



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Programa de Pós-Graduação em Engenharia Elétrica
Grupo de pesquisa MACRO - Mecatrônica, Controle e Robótica

Ana Christina Almada Campos

**INTERAÇÕES ENTRE HUMANOS E AGENTES VIRTUAIS: efeitos da combinação
de comunicações explícitas e implícitas**

Belo Horizonte
2022

Ana Christina Almada Campos

**INTERAÇÕES ENTRE HUMANOS E AGENTES VIRTUAIS: efeitos da combinação
de comunicações explícitas e implícitas**

Versão final

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Bruno Vilhena Adorno

Belo Horizonte
2022

C198i

Campos, Ana Christina Almada.

Interações entre humanos e agentes virtuais [recurso eletrônico] : efeitos da combinação de comunicações explícitas e implícitas / Ana Christina Almada Campos. - 2022.

1 recurso online (155 f. : il., color.) : pdf.

Orientador: Bruno Vilhena Adorno.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Apêndices: f. 129-155.

Bibliografia: f. 114-128.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Interação homem-máquina - Teses. 3. Robôs - Teses. 4. Sistemas de comunicação - Teses. I. Adorno, Bruno Vilhena. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 621.3(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

FOLHA DE APROVAÇÃO

"INTERAÇÃO ENTRE HUMANOS E AGENTES VIRTUAIS: EFEITOS DA COMBINAÇÃO DE COMUNICAÇÕES EXPLÍCITAS E IMPLÍCITAS"

ANA CHRISTINA ALMADA CAMPOS

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica. Aprovada em 21 de julho de 2022. Por:

Prof. Dr. Bruno Vilhena Adorno
DEE (UFMG) - Orientador

Prof. Dr. Felipe Campelo França Pinto
DEE (UFMG)

Dr. Luis Felipe da Cruz Figueredo
Munich Institute of Robotics and Machine Intelligence (MIRMI)



Documento assinado eletronicamente por **Bruno Vilhena Adorno, Professor do Magistério Superior**, em 22/07/2022, às 05:05, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1616943** e o código CRC **21F90311**.

*À minha mãe, Marília, por ter
quase tanto orgulho de mim quanto
eu tenho dela.*

AGRADECIMENTOS

Parece estranho começar a escrever em primeira pessoa depois de evitar essa conjugação ao longo de todo o texto. É mais engraçado ainda porque este trabalho foi feito definitivamente em primeira pessoa, porém do plural. “Eu”, no singular, não chegaria nem no título dessa dissertação.

Incluído nesse plural está, obviamente, o Prof. Bruno. Apesar de eu não o tratar usando esse termo, “professor”, esse é precisamente o papel que ele desempenha o tempo todo. Desde a sala de aula, onde eu ficava constantemente fascinada pelos desenvolvimentos matemáticos (que nós tínhamos que aprender para se um dia fôssemos a última pessoa a ter aquele conhecimento e precisássemos registrá-lo nas paredes de uma caverna com nosso próprio sangue), até em cada uma das reuniões que tive com ele ao longo desses anos, eu aprendi muito além de robótica e de pesquisa. Minha primeira reunião como integrante do grupo foi a primeira vez na minha trajetória na engenharia em que eu entrei em uma sala com metade de alunas mulheres e eu não tenho como explicar a importância disso para mim. Bruno acreditou e confiou em mim, sempre fez eu me sentir à vontade para dar ideias e admitir quando errava ou não sabia alguma coisa e todas as vezes que eu achei que estava dando tudo errado neste trabalho, fazia com que eu voltasse a pensar com clareza. Eu sou muito grata por todo aprendizado e inspiração.

Entretanto, esse “nós” não está restrito dentro de paredes de universidades. Obrigada aos meus pais, Marília e Hideraldo, à minha irmã, Márcia Helenna, e ao meu irmão, Luiz Henrique, por me acompanharem e permanecerem ao meu lado, fazendo com que meus dias sejam mais do que olhos fixos em uma tela de computador e dedos nervosos digitando incessantemente. O apoio de vocês, traduzido em companhias para café e filmes, orgulho e preocupação e fazendo questão de saber dizer exatamente o que eu faço, mesmo que ninguém vá entender, é essencial. Devo dizer que desde em detalhes dos experimentos até em figuras deste documento existe um dedo (ou mais) do meu irmão, a primeira pessoa a quem eu recorria quando precisava de opiniões e ideias nos mais diversos assuntos, então obrigada.

