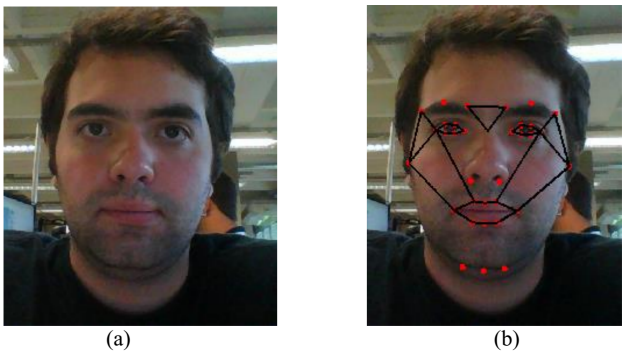


Figura 17 – Etapas desempenhadas pelo reconhecedor de emoções adotado.

A imagem, tomada como entrada do sistema, era submetida ao algoritmo em cascata de Viola e Jones (2001), para detecção de face. Em seguida, pontos de interesse eram obtidos com uso do FaceTracker. Por padrão, 66 pontos poderiam ser extraídos, porém considerou-se um subconjunto de 33 desses pontos, selecionados de acordo com evidências psicológicas. A etapa de extração de características consistia em obter um modelo geométrico a partir desse subconjunto. Isso seria feito calculando distâncias e ângulos entre as coordenadas, conforme ilustrado na Figura 18, e armazenando os resultados em vetores de características geométricas. Cada vetor de características poderia ser, finalmente, classificado por meio de uma técnica de aprendizado de máquina.



(a) (b)

Figura 18 – Extração de características, com uso do *FaceTracker*. (a) imagem original; (b) conjunto de 33 pontos extraídos e relações de interesse para o reconhecedor de emoções.

Foram considerados seis vetores de características, denominados FS_1, FS_2, FS_3, FS_4, FS_{1-3} e FS_{2-4}. O vetor FS_1 era obtido calculando todas as distâncias entre os 33 pontos de interesse e os ângulos que cada linha conectando dois pontos distintos faz com a horizontal. O vetor FS_2 era obtido por um procedimento semelhante, porém trabalha com um subconjunto dessas distâncias e ângulos, a fim de reduzir a quantidade de atributos e melhorar a capacidade de generalização dos algoritmos de classificação. Os vetores FS_3 e FS_4 eram obtidos subtraindo-se os vetores FS_1 e FS_2 dos vetores correspondentes na imagem da expressão facial neutra. Finalmente, FS_{1-3} e FS_{2-4} correspondiam a combinações dos vetores FS_1 com FS_3 e FS_2 com FS_4.

Transformações, tais como rotação, translação e normalizações, precediam a aplicação das técnicas de aprendizado de máquina, na etapa de classificação. Três técnicas foram avaliadas: *Support Vector Machines* (SVM), *Multi-Layer Perceptron* (MLP) e o algoritmo C4.5. Como já mencionado, essa etapa foi implementada com aplicação do simulador Weka.

Nos experimentos apresentados em Libralon e Romero (2013), os modelos treinados foram avalidados com uso de um conjunto de testes, o qual continha 67 imagens da base de dados RaFD, já citada, representativa tanto de crianças como de adultos. Com a técnica SVM, obteve-se 80,22% e 75,82% de acurácia, respectivamente, para os vetores de características FS_1 e FS_2. Com MLP, esses resultados foram 81,02% e 80,21%, e, com o algoritmo C4.5, 71,43% com ambos os vetores. Em todos os casos, os outros vetores de características produziram resultados

melhores, chegando a 94,51%, para FS_{1-3} com SVM. Entretanto, a elaboração desses outros vetores iria requerer uma imagem do sujeito com expressão facial neutra para referência, o que não se tem de antemão em aplicações em tempo real. Por esse motivo, foram considerados somente FS_1 e FS_2.

3.3 Considerações Finais

O problema do reconhecimento de emoções tem recebido uma série de desenvolvimentos e aperfeiçoamentos ao longo da última década. Com isso, algumas aplicações têm sido propostas, em campos que vão dos sistemas de recomendação (TKALCIC; KOSIR; TASIC, 2011) a sistemas interativos de diversas naturezas, sobretudo robôs sociais (SPAULDING; BREAZEAL, 2015). Explorar emoções e empatia tem sido uma diretriz importante dos trabalhos recentes em robótica social e agentes virtuais (LEITE; MARTINHO; PAIVA, 2013).

Foram apresentados estudos empíricos que investigaram as reações de humanos diante de robôs, em dadas circunstâncias. Esses estudos contribuíram para este trabalho na medida em que forneceram arcabouço experimental de que existe uma tendência de humanos entenderem seres artificiais como entidades sociais capazes de possuir emoções, diante das quais podem, inclusive, apresentar empatia.

O trabalho de Gácsi *et al.* (2016), em que o robô expressa emoções reproduzindo movimentos de um cão, mostrou que humanos podem atribuir emoções a agentes robóticos e reconhece-las. O sentimento de empatia foi explorado em Putten *et al.* (2013) e Darling, Nandy e Breazeal (2015), mostrando a medida em que seres humanos são capazes de empatizar com agentes robóticos. Jo *et al.* (2013), por outro lado, mostraram que um robô pode ser entendido como companhia real. Todos esses experimentos foram realizados com robôs com corpo físico, o que, conforme atestado por Wainer *et al.* (2006), é uma vantagem diante de agentes virtuais, como o desenvolvido como parte deste trabalho. Ainda assim, agentes virtuais podem ser importantes para realizar experimentos acerca da interface social, inerente a robôs sociais, como alguns desdobramentos tecnológicos têm atestado (KIRBY; FORLIZZI; SIMMONS, 2010), ou mesmo adaptar esses agentes para integra-los a robôs reais (RANIERI *et al.*, 2014).

Em seguida, discutiram-se desenvolvimentos de técnicas e tecnologias para explorar a dimensão afetiva em interações humano-robô. O robô Kismet (BREAZEAL, 2003), capaz de expressar emoções por meio de uma variedade de graus de liberdade na face, foi uma contribuição notável para este campo. Seguiram-se diversos trabalhos com finalidade semelhante, dentre os quais se pode citar Arnold (2016).

Sintetizar emoções foi um conceito apenas marginalmente abordado no trabalho desta dissertação. O mais importante, no contexto, foi identificar e reagir às emoções do interlocutor humano, como foi realizado em Leite *et al.* (2012), um trabalho que inspirou, em grande medida, os experimentos conduzidos neste projeto. Difere deste trabalho porque, além de se tratar de

um robô físico, usa os estados emocionais reconhecidos para selecionar os comportamentos exibidos, mas não a forma de executar esses comportamentos. A contribuição desta dissertação, nesse sentido, foi produzir pistas verbais e não verbais para reagir diferentemente a cada estado emocional observado.

Seguiu-se a apresentação de alguns agentes virtuais, abordagem adotada neste trabalho. Gockley *et al.* (2005) e Kirby, Forlizzi e Simmons (2010) foram trabalhos notáveis desenvolvidos com a recepcionista virtual Valerie. No mesmo contexto, Trovato *et al.* (2015) desenvolveram a recepcionista virtual Ana, com traços físicos característicos de recepcionistas brasileiras. Um agente virtual para *smartphones* foi apresentado em Yoon e Cho (2012), conforme discutido. O agente em questão não buscava reconhecer o estado emocional dos usuários, mas sim usar informações de contexto para sintetiza-las. Nesse sentido, diferiu da proposta desta dissertação.

Também foram apresentadas algumas possibilidades de se integrar *smartphones* a robôs, o que poderá ser uma futura extensão do agente virtual desenvolvido, com adaptações pontuais. Foram apresentados, então, Adalgeirsson e Breazeal (2010), Ranieri *et al.* (2014) e Alves e Ferasoli (2016). Finalmente, foi detalhado o sistema para reconhecimento de emoções adotado neste trabalho (LIBRALON; ROMERO, 2013).

No próximo capítulo, será apresentada a proposta que motivou o desenvolvimento desta dissertação. Serão detalhadas diversas considerações sobre o projeto e implementação do ambiente.

CAPÍTULO

4

DESENVOLVIMENTO DO PROJETO

Neste capítulo, serão relatadas as atividades de projeto e desenvolvimento realizadas neste trabalho. Será especificado o sistema proposto, detalhando as diretrizes de seu desenvolvimento e o comportamento dele esperado. Informações técnicas sobre o sistema operacional Android, importantes para compreensão da implementação do ambiente, precederão a apresentação da arquitetura de software, em camadas, a qual mostra a organização dos diferentes componentes de software do ambiente e suas relações com os sensores e atuadores do *smartphone*. Em seguida, será detalhada a arquitetura cliente-servidor usada para que o aplicativo do *smartphone* troque mensagens com o reconhecedor de emoções, um executável compilado para Windows. Por fim, será apresentado o desenvolvimento dos principais componentes do aplicativo.

4.1 Visão Global do Ambiente

O ambiente em desenvolvimento é uma personagem virtual interativa para *smartphones*, cujo padrão comportamental, ou *estratégia de interação*, adapta-se ao estado emocional inferido das expressões faciais do usuário. Como será detalhado adiante, o estado emocional resulta de um mapeamento, a partir dos resultados do reconhecedor de emoções, para três possíveis estados emocionais: positivo, negativo ou neutro. Isso foi feito para tornar o ambiente menos suscetível a erros de classificação, bem como para resuzir a quantidade de variáveis com que se lidar ao desenvolver a estratégia de interação.

A estratégia de interação é caracterizada pela alternância entre dois paradigmas independentes, chamados *normal* e *amigável*. Foram introduzidos apenas dois paradigmas por ser esse o cenário mais simples possível, tornando mais factível a realização de experimentos. A comunicação entre o humano e uma personagem virtual, desenvolvida neste projeto e chamada arbitrariamente de Claire, ocorre preferencialmente por fala, usando os serviços de síntese e reconhecimento de fala do Android. A representação visual da personagem consta na Figura 19. Para que a animação fique mais crível, a personagem pisca intermitentemente em todas as

situações.

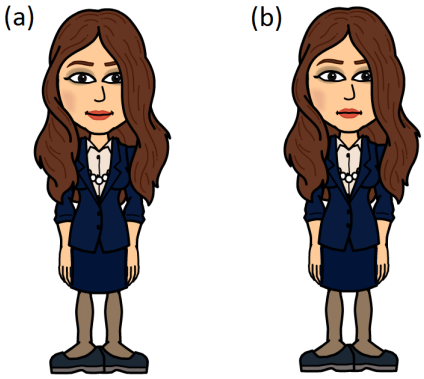


Figura 19 – Personagem virtual: (a) paradigma *amigável*; (b) paradigma *normal*.

Além de um módulo *reconhecedor de emoções*, apresentado em Libralon e Romero (2013) e descrito na seção 3.2, o aplicativo contém outros dois módulos, desenvolvidos como parte deste projeto de mestrado: o *motor de interação*, e o *motor de conteúdo*. A Figura 20 ilustra, em alto nível de abstração, as relações desses três módulos entre si e com o usuário.

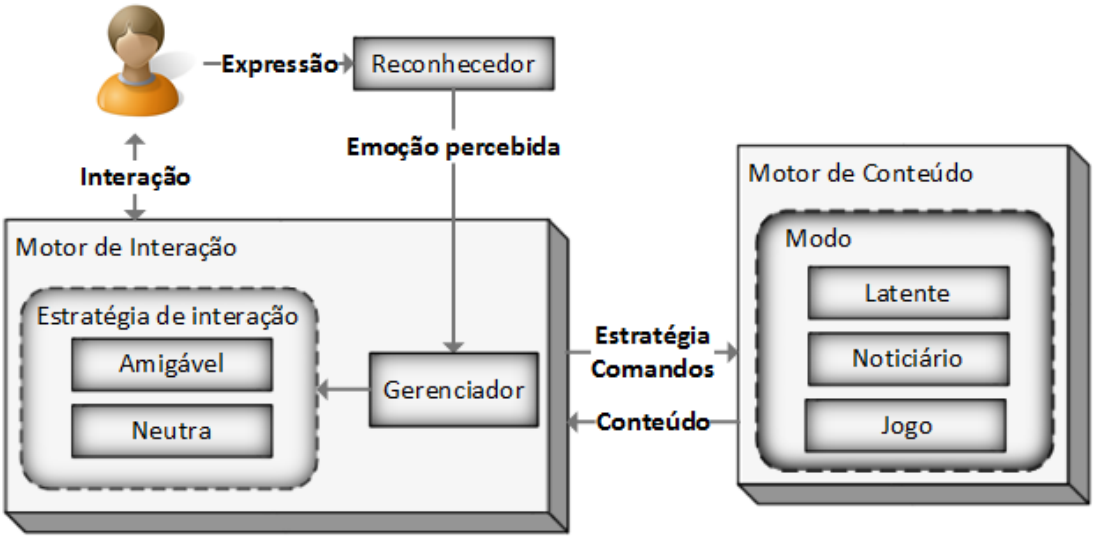


Figura 20 – Arquitetura do sistema proposto, composto por um classificador de emoções, um motor de interação e um motor de conteúdo.

Imagens da face do usuário, capturadas em tempo real, são usadas como entradas para o reconhecedor de emoções. Como resultado, é obtida a emoção percebida pelo sistema, que corresponde a uma das seis emoções básicas categorizadas por Ekman (1999) ou à emoção neutra. A saída desse procedimento de classificação é enviada ao motor de interação.

O motor de interação agrega o sintetizador e o reconhecedor de fala, além da representação visual da personagem. Cabe a ele selecionar a estratégia de interação, conforme o gerenciador da figura arbitra sobre as emoções inferidas, e aplica-la de forma apropriada. O padrão comportamental a ser exibido pela personagem, segundo cada uma dessas duas estratégias, é descrito a seguir:

- **Amigável:** o ambiente é mais proativo do que na estratégia normal. Além disso, exibe-se uma linguagem mais pessoal e expansiva, usando pronomes e verbos em primeira pessoa, em alguns momentos, e manifestando intencionalidade. Por exemplo: “eu ficaria muito feliz em ajudar, caso esteja precisando. Posso lhe mostrar os comandos disponíveis? ”.
- **Normal:** o ambiente é menos proativo do que na estratégia amigável e usa linguagem mais objetiva e desprovida de traços acentuados de personalidade. Por exemplo: “gostaria de verificar os comandos disponíveis? ”.

O motor de conteúdo, responsável por selecionar os comportamentos ou conteúdos propriamente ditos, conhece qual é a estratégia do motor de interação. Além disso, cabe a ele receber e processar os comandos e respostas do usuário, determinando e retornando o conteúdo apropriado. Esse motor pode operar em três *modos* distintos. Cada modo é a implementação de uma funcionalidade específica do sistema, sendo caracterizado pelos seus próprios comportamentos de suporte à interação. Esses modos, denominados *latente*, *noticiário* e *jogo*, têm suas transições representadas na Figura 21. Seguem suas descrições:

- **Latente:** o ambiente permanece sem ação, podendo, eventualmente, ser proativo e sugerir alguma atividade. Caso o usuário aceite uma sugestão ou dê um comando explícito, o modo do sistema é alterado para um dos outros, de acordo com a opção escolhida.
- **Noticiário:** as últimas notícias são obtidas de um portal online e selecionadas para o usuário, de acordo com a categoria. O sistema deve aprender quais categorias de notícias o usuário tem preferência em cada contexto, de modo que o contexto é dado pela estratégia de interação vigente.
- **Jogo:** um jogo simples, baseado em palavras, é iniciado. Esse jogo não apresentará interface gráfica própria, fazendo uso do mecanismo de diálogo para se desempenhar. De acordo com a estratégia de interação, a personagem pode dizer comentários motivadores, elogiar boas jogadas, aumentar o nível de dificuldade ou mesmo auxiliar o jogador, sugerindo-lhe uma jogada.

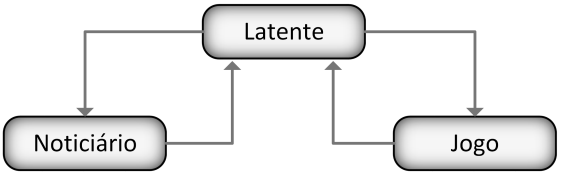


Figura 21 – Transições entre os modos segundo os quais o ambiente pode operar.

O noticiário é obtido do portal da Folha de São Paulo, que disponibiliza as suas reportagens, divididas em categorias, em formato RSS. Quanto ao jogo, foi implementado o clássico Doublets, criado por Lewis Carroll (CARROLL, 1879). Esse jogo consiste em, dadas duas

palavras com o mesmo número de letras, partir de uma delas e chegar à outra alterando uma letra de cada vez, de modo que, após cada alteração, tenha-se uma palavra existente no dicionário. Mais detalhes serão apresentados na seção 4.6.

4.2 Sistema Operacional Android

Como o aplicativo proposto foi desenvolvido para sistema operacional Android, as considerações a respeito de sua implementação não podem ser realizadas antes de uma breve apresentação de aspectos técnicos particulares dessa plataforma. Os conceitos apresentados nesta seção podem ser encontrados na documentação oficial das ferramentas para desenvolvimento Android[1].

Em termos de programação, um aplicativo para Android é dividido em unidades denominadas *activities*. Cada *activity* caracteriza-se por uma tela do aplicativo, tendo um contexto a ela atribuído. Uma *activity* pode invocar métodos que iniciam outra *activity* ou algum serviço disponibilizado no *smartphone*, retornando ou não um resultado quando estes terminam de executar.

No que concerne à implementação, as telas de um aplicativo são compreendidas por uma classe descendente de `Activity`, responsável por responder a eventos, e por um documento `xml` a ela associado, responsável por indicar os componentes, o layout e as referências para alguns eventos da *activity*. O aplicativo também deve levar em consideração os eventos que fazem parte do ciclo de vida da *activity*, ilustrado na máquina de estados da Figura 22. Os eventos associados a cada método um dos métodos representados pelas transições dessa máquina de estados são descritos a seguir:

- `onCreate()`: invocado quando a *activity* é criada pela primeira vez.
- `onRestart()`: invocado após a *activity* ter parado, antes de ser iniciada novamente.
- `onStart()`: invocado quando a *activity* está se tornando visível para o usuário.
- `onResume()`: invocado quando a *activity* vai começar a interagir com o usuário.
- `onPause()`: invocado quando o sistema vai continuar a execução uma *activity* anterior.
- `onStop()`: invocado quando a *activity* não está mais visível para o usuário, porque outra *activity* continuou a executar e encobriu o seu lugar.
- `onDestroy()`: a última chamada antes de a *activity* ser destruída.

[1] https://developer.android.com/index.html, acesso em 30/05/2016.

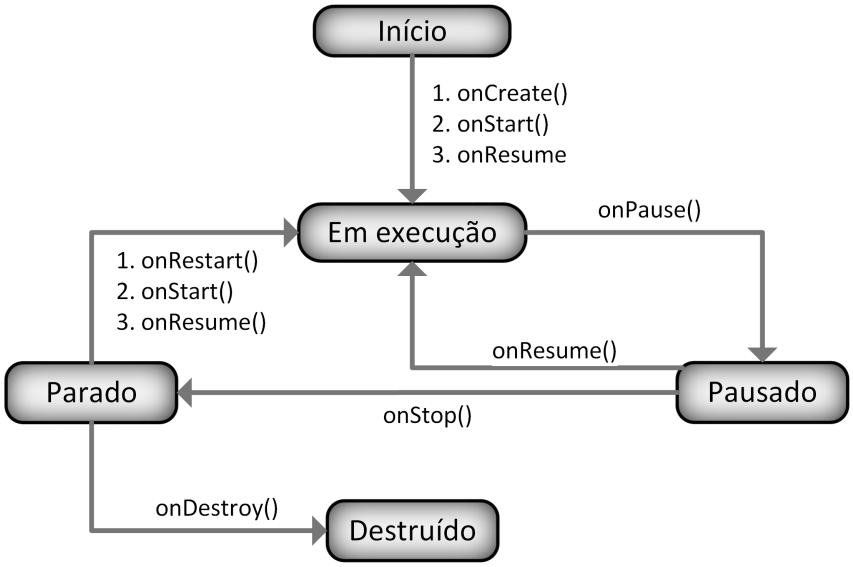


Figura 22 – Ciclo de vida de uma *activity*.

No aplicativo desenvolvido neste trabalho, foram incluídos alguns serviços e bibliotecas de *software*. Para visão computacional, foi incorporado o `OpenCv4Android`. Essa biblioteca, desenvolvida em linguagem `C++`, oferece uma grande variedade de métodos para processamento e análise de imagens, tendo sido exaustivamente utilizada em diferentes projetos que demandam recursos de visão computacional. A interface para Android faz uso de chamadas de função em linguagem `Java`, além de tratar certas particularidades do ambiente Android, como os ciclos de vida das *activities*. A forma mais prática de usar a biblioteca `OpenCv4Android` é instalar o aplicativo *OpenCv Manager*, que contém as funções a serem chamadas dinamicamente pela aplicação. Esse aplicativo está disponível na loja de aplicativos oficial do Google.

O `OpenCv4Android` captura quadros usando as câmeras traseira ou frontal do *smartphone*, exibe esses quadros no *display* e executa um método, definido pelo programador, toda vez que um quadro é capturado. Mais tecnicamente, para definir esse método, o programador deve implementar a interface `CvCameraViewListener2`.

Além de uma biblioteca para visão computacional, serviços para síntese e reconhecimento de fala são de interesse fundamental para este trabalho. Felizmente, as versões mais recentes do Android já disponibilizam, para uso *offline*, serviços para síntese e reconhecimento de fala, que foram aplicados ao presente trabalho.

4.3 Arquitetura de Software

Em nível de implementação, o ambiente desenvolvido é distribuído entre duas plataformas: um *smartphone* e uma base computacional. O *smartphone* é responsável tanto pela interação com o usuário, por meio de seus sensores e atuadores, quanto por quase todas as funcionalidades do aplicativo. As exceções são o reconhecimento de emoções e a obtenção de

notícias na Internet, que ficam a cargo da base computacional, na qual foi implementado um programa servidor.

Ainda que, em princípio, não existam restrições técnicas para implementar essas duas funcionalidades diretamente no *smartphone*, optou-se por essa forma distribuída para evitar a necessidade de reimplementar o reconhecedor de emoções, originalmente desenvolvido para sistema operacional Windows. A implementação do paradigma cliente-servidor, planejada especialmente para o reconhecedor de emoções, foi aproveitada para obter as notícias através da base computacional, visto que a linguagem `Python` dispõe de recursos bastante práticos para obter e processar notícias em formato `RSS`.

A arquitetura de software do ambiente é representada na Figura 23, que ilustra a organização dos componentes dentro das duas plataformas e a comunicação entre elas. Essa comunicação ocorre por uma conexão de rede, omitida da representação. O *smartphone* e a base computacional devem conectar-se à mesma rede local. O *smartphone* pode fazê-lo por meio de sua interface Wi-Fi, e a base computacional, por meio de qualquer tipo de interface. A nuvem representa uma conexão à Internet, por meio da qual as notícias são obtidas.

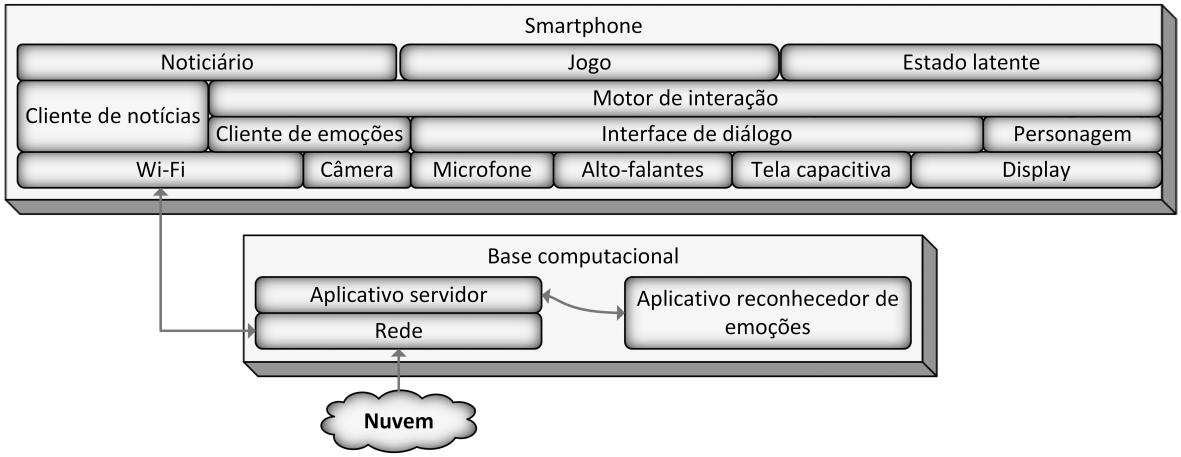


Figura 23 – Arquitetura de software do ambiente, contemplando tanto o smartphone quanto a base computacional. A conexão está representada como uma seta bidirecional entre as interfaces de ambos os dispositivos, ignorando os dispositivos e protocolos que se possam interpor.

Na base computacional, dois programas são operados: o programa servidor, codificado em linguagem `Python`, e o aplicativo reconhecedor de emoções, apresentado na seção 3.2. O programa servidor, que executa indefinidamente, é incumbido de receber e responder a solicitações, através de trocas de mensagens. Quando solicitado, portanto, cabe a ele criar um subprocesso para executar o reconhecedor de emoções, fazendo-o de modo síncrono (i.e., bloqueante) e retornando as emoções reconhecidas.

O restante desta seção será concentrado em descrever os módulos da arquitetura de *software* do aplicativo para *smartphone*, desenvolvido em linguagem Java, com uso do *Software Development Kit* (SDK) para Android. Essa arquitetura é ilustrada na Figura 23. Os sensores de que a aplicação faz uso são a câmera frontal, o microfone e a tela capacitiva, enquanto os

atuadores são os alto-falantes e o *display*. O primeiro nível da arquitetura é compreendido pelas interfaces para acesso a esses componentes. No segundo nível, têm-se o cliente de notícias, o cliente de emoções, a interface de diálogo e a representação visual da personagem. No terceiro nível, tem-se o motor de interação. No quarto e último nível, tem-se o motor de conteúdo, em que cada um dos modos de operação do ambiente é definido.

O cliente de notícias tem finalidade de, quando requisitado, usar a interface Wi-Fi para atualizar as notícias, armazenadas em uma estrutura de dados no *smartphone*, e tornar essas informações disponíveis diretamente para o noticiário. O cliente de emoções, por sua vez, obtém imagens da câmera frontal a cada segundo, com uso da biblioteca `OpenCv4Android`, e comunica-se com a base computacional de modo a obter a classificação da emoção do usuário. Essa classificação é disponibilizada para o motor de interação.

A interface de diálogo agrega o reconhecimento e a síntese de fala, disponibilizados pelo Android SDK, adequando essas funcionalidades para uso pelo motor de interação. Um botão na interface visual, acessado pela tela sensível ao toque, pode ser usado para alternar entre os estados *escutando*, em que informação proveniente do microfone é obtida para análise, *falando*, em que informação proveniente do motor de interação é traduzida em áudio para o alto-falante, ou *passivo*, em que a interface de diálogo está ociosa. Além de integrar essas funcionalidades, cabe à interface de diálogo fazer com que a síntese e o reconhecimento de fala ocorram de modo síncrono, interrompendo a aplicação até que cada uma dessas tarefas seja concluída. Esse mecanismo será discutido com detalhe na seção 4.5. Uma interface por texto, caracterizada por um botão que aciona uma caixa de diálogo, foi incluída para prevenir situações em que a interação esteja sendo prejudicada por problemas com o reconhecedor de fala.

O *display* é usado para explicitar, em uma legenda, o que é dito pelo sintetizador de fala, assegurando a compreensão, pelo usuário, de tais sentenças. Além disso, é nele que a representação visual da personagem é mostrada. Essa representação evidencia qual é a estratégia de interação em curso, por meio da animação selecionada (ver Figura 19).

Conforme já citado na seção 4.1, o motor de conteúdo agrega os três modos de operação definidos para o sistema: o estado latente, o noticiário e o jogo. Cada um desses modos e suas estratégias de aprendizado serão discutidos na seção 4.6.

4.4 Adaptação do Reconhecedor de Emoções

Quando do início da concepção deste projeto, o reconhecedor de emoções era um programa para Windows dotado de interface gráfica. Essa interface permitia obter uma imagem da *webcam* ou de um arquivo, selecionar conjuntos de características para análise e obter os resultados em componentes gráficos na tela. Entretanto, faz-se necessário que o reconhecimento de emoções ocorra em segundo plano e de forma iterativa, devendo, portanto, ser executado de forma automática e independente de intervenção do usuário.

Para isso, a interface gráfica foi removida, e argumentos para executar do programa foram introduzidos. O objetivo foi transforma-lo em um programa que pudesse ser executado como subprocesso de outra aplicação. Os argumentos são os conjuntos de características a serem usados (ver seção 3.2), o caminho da imagem de entrada e o caminho do arquivo de saída, que deverá conter os resultados da classificação. A Tabela 1 mostra os argumentos para execução do programa.

Tabela 2 – Argumentos para execução do reconhecedor de emoções.

Argumento	Descrição
-i <imagem >	Nome do arquivo da imagem de entrada
-o <saída >	Nome do arquivo de saída, com os resultados
-fs1	*Flag* para o conjunto de características FS_1
-fs2	*Flag* para o conjunto de características FS_2

Realizada essa adaptação, foi necessário criar a interface para realizar a chamada a esse executável, dentro da base computacional Windows, por meio de um processo que se comunicasse com o aplicativo em execução no *smartphone*. Para isso, foi criado um programa em `Python`, cuja finalidade é atuar como servidor nessa comunicação. Esse programa permanece em um laço, no qual aguarda uma conexão na porta 9999, selecionada arbitrariamente. Recebida a conexão, o procedimento para obter a imagem, reconhecer a emoção e retornar o resultado é o ilustrado no diagrama de sequência da Figura 24.

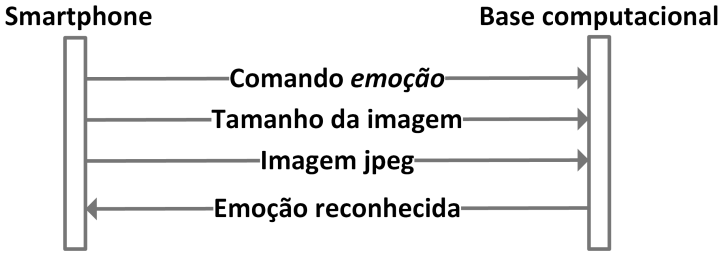


Figura 24 – Sequência de envios e recebimentos para que o aplicativo para smartphone (cliente) possa usar a base computacional (servidor) para obter a classificação da expressão facial obtida pela câmera frontal.