Obrigada ao Franco, por nunca parar de me incentivar, por me trazer segurança, ser alguém com quem eu posso contar e por ter se habituado a perguntar se eu tinha muita coisa para fazer e me ouvir repetir todas as vezes que “eu sempre tenho muita coisa para

fazer”. Às minhas amigas maravilhosas de tantos anos, Sara e Melissa, obrigada por me colocarem para cima quando eu preciso e por me ouvirem reclamar e me acolherem todos os dias. A todas as outras pessoas que eu sei que torcem por mim (e as que eu não sei também), muito obrigada!

Aos colegas do MACRO, especialmente aos alunos do Prof. Bruno, obrigada pelas conversas, às vezes complexas e filosóficas, durante o café, por toda a ajuda para resolver desde problemas no laboratório até no Lyx e pela companhia, mesmo que virtual, nesse dia a dia louco de pós-graduação. Um agradecimento especial para a Vera, que por muitos meses, nos períodos de restrição na UFMG, era a única pessoa com quem eu conversava o dia todo, além de cuidar bem dos laboratórios, sempre se preocupar se eu precisava de alguma coisa e sentir minha falta quando eu não aparecia.

Por fim, eu preciso agradecer a todas as pessoas que participaram de experimentos deste trabalho. Durante meses, era comum me ver andando exaustivamente pela Escola de Engenharia, entrando em todo laboratório, secretaria, sala de faxina ou qualquer lugar que estivesse aberto para convidar as pessoas a participarem. Fazer esses experimentos foi desesperador, mas ao mesmo tempo fascinante. Obrigada a quem ouviu com atenção meu convite, por quem se interessou e por quem se dispôs a ajudar em qualquer fase desta pesquisa, virtual ou presencialmente.

Tem muitas mãos nesse “nós”. Muito obrigada a cada uma delas!

RESUMO

Com a crescente aplicação de robôs e agentes virtuais em contato próximo com humanos, a área de interação humano-robô se torna cada vez mais importante. A comunicação entre humano e robô (ou agente virtual), essencial para a interação, muitas vezes é inspirada na comunicação humana, que utiliza gestos, expressões faciais, olhar e outros meios explícitos e implícitos. É importante investigar os efeitos de diferentes tipos de comunicação na interação. O objetivo deste trabalho é comparar a utilização apenas de comunicações explícitas com a combinação de comunicações explícitas e implícitas de humano e agente virtual, observando medidas de desempenho e a percepção das pessoas, enriquecendo a literatura sobre comunicação humano-robô. Foi desenvolvida uma infraestrutura de comunicação multimodal contendo sistemas de reconhecimento de gestos e direção do olhar do humano e uma agente virtual fazendo o papel de robô. Um experimento de interação semelhante a um jogo foi proposto e as variáveis observadas foram tempo e número de erros na execução da tarefa, aceitação, sociabilidade e transparência da agente virtual e percepção de eficiência da interação. Os resultados foram avaliados com testes de hipóteses frequentistas e com estimação de parâmetros Bayesiana. As variáveis mais relacionadas com a tarefa (tempo, número de erros e percepção de eficiência da interação) não parecem ter sido influenciadas pelo tipo de comunicação, o que acredita-se que pode ser devido a características específicas da interação proposta. Entretanto, observou-se que a combinação de comunicações explícitas e implícitas na interação pode melhorar a aceitação, a sociabilidade e a transparência da agente virtual. Melhorias no protocolo experimental podem ajudar a investigar medidas mais relacionadas à tarefa e as informações obtidas com a estimação Bayesiana de parâmetros associados à interação podem ser incorporadas em estudos futuros.

Palavras-chave: Interação humano-robô. Comunicação humano-robô. Comunicações explícitas e implícitas. Agente virtual.