Primeiro, cabe ao cliente (i.e., ao aplicativo para *smartphone*) enviar um comando. Essa etapa é necessária para informar ao servidor qual das suas funcionalidades está sendo requisitada, entre reconhecer emoções, obter notícias ou qualquer outra funcionalidade que se pretenda incluir. Enviado o comando `emoção`, o cliente deve enviar o tamanho da imagem, em *bytes*, seguido do próprio arquivo contendo a imagem, em formato `jpeg`. O servidor processa a requisição, por meio de uma chamada ao programa executável para reconhecimento de emoções, e envia o resultado para o cliente.

Por serem os únicos vetores de características viáveis para utilização em tempo real por imagens obtidas de uma câmera, dentre os vetores apresentados na seção 3.2, somente FS_1 e FS_2 foram analisados. Essa decisão foi tomada porque todos os outros vetores não podem ser elaborados sem a referência de uma imagem da face do indivíduo com expressão neutra, obtida

em condições semelhantes à da expressão em análise. Em aplicações práticas, não é possível obter tal imagem.

Para gerar o resultado final, foram comparados os resultados dos classificadores para cada vetor de características. Além disso, considerou-se somente a polaridade da emoção em análise. Em outras palavras, emoções com valência negativa (medo, aversão, raiva e desgosto) foram classificadas similarmente como negativas, ao passo que emoções com valência neutra (surpresa e neutro) foram classificadas como neutras, e a única emoção com valência positiva do modelo (alegria) foi classificada como positiva. Essa abordagem, justificada na Figura 3, é útil para eliminar algumas ambiguidades presentes no reconhecedor de emoções, além de tornar mais prática a análise do estado emocional do usuário.

Dessa forma, há três respostas possíveis para serem enviadas do servidor para o cliente: negativa, neutra ou positiva. Portanto, não se trabalhou com cada uma das emoções possíveis segundo o modelo de Ekman (1999), mas somente com as polaridades dessas emoções.

4.5 Projeto e Desenvolvimento do Motor de Interação

A interface do aplicativo, na qual foram incorporadas as funcionalidades do motor de interação, está representada na Figura 25. Na região superior, em fundo preto, situa-se a legenda, em que as falas sintetizadas pela personagem são reproduzidas, conforme mencionado na seção 4.1. O tamanho da legenda foi superdimensionado na representação, a fim de favorecer a visualização do texto. Na região central da tela, com fundo branco, está posicionada a personagem. Três botões podem ser visualizados abaixo dela: um botão com ícone de engrenagem, um botão cujo ícone alterna entre microfone (como na imagem) e alto-falante e um botão com ícone de ponta de caneta. Do lado direito da parte superior da personagem, há um cronômetro, aplicado em uma das versões do jogo descrito na subseção 4.6.2, e uma luz, que pode ser vermelha (como na imagem) ou verde.

A luz ao lado da personagem, que na figura é vermelha, indica se a face do usuário está sendo detectada pelo reconhecedor de emoções. Se essa luz estiver verde, a face está sendo detectada. Se estiver vermelha, o usuário deve reposicionar o dispositivo em relação a seu rosto, a fim de que sua face possa ser detectada. Esse mecanismo foi concebido como uma maneira prática de informar ao usuário se sua face está corretamente posicionada em relação ao *smartphone*.

O botão com ícone de engrenagem refere-se às configurações do aplicativo. No caso, a única configuração modificável é o endereço `IP` (*Internet Protocol*) da base computacional que contém o programa servidor. Pressionado o botão, uma caixa de diálogo é disponibilizada para permitir alterar a variável correspondente.

Os outros dois botões referem-se à interface de diálogo. O botão com ícone de microfone

ou alto-falante tem finalidade de acionar a interface por fala ou interromper a reprodução do áudio referente a uma comunicação da personagem. Já o botão com ícone de caneta, à direita do microfone, aciona a interface de texto, alternativa à interface por fala. No caso, a entrada é realizada por uma caixa de diálogo.

A seguir, na subseção 4.5.1, a implementação das funcionalidades propostas para o motor de interação, mencionadas na seção 4.1, será discutida. O desenvolvimento foi realizado seguindo o paradigma de orientação a objetos, provendo métodos de alto nível de abstração. O objetivo é que o módulo em questão possa ser expandido e usado em pesquisas futuras, ainda que independentemente da aplicação específica apresentada neste trabalho.

4.5.1 Detalhes de Implementação

Um dos objetivos do motor de interação é prover interfaces de *software* para manipular a interface gráfica. Para entrar em detalhes quanto ao modo como isso é feito, é oportuno apresentar um diagrama de classes simplificado, a fim de dar informações a respeito do ambiente em nível de implementação. Esse diagrama consta na Figura 26.

A classe `MainActivity` é a *activity* cujo *design* foi apresentado na Figura 25. Essa classe, herança da classe nativa `Activity`, implementa os dois primeiros níveis da arquitetura de software ilustrada na Figura 23. A interface de diálogo é caracterizada pelos métodos `listen()` e `speak()`, que integram a legenda da interface e as funcionalidades assíncronas de reconhecimento e síntese de fala do Android, implementando-as de forma síncrona. Em outras palavras, os serviços de reconhecimento e síntese de fala do Android são executados em um fluxo de controle (*thread*) distinto daquele que invocou o método correspondente. Na aplicação deste trabalho, foi necessário sincronizar tal procedimento.

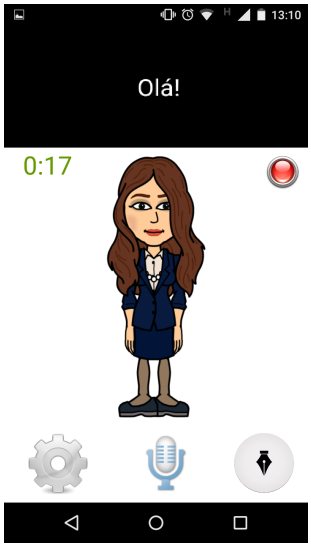


Figura 25 – *Activity* responsável pela interação entre humano e agente no aplicativo para smartphone. Para favorecer a visualização, o tamanho da legenda está superdimensionado nesta imagem.

Esse problema foi resolvido da seguinte maneira: quando os métodos assíncronos de reconhecimento ou síntese de fala são invocados no código, invoca-se o método bloqueante `acquire()` em um semáforo inicializado com valor zero. Com isso, o fluxo de controle que invocou o método é bloqueado. Quando o serviço retorna, em uma função de *callback*, invoca-se o método `release()` do semáforo em questão, desbloqueando tal fluxo de controle. As demais sincronizações realizadas neste trabalho foram implementadas segundo essa mesma técnica.

A classe `MainActivity` também agrega os clientes de emoções e de notícias. Um fluxo de controle dedicado, contido na classe `ClientThread`, troca mensagens periodicamente para obter o reconhecimento das emoções, e procede de modo a receber notícias quando um *flag* correspondente é sinalizado.

A imagem a ser enviada para o reconhecedor de emoções é obtida como parâmetro no método `onCameraFrame(CvCameraViewFrame inputFrame)`, um *callback* invocado dentro de um fluxo de controle do `OpenCv4Android` e que deve ser reimplementado em aplicações que fazem uso dessa biblioteca para obter quadros das câmera frontal ou traseira. Para fazê-lo, é preciso que a *activity* em questão implemente a interface `CvCameraViewListener2`.

Quanto à representação visual da personagem, a `MainActivity` limita-se a implementar as diferentes animações e prover o método público `setCartoon(int newCartoon)`, que permite ao motor de interação alterar a animação conforme o estado emocional inferido. Essa determinação do estado emocional, porém, cabe à classe `InteractionMotor`, responsável pelo terceiro nível da arquitetura de software da Figura 23.

A classe `InteractionMotor` contém um fluxo de controle que monitora os resultados do reconhecedor de emoções e modifica a estratégia de interação de acordo. Essa seleção de paradigma ocorre de forma determinística, considerando os n últimos resultados do reconhecedor. Se mais de $n/2$ desses resultados são emoções com polaridade positiva, a estratégia de interação é configurada para amigável, do contrário, é configurada para normal. Além disso, essa classe

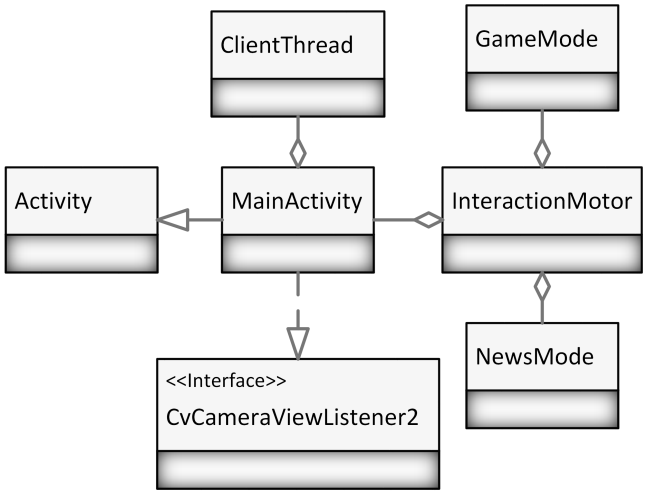


Figura 26 – Diagrama de classes simplificado do sistema. Algumas classes, com relevância restrita, foram omitidas, para facilitar a compreensão da estrutura global do desenvolvimento do aplicativo.

provê um fluxo de controle principal, que seleciona, com certa probabilidade a cada segundo, um comportamento proativo. Essa probabilidade é de $0,07$ na estratégia *amigável*, e de $0,035$ na estratégia *normal*. Os comportamentos proativos são sugerir verificar os comandos disponíveis, sugerir noticiário ou sugerir jogo.

Se o usuário aceita uma sugestão ou seleciona um comando, os modos noticiário ou jogo podem ser ativados. Esses modos fazem parte do motor de conteúdo, apresentado na próxima seção, junto às classes `NewsMode` e `GameMode`.

4.6 Projeto e Desenvolvimento do Motor de Conteúdo

O motor de conteúdo compreende o quarto nível da arquitetura de software do aplicativo. No diagrama de classes da Figura 26, é compreendido pelas classes `GameMode` e `NewsMode`. Uma vez que a interface de diálogo e todas as funcionalidades do motor de interação foram implementados no terceiro nível da arquitetura, a implementação de cada modo é bastante direta, como se poderá verificar nas próximas subseções.

4.6.1 Estado latente

O modo *estado latente* é gerenciado pela própria classe `InteractionMotor`, com seus comportamentos proativos. As sentenças ditas nos modos *amigável* e *normal* são diferentes para cada comportamento proativo. A finalidade deste modo é oferecer interfaces para que o usuário possa acessar algum dos outros modos, o que pode ocorrer através de uma sugestão, fornecida por meio de um comportamento proativo e acatada pelo usuário, ou através de um comando explícito, fornecido por meio da interface de diálogo. Os comandos disponibilizados foram *mostrar comandos*, *jogo* e *noticiário*, responsáveis por, respectivamente, dizer os comandos disponíveis, iniciar o modo *jogo* e iniciar o modo *noticiário*.

4.6.2 Jogo

O modo *jogo* foi implementado na classe `GameMode`. Consiste em um jogo de palavras clássico chamado *Doublets* (CARROLL, 1879), o qual não requer uma interface gráfica dedicada. As regras são as seguintes: dadas duas palavras A e B, iniciar pela palavra A e modifica-la uma letra de cada vez, de modo que cada alteração resulte, sempre, em uma palavra existente no vocabulário do idioma em que se estiver jogando, até se chegar à palavra B. Essa dinâmica é ilustrada na Figura 27, em um exemplo com as palavras *fogo* e *gelo*.

Não se trata de um jogo competitivo, portanto a personagem atua junto ao usuário na condição de mediador. Em todas as rodadas, o mediador enuncia a palavra atual e a palavra meta. As entradas válidas na interface de diálogo são as palavras válidas, modificando a palavra atual. Como parte da estratégia de interação, quatro comportamentos proativos foram introduzidos,

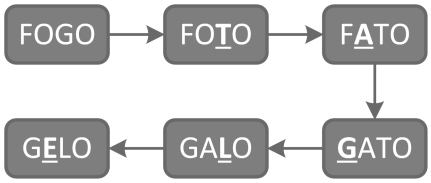


Figura 27 – Exemplo do jogo *Doublets*, com as palavras *fogo* e *gelo*.

sendo eles *proferir comentários encorajadores*, *dar informações gerais*, como quantas palavras o usuário já atingiu, *sugerir uma palavra viável* e *não fazer nada*. Esses comportamentos são caracterizados por sentenças diferentes, de acordo com o paradigma de interação em curso.

Embora a decisão de tomar iniciativa seja tomada com base em uma variável aleatória de distribuição uniforme, um algoritmo de aprendizado por reforço arbitra a seleção do comportamento. O processo de arbitração foi modelado como um *multiarmed bandit problem*, um algoritmo de aprendizado por reforço que consiste em, dado um conjunto finito de estratégias possíveis, chamadas *máquinas de apostas*, escolher a próxima estratégia, de modo a maximizar a recompensa média recebida. A solução aplicada neste trabalho é a do Algoritmo 1, adaptado de Auer, Cesa-Bianchi e Fischer (2002).

Algoritmo 1: Solução adotada para o *multiarmed bandit problem*.

```
1  if NOT todas as estratégias inicializadas then
2      for i = 1 → num_estratgias do
3          estratégia = estratégias[i]
4  else
5      while TRUE do
6          Selecionar estratégia i que maximiza x̄_i + sqrt(2 ln n / n_i)
7          Onde
8          x̄_i → recompensa média obtida com a estratégia i
9          n_i → número de vezes que a estratégia i foi selecionada até o momento
10         n → número total de estratégias selecionadas até o momento.
```

A regra de reforço selecionada foi baseada nos resultados do reconhecedor de emoções. Partiu-se da premissa de que se, logo após a execução de um comportamento, o estado emocional do usuário for alterado para um mais positivo, então aquele comportamento foi, provavelmente, adequado. Para realizar o cômputo, atribuiu-se um valor e a cada tipo de emoção. Para emoções positivas, $e = 1$; para emoções neutras, $e = 0$; para emoções negativas, $e = -1$. Toda vez que um comportamento proativo é executado, somam-se os valores das n últimas emoções reconhecidas antes de sua execução e e subtrai-se esse valor da soma das n emoções inferidas logo depois, normalizando o resultado em seguida. Esse procedimento é descrito pela Equação 4.1. Na

implementação realizada, foi fixado $n = 2$

$$recompensa = \sum_{i=1}^{n} \frac{e_i}{n} - \sum_{i=-n+1}^{0} \frac{e_i}{n} \tag{4.1}$$

Em uma versão desse jogo, aplicada a alguns experimentos do Capítulo 5, foi imposta uma limitação de tempo para completar a partida. Embora essa alteração introduza um elemento adicional de desafio, ela foi introduzida por motivo meramente prático: durante a realização dos experimentos, o tempo disponível era escasso.

4.6.3 Noticiário

Como discutido na seção 4.3, as notícias são obtidas pela base computacional e transferidas ao *smartphone* segundo uma arquitetura cliente-servidor. O cliente de notícias pertence ao segundo nível da arquitetura de *software* da Figura 23 e foi implementado dentro da classe `ClientThread`, interna a `MainActivity`. A forma como as notícias são obtidas do servidor é detalhada no diagrama de sequência da Figura 28.