ABSTRACT

The field of human-robot interaction has become more important with the increasing application of robots and virtual agents interacting with humans. Communication between humans and robots (or virtual agents) is essential for interaction and is often inspired by human communication, which uses gestures, facial expressions, gaze direction, and other explicit and implicit means. This work aims to compare the use of only explicit communication with the combination of explicit and implicit communication, and observe performance measures and people's perceptions, contributing to the human-robot communication literature. A multimodal communication infrastructure with systems for human gestures and gaze direction recognition and a virtual agent playing the role of a robot was developed. An interaction experiment similar to a game was proposed and the observed variables were time and number of errors in the task execution, acceptance of the virtual agent, the virtual agent's sociability and transparency, and perception of efficiency of the interaction. The results were evaluated with frequentist hypothesis tests and Bayesian parameter estimation. Measures more related to the task (time, number of errors and perception of efficiency of the interaction) appear not to have been influenced by the type of communication, which can be attributed to the specific characteristics of the proposed interaction. However, it was observed that the combination of explicit and implicit communications can improve people's acceptance of the virtual agent and its sociability and transparency. Improvements in the experimental protocol could help in the investigation of variables more related to the task, and information obtained with the Bayesian estimation of parameters associated with the interaction can be incorporated in future studies.

Keywords: Human-robot interaction. Human-robot communication. Explicit and implicit communications. Virtual agent.

LISTA DE FIGURAS

1.1	Exemplos de robôs da ficção.	19
1.2	Exemplos de robôs reais.	20
2.1	Exemplos de sistemas que utilizam comunicações naturais.	28
3.1	Transformações sucessivas entre sistemas de coordenadas.	36
3.2	Primitivas geométricas.	37
3.3	Juntas da cadeia cinemática do humano rastreadas.	38
3.4	Representação de gestos de apontar.	39
3.5	Representação de objetos e regiões de interesse.	39
3.6	Cruzamento entre linha do gesto de apontar e objeto.	40
3.7	Distâncias entre planos do objeto e ponto de cruzamento com linha.	41
3.8	Linha do gesto cruzando nenhum ou mais de um objeto ao mesmo tempo.	42
3.9	Exemplo de detecção de rosto, olhos e boca.	43
3.10	Gradientes de cor na imagem do olho.	44
3.11	Determinação do centro da íris do olho.	44
3.12	Valores cinza em imagem original e invertida.	45
3.13	Exemplo de funcionamento do algoritmo de detecção dos pontos no rosto.	46
3.14	Exemplos de variações nos atributos do rosto com a movimentação da cabeça.	47
3.15	Distância entre os olhos usada para determinar ângulo de rotação.	48
3.16	Distância entre olhos e boca usada para determinar ângulo de rotação.	48
3.17	Expressões faciais da agente virtual.	51
3.18	Diferentes formatos de boca para a agente virtual.	51
3.19	Ângulos de rotação para olhar da agente virtual.	52
3.20	Determinação de pontos para transformar imagem da agente virtual.	53
3.21	Exemplo de transformação de imagem da agente virtual.	53
3.22	Outras opções de olhos para a agente virtual.	54
4.1	Agentes virtuais para os experimentos.	58
4.2	Fluxo da interação entre humano e uma das agentes virtuais.	59
4.3	Nuvem de possibilidades da estatística de teste sob a hipótese nula.	63

4.4	Distribuições de probabilidade <i>a priori</i> e <i>a posteriori</i> na estimação Bayesiana.	67
4.5	Exemplos de situações entre o HDI e a ROPE e a decisão tomada.	68
4.6	Exemplo de distribuição da estatística do teste <i>t</i> sob a hipótese nula.	70
4.7	Exemplos de distribuições <i>t</i> .	74
4.8	Diagrama da estimação Bayesiana das variáveis métricas.	75
4.9	Modelo ordinal para análise Bayesiana das variáveis de percepção.	76
4.10	Respostas ordinais e distribuições latentes que se adequariam a elas.	77
4.11	Diagrama da estimação Bayesiana das variáveis ordinais.	79
5.1	Resumo do perfil de participantes do experimento de interação.	84
5.2	Diagramas de caixas do conjunto de dados de cada variável.	87
5.3	Exemplo de distribuição <i>a posteriori</i> para o tamanho de efeito com ROPE.	91
5.4	Resultados da inferência Bayesiana para a diferença de tempo.	92
5.5	Resultados da inferência Bayesiana para a diferença de erros.	94
5.6	Histogramas das respostas nas escalas Likert de variáveis de percepção.	95
5.7	Distribuições <i>a posteriori</i> de parâmetros de aceitação das agentes virtuais.	96
5.8	Distribuições <i>a posteriori</i> dos limites do item 3 da escala de aceitação.	98
5.9	Distribuições <i>a posteriori</i> dos limites dos itens de aceitação.	99
5.10	Checagem posterior de adequação do modelo para aceitação.	100
5.11	Dados com perfis de idade, gênero e escolaridade.	107
5.12	Dados com perfis de preferência, primeira condição e familiaridade com AVs.	108
A.1	Ambiente para experimentos com o sistema de gestos de apontar.	130
A.2	Exemplo de um gesto de apontar.	130
A.3	Acertos e erros do sistema sobre os gestos de cada participante.	131
A.4	Exemplos de erros na detecção de atributos faciais.	133
A.5	Imagens usadas para avaliação do reconhecimento das expressões faciais.	133
A.6	Resultados da avaliação do reconhecimento das expressões faciais.	134
A.7	Reconhecimento de expressões faciais por idade, gênero e escolaridade.	136
A.8	Reconhecimento de expressões faciais por familiaridade e preferência.	137
A.9	Respostas sobre as expressões faciais para grupo com interação.	138
A.10	Respostas sobre as expressões faciais para grupo sem interação.	138
B.1	Fluxo de participação no experimento de interação.	140
B.2	Contextualização da interação e instruções iniciais.	141
B.3	Fluxo de telas durante a interação.	143
B.4	Ambiente dos experimentos de interação.	144
B.5	Visão de participantes durante a fase 2 e imagens para contagem.	144
B.6	Conjuntos de objetos para contagem.	144
C.1	Distribuições <i>a posteriori</i> de parâmetros de sociabilidade das agentes virtuais.	147

C.2	Distribuições <i>a posteriori</i> dos limites dos itens de sociabilidade.	148
C.3	Checagem posterior de adequação do modelo para sociabilidade.	149
C.4	Distribuições <i>a posteriori</i> de parâmetros de transparência das agentes virtuais.	150
C.5	Distribuições <i>a posteriori</i> dos limites dos itens de transparência.	151
C.6	Checagem posterior de adequação do modelo para transparência.	152
C.7	Distribuições <i>a posteriori</i> de parâmetros de percepção de eficiência.	153
C.8	Distribuições <i>a posteriori</i> dos limites dos itens de percepção de eficiência. .	154
C.9	Checagem posterior de adequação do modelo para percepção de eficiência.	155

LISTA DE TABELAS

2.1	Exemplos de comunicações explícitas e implícitas.	25
3.1	Tipos de comunicações disponíveis para humano e agente virtual.	55
4.1	Itens das escalas Likert para medição de variáveis de percepção.	61
4.2	Possibilidades para a estatística do teste de Wilcoxon sob a hipótese nula. .	71
4.3	Vantagens e desvantagens de métodos para análise de dados com níveis vazios.	80
4.4	Resumo das análises estatísticas para cada variável de interesse no trabalho.	82
5.1	Descrição de desvios do protocolo experimental e suas resoluções.	85
5.2	Resultados de NHST para a diferença de tempo e número de erros.	89
5.3	Resultados de NHST para a diferença das variáveis de percepção.	90
5.4	Resultados da estimação Bayesiana de parâmetros das medidas subjetivas.	102
5.5	Resumo das análises para as variáveis de interesse do trabalho.	103

LISTA DE ABREVIATURAS

AV	Agente virtual
EX	Configuração de comunicação para interação apenas com comunicações explícitas
EXIM	Configuração de comunicação para interação com comunicações explícitas e implícitas
HDI	<i>Highest Density Interval</i> (intervalo de maior densidade)
IHR	Interação Humano-Robô
JAGS	<i>Just Another Gibbs Sampler</i>
MACRO	<i>Mechatronics, Control and Robotics</i>
MCMC	<i>Markov Chain Monte Carlo</i> (Monte Carlo via cadeias de Markov)
NHST	<i>Null Hypothesis Significance Testing</i> (teste de significância de hipótese nula)
ROPE	<i>Region of Practical Equivalence</i> (região de equivalência prática)
ROS	<i>Robot Operating System</i>