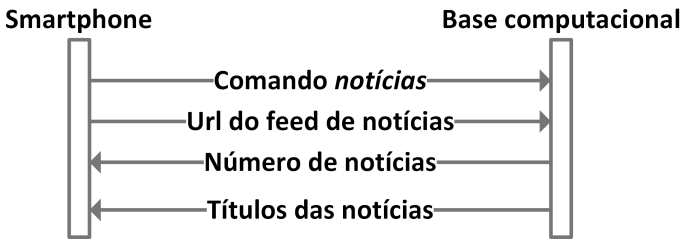


Figura 28 – Sequência de envios e recebimentos para que o aplicativo para *smartphone* (cliente) possa usar a base computacional (servidor) para obter notícias de um *feed* específico.

Segundo o protocolo estabelecido, o aplicativo no *smartphone* requisita o serviço para aquisição de notícias por meio do comando notícias. Enviado tal comando, deve ser enviada a `url` do arquivo `RSS` de notícias a ser obtido. Esse endereço pode determinar não somente o portal de onde essas notícias serão extraídas, mas também a categoria dessas notícias, uma vez que um mesmo portal, em geral, fornece diversos arquivos diferentes.

A base computacional, então, retorna duas mensagens. A primeira é o número de notícias que obteve por meio do arquivo `RSS`. A segunda corresponde a um vetor de cadeias de caracteres com os títulos de cada uma das notícias. O cliente de notícias armazena esses títulos em uma estrutura de dados estática, para que eles possam ser mostrados ao usuário em um momento oportuno, conforme mostrado na Figura 29.

No modo *noticiário*, implementado na classe `NewsMode`, o ambiente exibe a sequência de etapas ilustrada no fluxograma da Figura 29 Esse procedimento se repete n vezes, onde n é determinado como parâmetro na implementação do aplicativo. Todas as etapas ocorrem por meio da interface de diálogo, sem necessidade de uma interface gráfica específica para essa finalidade.

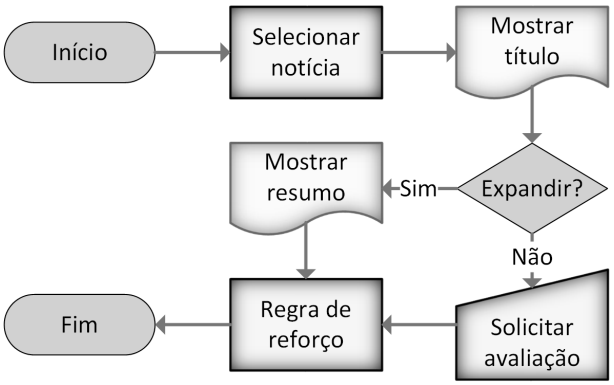


Figura 29 – Fluxograma mostrando a sequência de procedimentos executada para cada exibição de notícia do ambiente no modo *noticiário*.

O sistema seleciona uma notícia, com base no seu gênero e nas preferências aprendidas do usuário. Em seguida, o título da notícia é exibido, com uso da interface de diálogo, e é oferecida ao usuário a opção de expandi-la. Caso essa opção seja acatada, o resumo da notícia é enunciado, também através da interface de diálogo. Do contrário, a personagem solicita que o usuário avalie explicitamente o seu interesse por ela, com uma nota de 1 a 3. De qualquer modo, segue-se a aplicação da regra de reforço e, caso o número de notícias enunciadas até então não tenha superado n, o sistema seleciona uma próxima notícia.

A regra de reforço consiste em usar o *feedback* do usuário para atualizar, em tempo de execução, o modelo usado para recomendação de notícias. Assim como ocorreu com os comportamentos do modo *jogo*, o sistema de recomendação foi modelado como um *multiarmed bandit problem*, em que cada categoria de notícias - *poder*, *esporte*, *tecnologia* e *cultura* - foi mapeada para uma estratégia. A solução adotada, também aplicada na seleção de comportamentos do modo *jogo* consta no Algoritmo 1. O valor da recompensa é computado da seguinte maneira: caso o usuário solicite que a notícia seja expandida, atribui-se uma $recompensa = 1.0$ para a categoria a que ela pertence. Do contrário, o valor da avaliação explícita, de 1 a 3, é normalizado, tal que $recompensa \in [-0.5, 0.5]$.

As notícias, obtidas do portal Folha de São Paulo, são periodicamente atualizadas em uma estrutura de dados, por meio da sucessiva aquisição de arquivos RSS, disponibilizados e mantidos atualizados pelo jornal. As informações desses arquivos, estruturados segundo uma linguagem de marcação `xml`, são processadas com uso do pacote `BeautifulSoup`, disponível para linguagem `Python`. Isso ocorre dentro da base computacional remota, que roda Windows e se comunica com o aplicativo para *smartphone* através de uma conexão de rede, conforme detalhado nas seções anteriores. O sistema de aprendizado, responsável por selecionar a categoria a que deve pertencer a próxima notícia a ser apresentada, foi implementado dentro do aplicativo para *smartphone*. Uma abordagem de recomendação de notícias mais sofisticada, também baseada em soluções para o *multiarmed bandit problem*, foi apresentada em Tang *et al.* (2015).

4.7 Considerações Finais

O sistema proposto consiste em uma personagem virtual dotada de comportamentos adaptáveis, de acordo com uma estratégia de interação baseada em emoções. Essa personagem foi agregada a um aplicativo para *smartphone*, entretanto parte das suas funcionalidades, inclusive o reconhecedor de emoções (LIBRALON; ROMERO, 2013), operam em uma base computacional remota, em ambiente Windows. A comunicação ocorre por meio de uma arquitetura cliente-servidor. A arquitetura de *software* pautou toda a discussão sobre o desenvolvimento do ambiente e suas funcionalidades, que incluem um sistema para recomendação de notícias e um jogo de palavras.

No próximo capítulo, serão apresentados e discutidos os experimentos realizados com o ambiente desenvolvido.

CAPÍTULO

5

EXPERIMENTOS

Neste capítulo, serão apresentados os experimentos realizados para validar o sistema proposto. Dois estudos principais serão detalhados nas próximas seções. No primeiro deles, calouros de Ciência da Computação interagiram com a aplicação por tempo ilimitado. No segundo, alunos de um programa de pré-iniciação científica interagiram por tempo restrito. Outros estudos, inacabados ou de caráter exploratório, serão apresentados brevemente durante a discussão.

Por consistir em enunciados de notícias extraídos diretamente dos portais especializados, a interação no modo *noticiário* limita-se, na maior parte do tempo, a monólogos desprovidos de personalização. Isso torna mais difícil estabelecer diferenças entre os dois paradigmas propostos na estratégia de interação. Uma vez que os conjuntos de participantes disponíveis eram escassos, optou-se por conduzir estudos com usuários somente com o modo *jogo*, que é mais interativo e permite comportamentos proativos em mais situações.

Devido à reduzida quantidade de voluntários, todos os experimentos foram baseados em abordagens *intra-sujeito*, em que todos os participantes experimentam todas as condições. Como controle, foi introduzida uma versão modificada do aplicativo, em que a estratégia de interação exibe somente o paradigma *amigável*. Foram atribuídos nomes arbitrários a cada versão do ambiente: a versão completa, com a estratégia de interação proposta, foi chamada Claire, enquanto a versão controle foi denominada Rachel.

No início da sessão experimental, os participantes foram informados de que o sistema analisaria suas expressões faciais. Foi-lhes assegurada total privacidade quanto às imagens coletadas pela câmera do *smartphone*, eliminadas tão logo cada imagem era analisada, bem como o sigilo quanto às suas identidades.

Antes do início da interação, foram dadas instruções sobre as regras do jogo e sobre a interface do aplicativo, de modo que a curva de aprendizado do usuário sobre esses itens exercesse a menor influência possível nos resultados. Para priorizar a forma mais natural de interação,

foi solicitado aos participantes que preferissem usar a interface por reconhecimento de fala. O mesmo dispositivo, um *smartphone* Motorola Moto G da primeira geração, foi manipulado, exclusivamente, em todas as sessões. Para permitir maior imersão no ambiente, foram usados fones de ouvido, fornecidos de antemão.

Os experimentos foram contrabalanceados, ou seja, a ordem de apresentação das versões do aplicativo era alternada para cada participante. Os modelos dos questionários, aplicados após a experimentação de cada versão do aplicativo por um participante, constam no Apêndice C e no Apêndice D. Na maior parte dos casos, as respostas a esses questionários foram requeridas na forma de avaliações de 1 a 5. Considerações sobre as decisões tomadas para a configuração dos experimentos, levando em conta os resultados obtidos, serão apresentadas na seção 5.4.

A fim de verificar se houve significância estatística nos resultados, por meio da aplicação do teste-t pareado, apresentado na subseção 2.2.1.2, foi necessário assegurar-se de que as amostras podiam ser ajustadas por uma função de distribuição normal. Isso foi realizado com uso do teste Kolmogorov-Smirnov de aderência à normalidade, apresentado na subseção 2.2.1.1.

Atestada a normalidade das amostras, o teste-t pareado foi aplicado sobre cada característica analisada, em cada experimento. Em todos os casos, sendo μ_D a média amostral das diferenças entre a condição com a estratégia de interação implementada e a condição controle, as hipóteses foram definidas como:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases}$$

Como discutido na subseção 2.2.1.2, essa configuração implica na aplicação do teste-t unicaudal, uma vez que se estipulou uma direção, isto é, a diferença deveria ser favorável à estratégia proposta neste trabalho.

5.1 Experimento A

O primeiro experimento realizado ocorreu nos dias 14 e 15 de janeiro de 2016, no Laboratório de Integração de Sistemas e Dispositivos Inteligentes (LISDI) e na sala multi-uso do Laboratório de Extensão e Pesquisa em Computação (LEPEC) da Faculdade de Ciências da Universidade Estadual Paulista (FC-UNESP). Contou com 11 participantes (10 homens, 1 mulher), calouros do curso de Ciência da Computação da universidade, com idade entre 18 e 21 anos.

Como previamente mencionado, foi-lhes solicitado que jogassem com as duas versões, preenchendo questionários após cada partida. Metade dos participantes jogou primeiro com Claire, depois com Rachel, e a outra metade fez o oposto. Não foi estipulado tempo limite para as interações, de modo que cada participante interagiu por aproximadamente 10 minutos com

cada versão. O tempo total de experimentação, para cada participante, foi de aproximadamente 25 minutos.

O questionário aplicado consta no Apêndice C. As questões relacionaram-se a características desejadas atribuídas à personagem, para as quais cada participante deveria atribuir uma avaliação de 1 a 5. Essas características, adaptadas de um subconjunto do questionário apresentado e discutido em Bartneck *et al.* (2009), eram *realista*, *autoconsciente*, *interativo*, *atento*, *amigável*, *gentil*, *agradável*, *competente* e *sensato*. O objetivo foi verificar se, embora a representação da personagem fosse um mesmo desenho caricatural nas duas versões, a alteração de comportamento conferida pela estratégia de interação produziria percepções diferentes. Como última questão, foi-lhes perguntado se, durante a interação, sentiram *empatia* diante da personagem, também em uma escala de 1 a 5.

5.1.1 Resultados

Algumas questões foram desconsideradas na análise, por terem obtido respostas muito semelhantes nas duas versões do aplicativo. Foram consideradas, então, a questão sobre a sensação de empatia e as avaliações quanto às seguintes características: *realista*, *agradável*, *gentil* e *competente*. A aplicação do teste Kolmogorov-Smirnov de aderência à distribuição normal consta na planilha contida no Apêndice A, junto aos dados brutos deste experimento. A Figura 30 mostra os histogramas correspondentes a cada amostra.

O teste Kolmogorov-Smirnov foi bem-sucedido, em todos os casos, para nível de significância $\alpha = 0,05$. A Tabela 3 mostra os valores D_n obtidos para cada caso, e pode-se verificar que todos se situam abaixo do valor crítico para 11 participantes, $D_{n,\alpha} = 0,391$, constante na Tabela 1. Em outras palavras, a distribuição normal é um bom ajuste para as amostras e o teste-t pode ser aplicado para verificar se há significância estatística.

Os valores-p obtidos pela aplicação do teste-t constam na Tabela 4, junto às diferenças das médias entre as condições de teste (Claire) e controle (Rachel) e seus respectivos desvios-padrão. Referindo-se à questão sobre empatia, obteve-se $p = 0,0519$, o que mostra uma forte tendência à significância estatística (por convenção, adota-se nível de significância $\alpha = 0,05$, isto é, diz-se que a diferença entre duas amostras é significante quando $p \leq 0,05$). Por outro lado, quanto às outras características levadas em conta, o teste-t não revelou diferença significativa entre as duas condições.

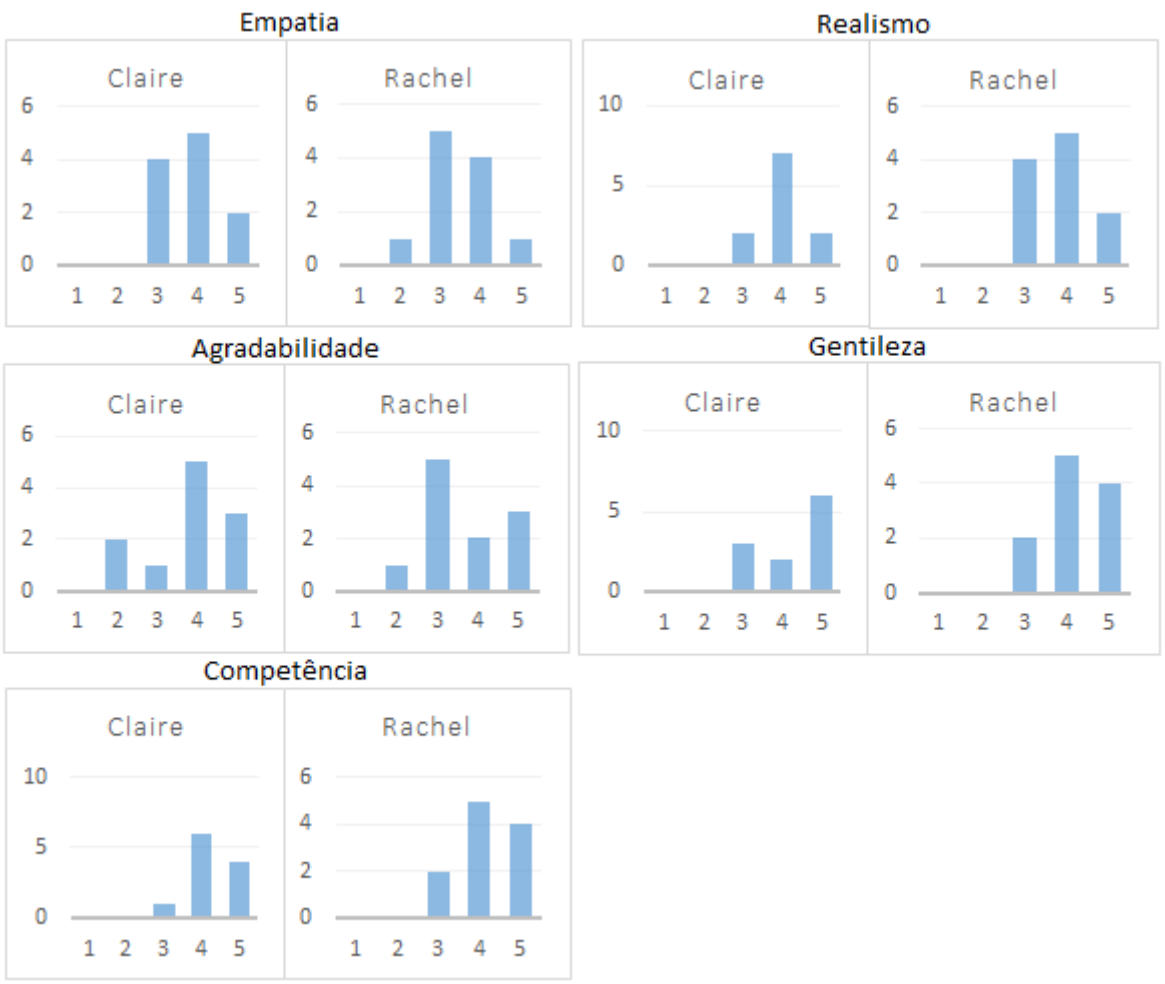


Figura 30 – Histogramas representando as frequências das notas em cada amostra do experimento A.