LISTA DE SÍMBOLOS E NOTAÇÕES

CAPÍTULO 3 E APÊNDICE A

\mathcal{F}	Sistema de coordenadas
\mathbf{h}^*	Conjugado do quatérnio \mathbf{h}
\mathbb{H}	Conjunto dos quatérnios
\mathbb{H}_p	Conjunto dos quatérnios puros
\mathcal{H}	Conjunto dos quatérnios duais
\mathbf{I}	Matriz identidade
$\text{Im}(\cdot)$	Parte imaginária
$\hat{i}, \hat{j}, \hat{k}$	Unidades imaginárias
\mathbb{R}	Conjunto dos números reais
$\text{Re}(\cdot)$	Parte real
\mathbb{S}^3	Conjunto dos quatérnios unitários
ε	Unidade dual
$\mathbf{0}$	Matriz de zeros
$\ \cdot\ $	Norma
$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle$	Produto interno entre \mathbf{h}_1 e \mathbf{h}_2
$\mathbf{h}_1 \times \mathbf{h}_2$	Produto externo entre \mathbf{h}_1 e \mathbf{h}_2

Letras simples minúsculas ou maiúsculas, como d ou D , representam escalares, exceto x , y e z , que são reservadas para rótulos de eixos de sistemas de coordenadas. Letras minúsculas em negrito representam vetores $\mathbf{v} \in \mathbb{R}^2$ ou quatérnios $\mathbf{h} \in \mathbb{H}$. Letras maiúsculas em negrito indicam matrizes $\mathbf{M} \in \mathbb{R}^{m \times n}$. Letras minúsculas em negrito e sublinhadas representam quatérnios duais $\underline{\mathbf{h}} \in \mathcal{H}$. Variáveis com circunflexo representam vetores normalizados $\hat{\mathbf{v}} = \mathbf{v} / \|\mathbf{v}\|$. Sobrescritos indicam o sistema de coordenadas de referência, por exemplo, \mathbf{p}^a é o ponto \mathbf{p} em relação ao sistema de coordenadas \mathcal{F}_a .

CAPÍTULOS 4 E 5 E APÊNDICE C

d	Tamanho de efeito
s	Desvio padrão amostral
x	Uma única observação do conjunto de dados
\bar{x}	Média amostral
\tilde{x}	Mediana amostral
θ_k	k -ésimo limite do modelo ordinal para estimação Bayesiana
μ	Parâmetro de localização de distribuição de probabilidade (média)
$\tilde{\mu}$	Parâmetro de localização de distribuição de probabilidade (mediana)
ν	Parâmetro de normalidade da distribuição t
σ	Parâmetro de escala de distribuição de probabilidade
$\Phi(\cdot)$	Função normal padrão acumulada
$\Psi(\cdot)$	Função t acumulada
$P(A)$	Probabilidade do evento A
$P(A B)$	Probabilidade do evento A condicionada ao evento B

Outras letras simples, minúsculas ou maiúsculas, representam escalares.

SUMÁRIO

1	INTRODUÇÃO	18
2	REVISÃO BIBLIOGRÁFICA E ESTADO DA ARTE	23
2.1	Comunicação humano-robô	24
2.1.1	Tecnologias de comunicação	26
2.1.2	Experimentos de interação e estudos comparativos	29
2.2	Conclusão	32
3	INFRAESTRUTURA DE COMUNICAÇÃO HUMANO-ROBÔ	34
3.1	Fundamentos matemáticos	35
3.1.1	Quatérnios	35
3.1.2	Quatérnios duais e primitivas geométricas	36
3.2	Gestos de apontar	37
3.3	Direção do olhar	42
3.3.1	Detecção de pontos no rosto	43
3.3.2	Determinação da direção do olhar e interpretação	46
3.4	Agente virtual	49
3.4.1	Voz	50
3.4.2	Rosto e expressões faciais	50
3.4.3	Direção do olhar	52
3.5	Outras comunicações	54
3.6	Conclusão	55
4	PLANEJAMENTO EXPERIMENTAL	56
4.1	Hipóteses	56
4.2	Interação proposta	57
4.3	Medidas e instrumentos de medição	60
4.4	Ferramentas de análises estatísticas	62
4.4.1	Testes de significância de hipótese nula	69
4.4.1.1	Medidas objetivas	69

4.4.1.2	Medidas subjetivas	72
4.4.2	Estimação de parâmetros Bayesiana	73
4.4.2.1	Medidas objetivas	73
4.4.2.2	Medidas subjetivas	76
4.5	Conclusão	82
5	RESULTADOS E AVALIAÇÃO DOS EXPERIMENTOS DE INTERAÇÃO	83
5.1	Resultados das análises frequentistas	87
5.2	Resultados das análises Bayesianas	91
5.2.1	Medidas objetivas	92
5.2.2	Medidas subjetivas	95
5.3	Discussão dos experimentos	103
5.4	Influência de fatores na diferença entre as configurações de comunicação . .	105
5.5	Conclusão	109
6	CONCLUSÃO E TRABALHOS FUTUROS	110
	REFERÊNCIAS BIBLIOGRÁFICAS	114
A	AVALIAÇÕES DOS SISTEMAS DE COMUNICAÇÃO	129
A.1	Gestos de apontar	129
A.2	Detecção de atributos faciais	132
A.3	Expressões faciais da agente virtual	133
B	DETALHES DA INTERAÇÃO PROPOSTA	140
C	DISTRIBUIÇÕES DE PROBABILIDADE E HISTOGRAMAS DA ESTIMAÇÃO DE PARÂMETROS BAYESIANA	146

1

INTRODUÇÃO

O ser humano há muito tempo idealiza robôs, procurando soluções que facilitem seu dia a dia e possibilitem a realização de tarefas antes difíceis ou impossíveis. Muitas vezes essas máquinas são inspiradas nos próprios humanos, recriando suas relações e imitando suas emoções e formas de interagir. Na ficção (ver Figura 1.1), os robôs C-3PO, tradutor e facilitador de interações sociais, e R2-D2, responsável pela manutenção e navegação de astronaves, da série *Star Wars* (1977-2019) e a inteligência artificial J.A.R.V.I.S (*Just A Rather Very Intelligent System*), que opera os sistemas internos da casa e o traje do Homem de Ferro na série de filmes sobre o herói (2008–2013), são exemplos de sistemas funcionais e prestativos. Além de todas as capacidades físicas e intelectuais, muitos robôs apresentam emoções e diferentes personalidades, como o carismático Bumblebee do filme homônimo (2018) e o frustrado Marvin da coleção de livros de Douglas Adams (1979) e do filme (2005) *O Guia do Mochileiro das Galáxias*, que possui cérebro do tamanho de um planeta, mas não tem a chance de usá-lo em tarefas além de abrir portas ou acompanhar visitantes. Os robôs ficcionais são criados para despertarem diferentes sentimentos nas pessoas, como o medo das Sentinelas na coletânea de filmes *Matrix* (1999–2021) e a afeição ao assistente pessoal de saúde Baymax, em *Operação Big Hero* (2014).

Na vida real, os robôs também assumem diversas funções e características. Máquinas como manipuladores robóticos e os robôs móveis de exploração são criadas para suprir necessidades como força e repetibilidade e serem usadas em ambientes desconhecidos e perigosos. As aplicações industriais atualmente incluem desde auxílio em tarefas de montagem e transporte de materiais (Unhelkar et al., 2018) até a realização de trabalhos colabora-