Tabela 3 – Valores D_n obtidos no teste de normalidade Kolmogorov-Smirnov aplicado ao experimento A.

Característica	D_n (Claire)	D_n (Rachel)
Empatia	0,2257	0,2557
Realismo	0,3182	0,2257
Agradabilidade	0,1603	0,2777
Gentileza	0,2107	0,2320
Competência	0,2998	0,2549

Tabela 4 – Resultados obtidos com aplicação do teste-t pareado unicaudal sobre as amostras do experimento A. As colunas *média das diferenças* e *desvio-padrão* referem-se às médias das diferenças entre as avaliações das duas condições ($\bar{x}_{\text{Claire}} - \bar{x}_{\text{Rachel}}$), tomadas par a par.

Característica	Média das diferenças	Desvio-padrão	Valor-p
Empatia	0,3636	0,6742	0,0519
Realismo	0,1818	0,9816	0,2764
Agradabilidade	0,1818	1,1677	0,3084
Gentileza	0,0909	0,7006	0,3380
Competência	0,0909	0,5394	0,2942

5.2 Experimento B

O segundo experimento foi realizado nos dias 2 e 3 de março de 2016, em uma sala de aula e na biblioteca do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP). Contou com 10 participantes (5 homens, 5 mulheres), com idade entre 16 e 18 anos. Trata-se de alunos de um programa de pré-iniciação científica do ICMC, os quais eram, também, estudantes de nível técnico em Informática.

A configuração do experimento, quanto aos procedimentos adotados, foi semelhante ao experimento A, porém com duas alterações. A primeira foi a introdução de um limite de tempo para finalizar cada partida, estipulado em 2min20seg, que resultou na inclusão de um cronômetro em contagem regressiva na interface do aplicativo. Essa alteração foi necessária para assegurar a conclusão do experimento, do contrário o tempo total disponível poderia ser insuficiente. A segunda foi a elaboração de um novo questionário, que contou com um enfoque maior nas sensações afetivas experimentadas pelo participante. Trata-se do questionário apresentado no Apêndice D. A Tabela 5 mostra as questões nele contidas, todas de múltipla escolha, junto às respostas disponibilizadas.

Tabela 5 – Questões e respostas disponibilizadas.

	Questão	**Respostas disponibilizadas**
Q1:	Que emoções você experimentou durante o jogo?	Qualquer conjunto das seis emoções de Ekman ou *nenhuma*
Q2:	Você acredita que a personagem reconheceu como você se sentiu durante o jogo?	Sim / Não
Q3:	Em uma escala de 1 a 5, o quão corretamente a personagem reagiu às suas emoções?	1 a 5
Q4:	Em uma escala de 1 a 5, o quanto você acredita que as suas emoções influenciaram as emoções da personagem?	1 a 5
Q5:	Em uma escala de 1 a 5, quanta empatia você acredita ter sentido diante da personagem?	1 a 5

A questão Q1 poderia ser respondida por um conjunto de emoções discretas do modelo de Ekman (1999). Ela foi introduzida para examinar, por meio de auto-relatos, a variedade de emoções vivenciadas no decorrer do experimento. Esperava-se que essas emoções fossem decorrentes, em especial, das circunstâncias elicitadas no decorrer jogo, estando associadas às frustrações e aos méritos percebidos pelo participante. No entanto, não seria possível verificar se as emoções reconhecidas pelo classificador corresponderiam às emoções auto-reportadas, uma vez que não haveria uma forma de analisar a evolução destas ao longo do tempo.

As outras questões avaliaram a percepção dos participantes em relação às características afetivas do agente contido no aplicativo. A questão Q2, que admitia uma resposta binária, do tipo "sim ou não", permitiu que cada participante reconhecesse explicitamente se a personagem

virtual causou a impressão de ter reconhecido a suas emoções. Já as questões Q3, Q4 e Q5, que admitiam avaliações escalares de 1 a 5, foram incluídas para permitir uma comparação mais refinada entre as duas abordagens. As respostas a essas questões, a exemplo das questões apresentadas no experimento A, foram submetidas a testes de significância estatística.

A questão Q5 foi acompanhada de uma definição da palavra *empatia*. Definiu-se empatia como "forma de identificação intelectual ou afetiva de um sujeito com uma pessoa, uma ideia ou uma coisa", conforme constava no Dicionário Priberam da Língua Portuguesa[1].

5.2.1 Resultados

Nos testes realizados, os resultados deste experimento não apresentaram significância estatística, provavelmente devido à limitação de tempo estipulada. Uma discussão mais aprofundada será apresentada na seção 5.4.

As respostas de cada participante à questão Q1, relativas a Claire (condição em análise) e a Rachel (condição controle), constam, respectivamente, na Tabela 6 e na Tabela 7. Em observações realizadas ao fim das sessões, os participantes atribuíram as causas das emoções reportadas tanto a eventos ocorridos durante o jogo quanto à sua percepção geral da personagem. Um dos participantes, por exemplo, irritou-se com a frequência com que se emitiam comportamentos proativos. Merece atenção o fato de seis participantes terem relatado sentir a emoção *surpresa* diante de Rachel, sendo que somente dois relataram essa emoção diante de Claire. Ademais, as emoções auto-reportadas por cada participante variaram pouco entre as duas condições.

Tabela 6 – Respostas à questão Q1, por participante, para a condição Claire.

	Nenhuma	Alegria	Tristeza	Aversão	Medo	Raiva	Surpresa
1		x					
2		x		x		x	x
3		x					
4		x					x
5		x					
6		x				x	
7	x						
8				x			
9		x				x	
10	x						

A Tabela 8 mostra as respostas de cada participante à questão Q2. Em relação a Claire, a proporção entre os participantes convencidos de que a personagem reconheceu suas emoções e os que não tiveram essa sensação foi 8:2, enquanto, para Rachel, essa proporção foi 6:4. Apesar desse dado, não se pode afirmar que as duas abordagens foram percebidas diferentemente, uma vez que o experimento é *intra-sujeito* e, como se pôde observar, 8 dos 10 participantes deram

[1] http://priberam.pt/dlpo/empatia, acesso em 01/03/2016

Tabela 7 – Respostas à questão Q1, por participante, para a condição Rachel.

	Nenhuma	Alegria	Tristeza	Aversão	Medo	Raiva	Surpresa
1		x					
2		x		x		x	x
3		x					
4		x					x
5		x				x	
6		x					x
7							x
8							x
9					x		x
10	x						

a mesma resposta para as duas abordagens (i.e., duas respostas *sim* ou duas respostas *não*), não tendo percebido, nesse quesito, diferença substancial entre as duas condições. As exceções foram os participantes 1 e 2, que relataram ter acreditado que Claire reconheceu suas emoções, diferentemente de Rachel. Alguns participantes afirmaram que sua percepção foi afetada porque eles reconheceram a voz sintetizada pela interface de diálogo, que é, de fato, a mesma voz emitida pelo Google Tradutor.

Tabela 8 – Respostas para a questão Q2, por participante. Embora a quantidade de respostas positivas tenha sido maior para Claire do que para Rachel, a maioria dos participantes não viu diferença entre as duas abordagens.

Participante	1	2	3	4	5	6	7	8	9	10	Total-Sim	Total-Não
Claire	S	S	S	S	S	S	N	S	S	N	8	2
Rachel	N	N	S	S	S	S	N	S	S	N	6	4

Para as questões Q3, Q4 e Q5, a exemplo do procedimento adotado no experimento A, foi aplicado o teste Kolmogovorv-Smirnov para verificar a aderência das amostras à normalidade. Os histogramas constam na Figura 31. A planilha usada para esse cômputo está reproduzida no Apêndice B.

Da Tabela 1, tem-se que o valor crítico, para 10 participantes e nível de significância $\alpha = 0,05$, é $D_{n,\alpha} = 0,410$. Como se pode observar na Tabela 9, todos os valores D_n situaram-se abaixo desse limiar, concluindo que a distribuição normal é, em todos os casos verificados, um bom ajuste para os dados das amostras. Portanto, seria adequado aplicar o teste-t sobre essas amostras.

Por meio da aplicação do teste-t, verificou-se que não há significância estatística em nenhuma das três questões. As questões Q4 e Q5 mostraram uma tendência mais aparente à significância, com $p = 0,0918$, porém a média das diferenças foi negativa. Isso significa que, caso se tivesse detectado significância, o resultado apontaria que a avaliação de Rachel teria superado a de Claire, contrariando as expectativas.

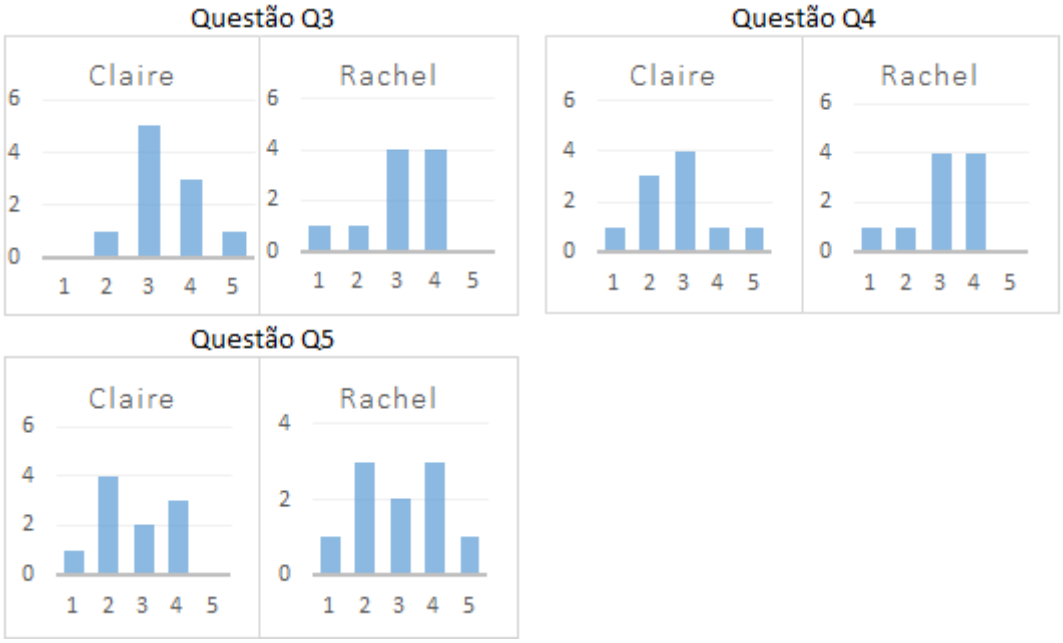


Figura 31 – Histogramas representando as frequências das notas em cada amostra do experimento B.

Tabela 9 – Valores D_n obtidos no teste de normalidade Kolmogorov-Smirnov aplicado ao experimento B.

Questão	D_n (Claire)	D_n (Rachel)
Q3	0,2278	0,0918
Q4	0,1574	0,0918
Q5	0,2002	0,1523

Tabela 10 – Resultados obtidos com aplicação do teste-t pareado unicaudal sobre as amostras do experimento B. As colunas *média das diferenças* e *desvio-padrão* referem-se às médias das diferenças entre as avaliações das duas condições ($\bar{x}_{\text{Claire}} - \bar{x}_{\text{Rachel}}$), tomadas par a par.

Questão	Média das diferenças	Desvio-padrão	Valor-p
Q3	0,3000	0,9487	0,1717
Q4	-0,3000	0,6749	0,0967
Q5	-0,3000	0,6749	0,0967

5.3 Outros Experimentos

Além dos experimentos A e B, outros dois foram realizados. Estes, todavia, não foram foram conduzidos com número de participantes suficiente para se realizar uma análise precisa de seus resultados. De todo modo, os eventos observados no seu decorrer foram importantes para traçar diretrizes para o seguimento do trabalho, ou mesmo para obter relatos que pudessem sugerir diferentes interpretações para os resultados gerais observados.

No dia 25 de fevereiro de 2016, cerca de uma semana antes do experimento B, foi realizado um estudo com crianças do Projeto Pequeno Cidadão, localizado na USP de São Carlos, com idade de 11 a 12 anos. Para fins de referência, esse experimento será identificado

como P1. Trata-se de crianças em situação de vulnerabilidade econômica, portanto, conforme se verificou, com pouca afinidade diante de artefatos tecnológicos. A ideia seria verificar os efeitos da estratégia proposta em atividades com crianças, o que poderia ser estendido, futuramente, a atividades pedagógicas.

Houve certa dificuldade das crianças em conseguir manipular a interface do aplicativo. Do total de seis participantes, quatro relataram ter sentido medo ou nervosismo, devido às dificuldades tanto em manipular o *smartphone*, quanto em resolver o jogo de palavras. Essa situação contribuiu para a perda do engajamento das crianças no decorrer de cada sessão. Além disso, o sistema para reconhecimento de emoções não pôde atuar corretamente, visto que, durante boa parte do tempo, os participantes posicionaram o *smartphone* de tal modo que sua face permanecia apenas parcialmente enquadrada na imagem da câmera frontal.

Além do experimento P1, um último experimento de caráter exploratório foi conduzido, após o experimento B, para verificar se as razões do resultado não significativo deste resultou de uma má atuação do reconhecedor de emoções, possivelmente devido às condições de iluminação. Esse experimento será identificado como P2.

Graduandos em Ciência da Computação e pós-graduandos em Ciências da Computação e Matemática Computacional pelo ICMC experimentaram as duas condições, nos mesmos moldes do experimento B. A diferença foi que o reconhecedor automático de emoções foi substituído por uma abordagem *Wizard of Oz* (RIEK, 2012), em que o autor desta dissertação se responsabilizou pelo reconhecimento de emoções, comparando as expressões faciais dos participantes às expressões apresentadas no modelo de Ekman. Por meio de análises de resultados parciais, notou-se que o resultado continuava não significativo. Através de entrevistas, os participantes revelaram que o tempo reduzido de interação era insuficiente para cultivar maior intensidade afetiva. Alguns sequer notaram diferença entre as duas condições.

5.4 Discussão

Apesar do relativo sucesso do experimento A em produzir a sensação de empatia, não se averiguou se o resultado observado dependeu das respostas da estratégia de interação às emoções do indivíduo, ou se o fato de haver alternância entre paradigmas seria suficiente para tal. Isso ocorreu porque, devido à quantidade insuficiente de participantes para realização de experimentos *entre sujeitos* e ao tempo relativamente longo decorrido para realizar cada sessão dos experimentos *intra-sujeito*, não seria possível introduzir mais de duas condições.