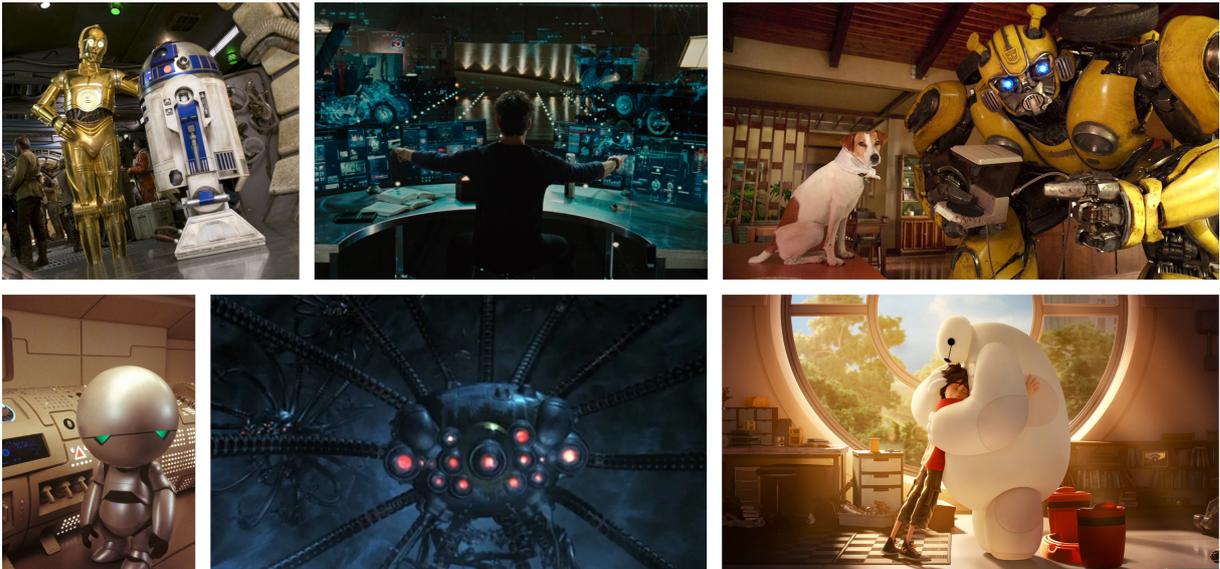


Figure 1.1: Exemplos de robôs da ficção¹. Em sentido *horário*: C-3PO e R2-D2, Tony Stark interagindo com J.A.R.V.I.S, Bumblebee, Baymax, Sentinela da Matrix e Marvin.

tivos de interação física com humanos (Ajoudani et al., 2018). Para o desenvolvimento desses robôs, não só medidas de desempenho, como tempo de execução e minimização de tempo ocioso, são levadas em consideração, mas também aspectos relacionados ao humano que participa da interação, buscando respeitar suas preferências, não sobrecarregá-lo e proporcionando satisfação com a colaboração (Shah et al., 2011; Gombolay et al., 2015; Gombolay et al., 2017). Já em ambientes menos controlados que os industriais e com maior proximidade com humanos, robôs podem, por exemplo, realizar tarefas domésticas (Asfour et al., 2006), transportar materiais dentro de hospitais (Ljungblad et al., 2012) e proporcionar auxílio físico a pessoas com deficiência (Lana et al., 2013).

De acordo com Breazeal (2003), robôs autônomos equipados com capacidades de comunicação, cooperação e aprendizado tendem a ser antropomorfizados, ou seja, as pessoas atribuem características humanas a eles. Com isso, as pessoas tendem a se basear na interação social entre humanos para criar o modelo mental de interação com esses robôs. Robôs que motivam a aplicação de modelos sociais de interação são chamados de robôs sociais (Breazeal, 2003) e aqueles em que a interação social tem um papel principal são conhecidos como robôs socialmente interativos (Fong et al., 2003). Um exemplo de robô socialmente interativo é o humanoide Pepper (Figura 1.2a), desenvolvido pela empresa SoftBank Robotics. Pepper² foi criado para interagir com pessoas, possui uma base móvel e é equipado com sensores que permitem navegação autônoma, percepção do ambiente e

¹Fontes: C-3PO e R2-D2 - <https://www.starwars.com/databank/r2-d2-biography-gallery>; J.A.R.V.I.S. - https://marvelcinematicuniverse.fandom.com/wiki/Iron_Man/Gallery; Sentinela - [https://villains.fandom.com/wiki/Machines_\(The_Matrix\)](https://villains.fandom.com/wiki/Machines_(The_Matrix)); Baymax - https://disney.fandom.com/wiki/Big_Hero_6/Gallery; Bumblebee - <https://www.imdb.com/title/tt4701182/>; Marvin - <https://www.imdb.com/title/tt0371724/>.