Para estabelecer a condição controle, consideraram-se três possibilidades. A primeira seria, a exemplo de Leite *et al.* (2012), alternar os paradigmas aleatoriamente, ignorando os resultados do reconhecimento de emoções. Essa condição poderia ser demasiadamente rigorosa, uma vez que, se a própria alternância entre paradigmas com diferentes características afetivas pudesse influenciar a percepção do participante diante do agente, isso já seria um resultado

interessante. Em outras palavras, o fato de essa alternância não ser resultado direto das emoções experimentadas pelo participante seria menos importante do que as melhorias produzidas por essa estratégia em relação a uma estratégia estática, que conservasse sempre um mesmo paradigma.

As outras duas possíveis condições consistiriam em conservar estaticamente somente um dos paradigmas, seja ele o *normal* ou o *amigável*. Caso o paradigma *normal* fosse selecionado como condição controle, um eventual resultado favorável à estratégia implementada poderia ser devido à maior intimidade produzida nos momentos em que o paradigma *amigável* atuasse, uma vez que esse paradigma foi concebido para produzir maior intimidade. Nesse caso, as alternâncias entre os dois paradigmas poderiam não ser o fator determinante, e dificilmente algum resultado poderia ser satisfatoriamente conclusivo. Por essa razão, optou-se por uma versão que permanecesse constantemente na estratégia *amigável*.

De fato, notadamente durante o experimento B, obteve-se evidência informal de que a personagem na condição controle, permanentemente no paradigma *amigável*, parecia "importar-se mais"com o participante, justificando algumas avaliações favoráveis a ela. Por outro lado, no experimento A, não houve relatos desse tipo, possivelmente porque o maior tempo de exposição à aplicação tenha permitido observar que esse paradigma também estava presente na estratégia implementada. Nesse sentido, espera-se que a condição controle adotada tenha sido a mais correta possível.

A estratégia para alternância de paradigmas baseada em reconhecimento de emoções também poderia ter sido implementada de outro modo, associando emoções positivas ao paradigma *normal*, e negativas, ao paradigma *amigável*. Havendo condições, teria sido oportuno incluir uma condição em que a estratégia de interação operasse dessa maneira. Como não houve essa possibilidade, foi preciso tomar uma decisão arbitrária nesse sentido. A abordagem adotada foi escolhida sob a hipótese de que reproduzir o estado afetivo do indivíduo pode produzir maior identificação do que emitir pistas emocionais de natureza oposta.

O experimento A, apresentado na seção 5.1. foi conduzido assim que a implementação do aplicativo foi finalizada. Foi a primeira vez em que o ambiente foi avaliado por usuários humanos. Obteve-se engajamento dos participantes, curiosos em relação a diversos aspectos da implementação do aplicativo. Essa receptividade permitiu sessões mais longas, com desafios de maior nível de dificuldade, o que contribuiu para que as diferenças entre as duas condições se tornassem mais perceptíveis a eles. Esperando obter informações adicionais sobre a dimensão do efeito produzido, decidiu-se realizar mais experimentos.

Embora o experimento P1, conduzido em seguida, não tenha produzido resultados quantificáveis, ele chamou à atenção alguns cuidados que precisariam ser tomados dali em diante. Em primeiro lugar, seria importante assegurar que os participantes posicionem o *smartphone* de modo que suas faces sejam enquadradas completamente na imagem da câmera frontal. Em segundo lugar, para que esse ambiente pudesse ser adequadamente experimentado com crianças, modificações precisariam ser realizadas. Uma alternativa seria alterar a imagem da personagem

virtual, mantendo somente uma face amigável para crianças, e introduzir o ambiente a um robô móvel, conforme realizado em Russell *et al.* (2014). A curiosidade e a presença física poderiam, ao menos em curto prazo, aliviar o efeito aparentemente intimidador da aplicação embarcada em um dispositivo móvel e aumentar o engajamento.

Seguiu-se o experimento B, relatado na seção 5.2. O questionário, conforme já apresentado, conteve questões mais variadas na forma de responder, porém mais restritas à dimensão afetiva da interação. A não significância dos resultados, em todas as questões, levantou duas suspeitas. A primeira foi de que, devido a condições adversas, como a iluminação dos ambientes, os sistemas para detecção de face e reconhecimento de emoções não funcionaram corretamente. A segunda foi de que, para se produzir bons resultados nesses quesitos, seria necessário manter a interação por mais tempo, como ocorreu no experimento A.

Com o experimento P2, esperava-se detectar se, substituindo o classificador automático de emoções por um reconhecimento realizado por um humano, as diferenças entre as duas abordagens voltassem a se tornar significativas. Nesse experimento, novamente os resultados não caminharam para a significância, e relatos apontaram para a segunda suspeita elencada acima. De todo modo, pôde-se observar uma dificuldade em identificar as emoções dos indivíduos com base em suas expressões faciais, que, em alguns casos, permaneceram praticamente impassíveis durante toda a sessão.

Ainda que, no experimento B, os participantes tenham reportado terem sentido uma variedade de emoções, verificou-se que essas emoções nem sempre se traduziram em expressões faciais muito identificáveis. Uma possível explicação seria que as pessoas estão habituadas a lidar com dispositivos móveis, e o fato de haver um agente virtual personificado pode ser insuficiente para eliminar a possível inibição de comunicar-se por expressões faciais nesse tipo se interação. Essa hipótese está em consonância com as descobertas apresentadas na literatura. Embora trabalhos como Jo *et al.* (2013) e Darling, Nandy e Breazeal (2015) tenham sugerido que humanos podem perceber robôs como companhia ou sentir empatia afetiva diante deles, esse efeito pode ser menos proeminente no caso de agentes virtuais, como Wainer *et al.* (2006) observaram.

De todo modo, esses resultados confirmam que, a despeito da acurácia de um sistema de reconhecimento de emoções unimodal em bases de dados controladas, expressões faciais podem ser insuficientes para detectar muitas das nuances afetivas geradas ao longo de interações conduzidas na prática. Para minimizar esse problema, seria adequada a introdução de sistemas multimodais para reconhecimento de emoções (DHALL *et al.*, 2015), que teriam subsídios para detectar mudanças de emoções mesmo diante de expressões faciais pouco afetadas.

Como já discutido, os experimentos foram conduzidos somente no modo *jogo*. O noticiário foi incluído para que a aplicação pudesse ter múltiplas funcionalidades, constituindo uma versão simplificada de um assistente pessoal virtual. O sistema empregado para recomendar notícias foi baseado em conteúdo e com o mesmo algoritmo de aprendizado por reforço usado

para selecionar comportamentos no jogo. Trata-se de um sistema limitado, se comparado aos sistemas de recomendação que têm sido desenvolvidos (TKALCIC; GEMMIS; SEMERARO, 2014), tendo sido introduzido apenas para diversificar as funcionalidades da aplicação final. Construir um sistema de recomendação no nível do estado da arte demandaria um investimento proibitivo de tempo e esforço de desenvolvimento. Como não há uma base de dados disponível nos moldes da aplicação proposta nesta dissertação, seria necessário obter uma grande quantidade de dados de usuários reais.

A única maneira de fazer isso seria, com todas as dificuldades e implicações de natureza ética, solicitar a um grupo substancial de voluntários que usasse, durante alguns meses, um aplicativo que monitorasse suas atividades, o qual poderia interrompe-los periodicamente para solicitar avaliações de itens. Isso ocorre porque, ao menos em alguma medida, seria necessário aplicar filtragem colaborativa (BOBADILLA *et al.*, 2013). Um trabalho como esse seria, por si, tema para um projeto de mestrado, voltado a recuperação de informação e mineração de dados. Por esse motivo, optou-se por dar enfase a aspectos de interação, da forma como toda a discussão tem sido conduzida no presente trabalho.

5.5 Considerações finais

Foram apresentados, em detalhes, dois experimentos conduzidos para avaliar a arquitetura proposta e implementada. O primeiro experimento revelou que o sistema pode produzir uma sensação de empatia mais proeminente, segundo avaliações realizadas pelos participantes através de questionários. O segundo, que contou com limitação de tempo e cronômetro em contagem regressiva, não mostrou diferença significativa entre as condições consideradas. Nos dois casos, a condição analisada, chamada arbitrariamente de Claire, foi comparada a uma condição controle, chamada de Rachel, dotada de uma estratégia de interação que permanecia, constantemente, no paradigma *amigável*. Em seguida, foi apresentada uma discussão sobre esses resultados, levando em conta alguns experimentos exploratórios ou inacabados.

Finalmente, no próximo capítulo, serão tecidas as conclusões desta dissertação.

CAPÍTULO

6

CONCLUSÃO

Nesta Dissertação, foi proposta uma estratégia baseada em emoções para interação humano-robô. Esta estratégia consistiu em conceber um ambiente de interação em que um agente virtual consiga adaptar seu comportamento de acordo com a análise de emoções expressas pelo usuário. A emoção detectada do usuário foi mapeada para um padrão de comportamento, a fim de aprimorar a sua percepção acerca do agente com que está interagindo. Isso pode ser importante em situações que requeiram maior engajamento do usuário, como guias robóticos em museus ou robôs projetados para convívio social, como educadores para crianças ou acompanhantes para idosos. Não se propôs tomar a resposta emocional do usuário para inferir algum tipo de significado semântico adicional na interação, tampouco representar emoções artificiais em modelos cognitivos.

O ambiente contou com uma base computacional Windows e com um *smartphone* Android, os quais se comunicam por conexões de rede. Na base computacional, um programa servidor oferecia rotinas para executar o reconhecedor de emoções ou para obter notícias de portais específicos. No *smartphone*, foi implementado o aplicativo responsável pela interação, que incorpora uma personagem virtual. Essa arquitetura, em que uma base computacional realiza algumas tarefas, enquanto a aplicação de um dispositivo móvel dedica-se às funcionalidades interativas, pode ser aplicada a uma infinidade de situações, sendo a base por trás de muitos sistemas de computação em nuvem.

Estudos com usuários foram conduzidos para validar a arquitetura apresentada e para testar se a estratégia de interação proposta produz uma interação prazerosa, bem como se ela faz com que os usuários sintam mais empatia diante da personagem. Como controle, foi experimentada uma versão do aplicativo em que a estratégia de interação exibia somente um dos paradigmas implementados. Os resultados sugeriram que o sistema pode produzir sensação de empatia, contanto que o tempo de interação seja suficiente para estabelecer uma relação. Isso foi corroborado tanto pelos resultados diferentes obtidos nos experimentos A e B, quanto por relatos

de participantes em experimentos exploratórios.

Com isso, foram apresentadas contribuições adicionais aos trabalhos relacionados na Seção 3, uma vez que foi produzida evidência experimental de que o nível de empatia experimentado por um humano diante de um personagem artificial é influenciado positivamente se esse agente reage ao seu estado emocional segundo a estratégia proposta. Mais especificamente, as contribuições foram uma proposta de ambiente interativo dotado de estratégia baseada em emoções inferidas por análise de expressões faciais, arquiteturas para implementação dessa estratégia, em diferentes níveis de abstração, e arcabouço experimental quanto à sua aplicação em sessões com participantes humanos.

Esta pesquisa deu origem a um artigo publicado e dois artigos submetidos, em análise para publicação. A publicação se deu em um congresso internacional, o *The Eighth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2016)*, realizado em Roma, na Itália. Os outros dois artigos em análise foram submetidos para o *13rd Latin American Robotics Symposium (LARS 2016)*, um congresso internacional a ser realizado em Recife, e para o *XXI Congresso Brasileiro de Automática (CBA 2016)*, um congresso nacional a ser realizado em Vitória. Seguem as referências, respectivamente, para estes três artigos.

- RANIERI, C. M.; FERASOLI, H.; ROMERO, R. A. F. A mobile virtual character with emotion-aware strategies for human-robot interaction. THE EIGHTH INTERNATIONAL CONFERENCE ON ADVANCED COGNITIVE TECHNOLOGIES AND APPLICATIONS (COGNITIVE 2016), International Academy, Research and Industry Association (IARIA). Roma, Itália, 2016, pp. 124-128.

- RANIERI, C. M.; FERASOLI, H.; ROMERO, R. A. F. An Emotion-Based Interaction Strategy to Improve Human-Robot Interaction. 13RD LATIN AMERICAN ROBOTICS SYMPOSIUM (LARS 2016). Recife, 2016. Manuscrito submetido para publicação.

- RANIERI, C. M.; ROMERO, R. A. F.. Ambiente baseado em Emoções para Interação entre Humanos e Robôs Sociais. XXI CONGRESSO BRASILEIRO DE AUTOMÁTICA (CBA2016). Vitória, 2016. Manuscrito submetido para publicação.

Em trabalhos futuros, o aplicativo desenvolvido poderá ser dotado de mais funcionalidades, a serem avaliadas individualmente por meio de experimentos. Um possível aprimoramento seria incluir um sistema de recomendação de notícias mais robusto, em um projeto dedicado a esta finalidade. Com uma quantidade suficiente de funcionalidades adicionais, o aplicativo implementado poderia se tornar um assistente pessoal, possibilitando disponibilização pública e obtenção de dados em maior escala. Para isso, porém, seria necessário se ter um reconhecedor de emoções que não dependesse de uma base computacional remota, o que poderia ser implementado em um trabalho dedicado a portar o sistema original para plataforma Android ou obtido por

meio de serviços em nuvem, como a recentemente disponibilizada Google Cloud Vision API[1], que oferece esse serviço por meio da nuvem.

Além disso, a estratégia proposta poderá ser integrada a situações mais específicas, como parte de ambientes maiores. Um possível passo nesse sentido seria adaptar o ambiente para operar em um robô móvel, posteriormente adaptando-o para ambientes domésticos. Trata-se de uma área com recente apelo não somente acadêmico, como também comercial, o que se tem traduzido no lançamento de robôs como o Jibo[2], o Zenbo[3] e o Cozmo [4]. Nesse sentido, a estratégia proposta consistiria em um módulo de uma arquitetura maior, que incluiria funcionalidades diversas tanto de casas inteligentes quanto de assistentes pessoais virtuais.

Outra possibilidade é realizar estudos de longo termo, isto é, conduzidos no decorrer de múltiplas sessões com um conjunto fixo de usuários. Estudos dessa natureza avaliam a evolução da interação ao longo do tempo, avaliando, entre outros, o quanto a sensação de novidade é importante para o efeito de uma proposta de robótica social. Com isso, podem-se propor e refinar estratégias para manter o engajamento, importantes para quaisquer ideias de se ter robôs integrados a diferentes ambientes domésticos, profissionais ou terapêuticos.

Aprimorar o reconhecedor de emoções, tornando-o multimodal, também poderia tornar as reações mais imediatas e corretas, portanto essa área é importante para trabalhos futuros que produzam reações a estímulos emocionais. Este trabalho detectou dificuldades oriundas de sistemas unimodais, apontando que novos desenvolvimentos nesse sentido são fundamentais e devem preceder iniciativas mais ambiciosas envolvendo estratégias de interação baseadas em reconhecimento de emoções. No entanto, para ser aplicável a estratégias interativas, além de apresentar boa acurácia, um reconhecedor de emoções multimodal deve ter flexibilidade para se adequar a diferentes situações.