²<https://www.softbankrobotics.com/emea/en/pepper>

das pessoas ao seu redor, é capaz de reconhecer rostos e emoções básicas, conversar em diferentes idiomas, realizar movimentos naturais e expressivos e possui uma tela sensível a toque, o que possibilita diferentes vias de comunicação entre humano e robô. Já Paro (Figura 1.2b) é um robô terapêutico que imita uma foca. Criado pela japonesa AIST, Paro³ é equipado com diferentes sensores que o permitem perceber quando alguém o bate ou acaricia, além de ser capaz de transmitir vivacidade e fazer sons que imitam uma foca bebê real. Buscando proporcionar os benefícios obtidos com terapia animal, Paro pode ser usado em aplicações como cuidado de idosos (Abdi et al., 2018) e de auxílio à saúde mental de crianças e adultos (Crossman et al., 2018; Scoglio et al., 2019) e é um exemplo dos chamados robôs de assistência social, *i.e.*, aqueles que dão assistência a seres humanos através de interações sociais (Feil-Seifer & Mataric, 2005). O robô social jibo (Figura 1.2c), criado por Cynthia Breazeal, inicialmente era um assistente pessoal com funcionalidades parecidas com as dos atuais *smart speakers*, porém capaz de estimular interações sociais. Recentemente adquirido pela NTT Disruption, o foco de jibo⁴ passou a ser no auxílio à educação e saúde. O robô é equipado com habilidades de interação focadas em fatores humanos, como, por exemplo, um sistema motor que o permite expressar diferentes emoções humanas através de movimentos. As pesquisas sobre as aplicações de robôs socialmente interativos incluem robôs recepcionistas, que fornecem informações úteis e conversam com visitantes (Gockley et al., 2005), robôs educacionais para crianças (Toh et al., 2016), robôs de companhia para utilização em hospitais infantis (Meghdari et al., 2018) e robôs educacionais e de terapia para crianças com autismo, possibilitando ambientes seguros e sem julgamentos para o desenvolvimento de habilidades (Robins et al., 2018).

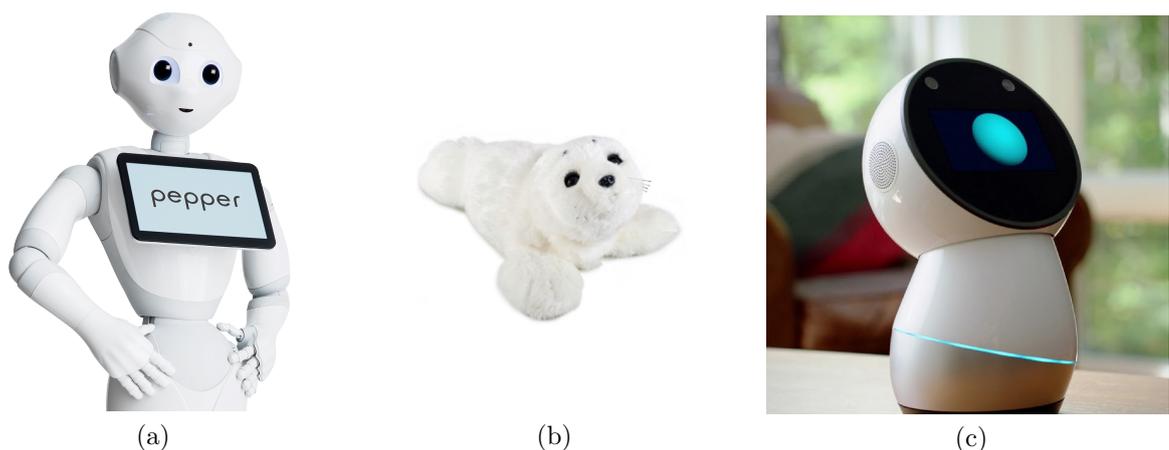


Figure 1.2: Exemplos de robôs reais⁵: (a) Pepper. (b) Paro. (c) jibo.

Tanto para aplicações com robôs colaborativos, onde a troca de informações entre

³<http://www.parorobots.com/>

⁴<https://jibo.com/>

⁵Fontes: (a) <https://www.softbankrobotics.com/emea/en> (b) <https://robots.ieee.org/robots/paro/> (