[1] https://cloud.google.com/vision/, acesso em 30/05/2016.
[2] https://www.jibo.com/, acesso em 07/06/2016.
[3] https://zenbo.asus.com/, acesso em 07/06/2016.
[4] https://anki.com/en-us/cozmo, acesso em 10/08/2016.

REFERÊNCIAS

ADALGEIRSSON, S. O.; BREAZEAL, C. Mebot: a robotic platform for socially embodied presence. In: PROCEEDINGS OF THE 5TH ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION. Osaka, Japan,, 2010. p. 15–22. Citado 2 vezes nas páginas 45 e 50.

ALONSO-MARTIN, F.; MALFAZ, M.; SEQUEIRA, J.; GOROSTIZA, J. F.; SALICHS, M. A. A multimodal emotion detection system during human–robot interaction. **Sensors**, v. 13, n. 11, p. 15549–15581, 2013. Citado na página 37.

ALVES, S. F. d. R.; FERASOLI, H. Intelligent control architecture for assistive mobile robots. **Journal of Control, Automation and Electrical Systems**, Springer, p. 1–12, 2016. Citado 3 vezes nas páginas 44, 46 e 50.

ARNOLD, L. Emobie^{TM}: A robot companion for children with anxiety. In: IEEE PRESS. **The Eleventh ACM/IEEE International Conference on Human Robot Interation**. [S.l.], 2016. p. 413–414. Citado 2 vezes nas páginas 42 e 49.

AUER, P.; CESA-BIANCHI, N.; FISCHER, P. Finite-time analysis of the multiarmed bandit problem. **Machine learning**, v. 47, n. 2-3, p. 235–256, 2002. Citado na página 63.

BAINBRIDGE, W. A.; HART, J. W.; KIM, E. S.; SCASSELLATI, B. The benefits of interactions with physically present robots over video-displayed agents. **International Journal of Social Robotics**, Springer, v. 3, n. 1, p. 41–52, 2011. Citado na página 30.

BARTNECK, C.; KULIĆ, D.; CROFT, E.; ZOGHBI, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. **International journal of social robotics**, Springer, v. 1, n. 1, p. 71–81, 2009. Citado na página 69.

BOBADILLA, J.; ORTEGA, F.; HERNANDO, A.; GUITIéRREZ, A. Recommender systems survey. **Knowledge-Based Systems**, v. 46, p. 109–132, 2013. Citado na página 78.

BOWER, G. H. How might emotions affect learning. In: ____. **The handbook of emotion and memory: Research and theory**. [S.l.]: Lawrence Erlbaum Associates, 1992. p. 3–31. Citado na página 25.

BRAVE, S.; NASS, C. Emotion in human-computer interaction. In: ____. **The human-computer interraction handbook**. [S.l.: s.n.], 2002. p. 53–58. Citado 3 vezes nas páginas 13, 24 e 26.

BREAZEAL, C. Emotion and sociable humanoid robots. **International Journal of Human-Computer Studies**, Elsevier, v. 59, n. 1, p. 119–155, 2003. Citado 3 vezes nas páginas 13, 42 e 49.

BREAZEAL, C.; BROOKS, R. Robot emotion: A functional perspective. **Who needs emotions**, Oxford University Press.: NY, p. 271–310, 2005. Citado na página 20.

BREAZEAL, C. L. **Designing sociable robots**. [S.l.]: MIT Press, 2004. Citado na página 29.

CARROLL, L. **Doublets: A word-puzzle**. [S.l.]: Macmillan, 1879. Citado 2 vezes nas páginas 53 e 62.

CAñAMERO, L. Bridging the gap between hri and neurosciencein emotion research: Robots as models. In: 9TH ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION. Bielefeld, Germany,, 2014. Citado na página 20.

DAMASIO, A. **Descartes' error: Emotion, reason, and the human brain**. [S.l.]: Penguin Books, 2005. Citado na página 19.

DAMIANO, L.; DUMOUCHEL, P.; LEHMANN, H. Towards human–robot affective co-evolution overcoming oppositions in constructing emotions and empathy. **International Journal of Social Robotics**, Springer, v. 7, n. 1, p. 7–18, 2015. Citado na página 29.

DARLING, K.; NANDY, P.; BREAZEAL, C. Empathic concern and the effect of stories in human-robot interaction. In: 24TH IEEE INTERNATIONAL SYMPOSIUM ON ROBOT AND HUMAN INTERACTIVE COMMUNICATION (RO-MAN). Kobe, Japan, 2015. p. 770–775. Citado 6 vezes nas páginas 13, 20, 31, 40, 49 e 77.

DARWIN, C.; EKMAN, P.; PRODGER, P. **The expression of the emotions in man and animals**. [S.l.]: Oxford University Press, USA, 1998. Citado 2 vezes nas páginas 19 e 24.

DECETY, J.; COWELL, J. M. The complex relation between morality and empathy. **Trends in Cognitive Sciences**, v. 18, n. 7, p. 337–339, 2014. Citado na página 30.

DHALL, A.; MURTHY, O. R.; GOECKE, R.; JOSHI, J.; GEDEON, T. Video and image based emotion recognition challenges in the wild: Emotiw 2015. In: ACM. **Proceedings of the 2015 ACM on International Conference on Multimodal Interaction**. [S.l.], 2015. p. 423–426. Citado 4 vezes nas páginas 20, 27, 38 e 77.

EKMAN, P. Basic emotions. **Handbook of cognition and emotion**, v. 4, p. 5–60, 1999. Citado 9 vezes nas páginas 13, 21, 25, 26, 37, 47, 52, 59 e 71.

EKMAN, P.; FRIESEN, W. V. **Unmasking the face: A guide to recognizing emotions from facial clues**. [S.l.]: Ishk, 2003. Citado na página 25.

GÁCSI, M.; KIS, A.; FARAGÓ, T.; JANIAK, M.; MUSZYŃSKI, R.; MIKLÓSI, Á. Humans attribute emotions to a robot that shows simple behavioural patterns borrowed from dog behaviour. **Computers in Human Behavior**, Elsevier, v. 59, p. 411–419, 2016. Citado 5 vezes nas páginas 13, 31, 38, 39 e 49.

GIL, G. B.; BERLANGA, A.; MOLINA., J. M. Emotioncontext: User emotion dataset using smartphones. In: AMBIENT ASSISTED LIVING AND HOME CARE: 4TH INTERNATIONAL WORKSHOP (IWAAL 2012). Vitoria-Gasteiz, Espanha,: Springer, 2012. p. 371–374. Citado na página 25.

GOCKLEY, R.; BRUCE, A.; FORLIZZI, J.; MICHALOWSKI, M.; MUNDELL, A.; ROSENTHAL, S.; SELLNER, B.; SIMMONS, R.; SNIPES, K.; SCHULTZ, A. C. *et al.* Designing robots for long-term social interaction. In: IEEE. **Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on**. [S.l.], 2005. p. 1338–1343. Citado 2 vezes nas páginas 43 e 50.

GOODRICH, M. A.; SCHULTZ, A. C. Human-robot interaction: a survey. **Foundations and trends in human-computer interaction**, v. 1, n. 3, p. 203–275, 2007. Citado na página 29.

HEGEL, F.; MUHL, C.; WREDE, B.; HIELSCHER-FASTABEND, M.; SAGERER, G. Understanding social robots. In: ACHI'09. SECOND INTERNATIONAL CONFERENCES ON ADVANCES IN COMPUTER-HUMAN INTERACTIONS. [S.l.], 2009. Citado 2 vezes nas páginas 13 e 29.

HESS, U.; THIBAULT, P. Darwin and emotion expression. **American Psychologist**, American Psychological Association, v. 64, n. 2, p. 120, 2009. Citado 2 vezes nas páginas 19 e 24.

HOCKSTEIN, N. G.; GOURIN, C. G.; FAUST, R. A.; TERRIS, D. J. A history of robots: from science fiction to surgical robotics. **Journal of robotic surgery**, v. 1, n. 2, p. 113–118, 2007. Citado na página 28.

HUNTER, W. G.; HUNTER, J. S.; GEORGE, E. P. **Statistics for experimenters: an introduction to design, data analysis, and model building**. [S.l.]: Wiley New York, 1978. Citado na página 33.

JO, D.; HAN, J.; CHUNG, K.; LEE, S. Empathy between human and robot? In: PROCEEDINGS OF THE 8TH ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION. Tokyo, Japan,, 2013. p. 151–152. Citado 3 vezes nas páginas 40, 49 e 77.

KÄCHELE, M.; SCHELS, M.; MEUDT, S.; PALM, G.; SCHWENKER, F. Revisiting the emotiw challenge: how wild is it really? **Journal on Multimodal User Interfaces**, Springer, p. 1–12, 2016. Citado na página 38.

KAHOU, S. E.; PAL, C.; BOUTHILLIER, X.; FROUMENTY, P.; GÜLÇEHRE, Ç.; MEMISEVIC, R.; VINCENT, P.; COURVILLE, A.; BENGIO, Y.; FERRARI, R. C. *et al.* Combining modality specific deep neural networks for emotion recognition in video. In: ACM. **Proceedings of the 15th ACM on International conference on multimodal interaction**. [S.l.], 2013. p. 543–550. Citado na página 38.

KIRBY, R.; FORLIZZI, J.; SIMMONS, R. Affective social robots. **Robotics and Autonomous Systems**, Elsevier, v. 58, n. 3, p. 322–332, 2010. Citado 4 vezes nas páginas 41, 43, 49 e 50.

LANGNER, O.; DOTSCH, R.; BIJLSTRA, G.; WIGBOLDUS, D. H.; HAWK, S. T.; KNIPPENBERG, A. van. Presentation and validation of the radboud faces database. **Cognition and Emotion**, v. 24, n. 8, p. 1377–1388, 2010. Citado na página 47.

LEITE, I.; CASTELLANO, G.; PEREIRA, A.; MARTINHO, C.; PAIVA, A. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In: PROCEEDINGS OF THE SEVENTH ANNUAL ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION. Boston, MA, USA,, 2012. p. 367–374. Citado 6 vezes nas páginas 13, 20, 42, 43, 49 e 75.

LEITE, I.; MARTINHO, C.; PAIVA, A. Social robots for long-term interaction: a survey. **International Journal of Social Robotics**, Springer, v. 5, n. 2, p. 291–308, 2013. Citado na página 49.

LIBRALON, G.; ROMERO, R. A. Geometrical facial modeling for emotion recognition. In: THE 2013 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN). Dallas, TX, USA,, 2013. p. 1–8. Citado 5 vezes nas páginas 47, 48, 50, 52 e 66.

LIBRALON, G. L. **Modelagem computacional para reconhecimento de emoções na análise facial**. São Carlos,, 2014. Citado 2 vezes nas páginas 21 e 47.

LUCEY, P.; COHN, J. F.; KANADE, T.; SARAGIH, J.; AMBADAR, Z.; MATTHEWS, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW). [S.l.], 2010. p. 94–101. Citado na página 47.

MATARIC, M. J. **The Robotics Primer**. [S.l.]: MIT Press, 2007. Citado na página 28.

MORI, M.; MACDORMAN, K. F.; KAGEKI, N. The uncanny valley. **Robotics & Automation Magazine, IEEE**, v. 19, n. 2, p. 98–100, 2012. Citado 2 vezes nas páginas 13 e 30.

NAVARRO, D. **Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners,(version 0.4)**. [S.l.]: Lulu. com, 2014. Citado 2 vezes nas páginas 31 e 33.

O'CONNOR, P.; KLEYNER, A. **Practical reliability engineering**. [S.l.]: John Wiley & Sons, 2011. Citado 3 vezes nas páginas 15, 32 e 33.

PERGHER, G. K.; GRASSI-OLIVEIRA, R.; ÁVILA, L. M. de; STEIN, L. M. Memória, humor e emoção. **Revista de psiquiatria do Rio Grande do Sul**, v. 28, n. 1, p. 61–68, 2006. Citado na página 23.

PICARD, R. W. **Affective computing**. [S.l.]: MIT Press, 2000. Citado na página 19.

PUTTEN, A. M. R. von der; KRAMER, N. C.; HOFFMANN, L.; SOBIERAJ, S.; EIMLER, S. C. An experimental study on emotional reactions towards a robot. **International Journal of Social Robotics**, v. 5, n. 1, p. 17–34, 2013. Citado 4 vezes nas páginas 13, 39, 40 e 49.

RABIEI, M.; GASPARETTO, A. A system for feature classification of emotions based on speech analysis; applications to human-robot interaction. In: SECOND RSI/ISM INTERNATIONAL CONFERENCE ON ROBOTICS AND MECHATRONICS (ICROM). [S.l.], 2014. Citado na página 37.

RANIERI, C. M.; ALVES, S. F. d. R.; SILVA, I. N.; FERASOLI, H. Navegação assistida de robôs móveis baseada em reconhecimento e síntese de fala. In: SBA. **Anais do XX Congresso Brasileiro de Automática**. [S.l.], 2014. p. 79–84. Citado 4 vezes nas páginas 41, 46, 49 e 50.

RIEK, L. D. Wizard of oz studies in hri: a systematic review and new reporting guidelines. **Journal of Human-Robot Interaction**, v. 1, n. 1, p. 119–136, 2012. Citado 2 vezes nas páginas 31 e 75.

RUSSELL, E.; STROUD, A.; CHRISTIAN, J.; RAMGOOLAM, D.; WILLIAMS, A. B. Smile: A portable humanoid robot emotion interface. In: 9TH ACM/IEEE INTERNATIONAL CONFERENCE ON HUMAN-ROBOT INTERACTION, WORKSHOP ON APPLICATIONS FOR EMOTIONAL ROBOTS, HRI14. Bielefeld, Germany,, 2014. p. 1–5. Citado 5 vezes nas páginas 13, 44, 45, 46 e 77.

RUSSELL, J. A.; LEWICKA, M.; NIIT, T. A cross-cultural study of a circumplex model of affect. **Journal of personality and social psychology**, v. 57, n. 5, p. 848, 1989. Citado 3 vezes nas páginas 13, 25 e 26.

SARAGIH, J. M.; LUCEY, S.; COHN, J. F. Deformable model fitting by regularized landmark mean-shift. **International Journal of Computer Vision**, v. 91, n. 2, p. 200–215, 2011. Citado na página 47.

SEBE, N.; COHEN, I.; GEVERS, T.; HUANG, T. S. Emotion recognition based on joint visual and audio cues. In: 18TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICOR 2006). [S.l.]: IEEE, 2006. p. 1136–1139. Citado na página 27.

SPAULDING, S.; BREAZEAL, C. Towards affect-awareness for social robots. In: **2015 AAAI Fall Symposium Series**. [S.l.: s.n.], 2015. Citado na página 49.

SPIEGEL, M. R.; SCHILLER, J. J.; SRINIVASAN, R. A. **Schaum Outline of Probability and Statistics**. 4. ed. [S.l.]: McGraw-Hill Education, New York, USA, 2012. Citado na página 31.

STEPHAN, A. Empathy for artificial agents. **International Journal of Social Robotics**, Springer, v. 7, n. 1, p. 111–116, 2015. Citado na página 21.

STRATOU, G.; MORENCY, L.-P.; DEVAULT, D.; HARTHOLT, A.; FAST, E.; LHOMMET, M.; LUCAS, G.; MORBINI, F.; GEORGILA, K.; SCHERER, S. *et al.* A demonstration of the perception system in simsensei, a virtual human application for healthcare interviews. In: IEEE. **Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on**. [S.l.], 2015. p. 787–789. Citado na página 30.

TANG, L.; JIANG, Y.; LI, L.; ZENG, C.; LI, T. Personalized recommendation via parameter-free contextual bandits. In: PROCEEDINGS OF THE 38TH INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL. [S.l.], 2015. Citado na página 65.

TIAN, Y.-L.; KANADE, T.; COHN, J. F. Handbook of face recognition. Londres,, p. 487–519, 2005. Citado na página 28.

TKALCIC, M.; GEMMIS, M. de; SEMERARO, G. Personality and emotions in decision making and recommender systems. In: **Proceedings of the First International Workshop on Decision Making and Recommender Systems (DMRS2014), Bolzano, Italy, September 18-19, 2014.** [S.l.: s.n.], 2014. p. 14–18. Citado na página 78.

TKALCIC, M.; KOSIR, A.; TASIC, J. Affective recommender systems: the role of emotions in recommender systems. In: THE RECSYS 2011 WORKSHOP ON HUMAN DECISION MAKING IN RECOMMENDER SYSTEMS. [S.l.], 2011. Citado na página 49.

TROVATO, G.; RAMOS, J. G.; AZEVEDO, H.; MORONI, A.; MAGOSSI, S.; ISHII, H.; SIMMONS, R.; TAKANISHI, A. "olá, my name is ana": A study on brazilians interacting with a receptionist robot. In: IEEE. **Advanced Robotics (ICAR), 2015 International Conference on**. [S.l.], 2015. p. 66–71. Citado 2 vezes nas páginas 43 e 50.

VALSTAR, M. F. **Timing is everything: A spatio-temporal approach to the analysis of facial actions**. Londres,, 2008. Citado na página 28.

VINCENT, J. Is the mobile phone a personalized social robot? **Intervalla**, v. 1, n. 1, p. 60–70, 2013. Citado 2 vezes nas páginas 20 e 44.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: PROCEEDINGS OF THE 2001 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2001). Kauai, HI, EUA,, 2001. p. 511–518. Citado na página 48.

WAINER, J.; FEIL-SEIFER, D. J.; SHELL, D. A.; MATARIĆ, M. J. The role of physical embodiment in human-robot interaction. In: IEEE. **Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on**. [S.l.], 2006. p. 117–122. Citado 3 vezes nas páginas 41, 49 e 77.

YOON, J.-W.; CHO, S.-B. An intelligent synthetic character for smartphone with bayesian networks and behavior selection networks. **Expert Systems with Applications**, Elsevier, v. 39, n. 12, p. 11284–11292, 2012. Citado 4 vezes nas páginas 13, 44, 45 e 50.

ZENG, Z.; PANTIC, M.; ROISMAN, G. I.; HUANG, T. S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 31, n. 1, p. 39–58, 2009. Citado na página 19.

APÊNDICE

A

TESTES DE NORMALIDADE - EXPERIMENTO A

Empatia		
Participante	Claire	Rachel
1	3	3
2	3	2
3	4	5
4	5	4
5	4	3
6	3	3
7	3	3
8	4	4
9	4	4
10	4	3
11	5	4
Média	**3,8182**	**3,4545**
Desvio-padrão	**0,7508**	**0,8202**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-3,7538	0,0001	0,0001
2	0	0	0,0000	-2,4218	0,0077	0,0077
3	4	4	0,3636	-1,0898	0,1379	0,2257
4	5	9	0,8182	0,2422	0,5957	0,2225
5	2	11	1,0000	1,5742	0,9423	0,0577
6	0	11	1,0000	2,9062	0,9982	0,0018

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-2,9926	0,0014	0,0014
2	1	1	0,0909	-1,7734	0,0381	0,0528
3	5	6	0,5455	-0,5542	0,2897	0,2557
4	4	10	0,9091	0,6650	0,7470	0,1621
5	1	11	1,0000	1,8842	0,9702	0,0298
6	0	11	1,0000	3,1035	0,9990	0,0010

Realista		
Participante	Claire	Rachel
1	4	3
2	4	3
3	4	4
4	5	5
5	4	4
6	4	4
7	4	3
8	3	4
9	4	5
10	3	4
11	5	3
Média	**4,0000**	**3,8182**
Desvio-padrão	**0,6325**	**0,7508**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-4,7434	0,0000	0,0000
2	0	0	0,0000	-3,1623	0,0008	0,0008
3	2	2	0,1818	-1,5811	0,0569	0,1249
4	7	9	0,8182	0,0000	0,5000	0,3182
5	2	11	1,0000	1,5811	0,9431	0,0569
6	0	11	1,0000	3,1623	0,9992	0,0008

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-3,7538	0,0001	0,0001
2	0	0	0,0000	-2,4218	0,0077	0,0077
3	4	4	0,3636	-1,0898	0,1379	0,2257
4	5	9	0,8182	0,2422	0,5957	0,2225
5	2	11	1,0000	1,5742	0,9423	0,0577
6	0	11	1,0000	2,9062	0,9982	0,0018

Agradável		
Participante	**Claire**	**Rachel**
1	2	3
2	4	2
3	4	5
4	5	4
5	5	3
6	2	3
7	4	3
8	4	5
9	4	4
10	3	3
11	5	5
Média	**3,8182**	**3,6364**
Desvio-padrão	**1,0787**	**1,0269**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-2,6125	0,0045	0,0045
2	2	2	0,1818	-1,6855	0,0459	0,1359
3	1	3	0,2727	-0,7585	0,2241	0,0486
4	5	8	0,7273	0,1685	0,5669	0,1603
5	3	11	1,0000	1,0956	0,8634	0,1366
6	0	11	1,0000	2,0226	0,9784	0,0216

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-2,5673	0,0051	0,0051
2	1	1	0,0909	-1,5935	0,0555	0,0354
3	5	6	0,5455	-0,6197	0,2677	0,2777
4	2	8	0,7273	0,3541	0,6384	0,0889
5	3	11	1,0000	1,3279	0,9079	0,0921
6	0	11	1,0000	2,3017	0,9893	0,0107

Gentil		
Participante	**Claire**	**Rachel**
1	5	4
2	4	4
3	4	5
4	5	4
5	5	5
6	3	3
7	5	4
8	3	3
9	5	5
10	3	4
11	5	5
Média	**4,2727**	**4,1818**
Desvio-padrão	**0,9045**	**0,7508**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-3,6181	0,0001	0,0001
2	0	0	0,0000	-2,5126	0,0060	0,0060
3	3	3	0,2727	-1,4071	0,0797	0,1930
4	2	5	0,4545	-0,3015	0,3815	0,0730
5	6	11	1,0000	0,8040	0,7893	0,2107
6	0	11	1,0000	1,9096	0,9719	0,0281

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-4,2381	0,0000	0,0000
2	0	0	0,0000	-2,9062	0,0018	0,0018
3	2	2	0,1818	-1,5742	0,0577	0,1241
4	5	7	0,6364	-0,2422	0,4043	0,2320
5	4	11	1,0000	1,0898	0,8621	0,1379
6	0	11	1,0000	2,4218	0,9923	0,0077

Competente		
Participante	**Claire**	**Rachel**
1	4	4
2	5	5
3	5	4
4	3	3
5	4	4
6	4	4
7	4	5
8	4	4
9	5	5
10	4	4
11	5	4
Média	**4,2727**	**4,1818**
Desvio-padrão	**0,6467**	**0,6030**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-5,0609	0,0000	0,0000
2	0	0	0,0000	-3,5145	0,0002	0,0002
3	1	1	0,0909	-1,9681	0,0245	0,0664
4	6	7	0,6364	-0,4217	0,3366	0,2998
5	4	11	1,0000	1,1246	0,8696	0,1304
6	0	11	1,0000	2,6710	0,9962	0,0038

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-5,2764	0,0000	0,0000
2	0	0	0,0000	-3,6181	0,0001	0,0001
3	2	2	0,1818	-1,9598	0,0250	0,1568
4	5	7	0,6364	-0,3015	0,3815	0,2549
5	4	11	1,0000	1,3568	0,9126	0,0874
6	0	11	1,0000	3,0151	0,9987	0,0013

APÊNDICE

B

TESTES DE NORMALIDADE - EXPERIMENTO B

Questão Q3

Participante	Claire	Rachel
1	2	2
2	3	1
3	5	4
4	3	4
5	4	3
6	3	4
7	3	3
8	4	4
9	4	3
10	3	3
Média	**3,4000**	**3,1000**
Std	**0,8433**	**0,9944**
	p-value	**0,1717**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	0	0	0,0000	-2,8460	0,0022	0,0022
2	1	1	0,0909	-1,6602	0,0484	0,0425
3	5	6	0,5455	-0,4743	0,3176	0,2278
4	3	9	0,8182	0,7115	0,7616	0,0566
5	1	10	0,9091	1,8974	0,9711	0,0620
6	0	10	0,9091	3,0832	0,9990	0,0899

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	1	1	0,0909	-2,1118	0,0174	0,0736
2	1	2	0,1818	-1,1062	0,1343	0,0475
3	4	6	0,5455	-0,1006	0,4599	0,0855
4	4	10	0,9091	0,9050	0,8173	0,0918
5	0	10	0,9091	1,9106	0,9720	0,0629
6	0	10	0,9091	2,9162	0,9982	0,0891

Questão Q4

Participante	Claire	Rachel
1	2	3
2	1	1
3	3	3
4	3	4
5	4	4
6	2	3
7	2	2
8	5	4
9	3	3
10	3	4
Média	**2,8000**	**3,1000**
Std	**1,1353**	**0,9944**
	p-value	**0,0967**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	1	1	0,0909	-1,5855	0,0564	0,0345
2	3	4	0,3636	-0,7047	0,2405	0,1231
3	4	8	0,7273	0,1762	0,5699	0,1574
4	1	9	0,8182	1,0570	0,8547	0,0366
5	1	10	0,9091	1,9378	0,9737	0,0646
6	0	10	0,9091	2,8187	0,9976	0,0885

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	1	1	0,0909	-2,1118	0,0174	0,0736
2	1	2	0,1818	-1,1062	0,1343	0,0475
3	4	6	0,5455	-0,1006	0,4599	0,0855
4	4	10	0,9091	0,9050	0,8173	0,0918
5	0	10	0,9091	1,9106	0,9720	0,0629
6	0	10	0,9091	2,9162	0,9982	0,0891

Questão Q5

Participante	Claire	Rachel
1	2	2
2	1	1
3	4	5
4	4	4
5	2	2
6	3	4
7	2	3
8	3	4
9	4	3
10	2	2
Média	**2,7000**	**3,0000**
Std	**1,0593**	**1,2472**
	p-value	**0,0967**

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	1	1	0,0909	-1,6048	0,0543	0,0366
2	4	5	0,4545	-0,6608	0,2544	0,2002
3	2	7	0,6364	0,2832	0,6115	0,0249
4	3	10	0,9091	1,2272	0,8901	0,0190
5	0	10	0,9091	2,1711	0,9850	0,0759
6	0	10	0,9091	3,1151	0,9991	0,0900

Nota	Freqüência	Cumulativo	S_n(x)	Z-Score	F(x)	Diferença
1	1	1	0,0909	-1,6036	0,0544	0,0365
2	3	4	0,3636	-0,8018	0,2113	0,1523
3	2	6	0,5455	0,0000	0,5000	0,0455
4	3	9	0,8182	0,8018	0,7887	0,0295
5	1	10	0,9091	1,6036	0,9456	0,0365
6	0	10	0,9091	2,4054	0,9919	0,0828

APÊNDICE

C

QUESTIONÁRIO - EXPERIMENTO A

Questionário – experimento na UNESP

1. Suas impressões quanto às características humanas do agente, nas seguintes características:
    a. Realista: (1) (2) (3) (4) (5)
    b. Autoconsciente: (1) (2) (3) (4) (5)

2. Sua impressão sobre a expressão de vida do agente nas seguintes características:
    a. Interativo: (1) (2) (3) (4) (5)
    b. Atento: (1) (2) (3) (4) (5)

3. Sua impressão sobre a simpatia do agente, nas seguintes características:
    a. Amigável: (1) (2) (3) (4) (5)
    b. Gentil: (1) (2) (3) (4) (5)
    c. Agradável: (1) (2) (3) (4) (5)

4. Sua impressão sobre a inteligência percebida do agente nas seguintes características:
    a. Competente: (1) (2) (3) (4) (5)
    b. Sensato: (1) (2) (3) (4) (5)

5. Você sentiu empatia diante do agente?
    (1) (2) (3) (4) (5)

---

1. Suas impressões quanto às características humanas do agente, nas seguintes características:
    a. Realista: (1) (2) (3) (4) (5)
    b. Autoconsciente: (1) (2) (3) (4) (5)

2. Sua impressão sobre a expressão de vida do agente nas seguintes características:
    a. Interativo: (1) (2) (3) (4) (5)
    b. Atento: (1) (2) (3) (4) (5)

3. Sua impressão sobre a simpatia do agente, nas seguintes características:
    a. Amigável: (1) (2) (3) (4) (5)
    b. Gentil: (1) (2) (3) (4) (5)
    c. Agradável: (1) (2) (3) (4) (5)

4. Sua impressão sobre a inteligência percebida do agente nas seguintes características:
    a. Competente: (1) (2) (3) (4) (5)
    b. Sensato: (1) (2) (3) (4) (5)

5. Você sentiu empatia diante do agente?
    (1) (2) (3) (4) (5)

APÊNDICE

D

QUESTIONÁRIO - EXPERIMENTO B

Questionário para pesquisa – ICMC/USP

1. Que emoções você experimentou durante o jogo?
   (a) Nenhuma (b) Alegria (c) Tristeza
   (d) Aversão (e) Medo (f) Raiva
   (g) Surpresa
2. Você acredita que a personagem reconheceu como você se sentiu durante o jogo?
   (a) Sim (b) Não
3. Em uma escala de **1 a 5**, o quão corretamente a personagem reagiu às suas emoções?
   (1) (2) (3) (4) (5)
4. Em uma escala de **1 a 5**, o quanto você acredita que as suas emoções influenciaram as emoções da personagem?
   (1) (2) (3) (4) (5)
5. Em uma escala de **1 a 5,** quanta **empatia**[1] você acredita ter sentido diante da personagem?
   (1) (2) (3) (4) (5)

---

1. Que emoções você experimentou durante o jogo?
   (b) Nenhuma (b) Alegria (c) Tristeza
   (e) Aversão (e) Medo (f) Raiva
   (h) Surpresa
2. Você acredita que a personagem reconheceu como você se sentiu durante o jogo?
   (a) Sim (b) Não
3. Em uma escala de **1 a 5**, o quão corretamente a personagem reagiu às suas emoções?
   (1) (2) (3) (4) (5)
4. Em uma escala de **1 a 5**, o quanto você acredita que as suas emoções influenciaram as emoções da personagem?
   (1) (2) (3) (4) (5)
5. Em uma escala de **1 a 5,** quanta **empatia**[1] você acredita ter sentido diante da personagem?
   (1) (2) (3) (4) (5)

---

[1]**Empatia:** forma de identificação intelectual ou afetiva de um sujeito com uma pessoa, uma ideia ou uma coisa (*ex.: a empatia entre os voluntários e a população local era evidente*; *assistimos à perfeita empatia entre piano e violino*).

Dicionário Priberam da Língua Portuguesa, 2008-2013, https://www.priberam.pt/DLPO/empatia [consultado em 01-03-2016].

Participante:

Ordem: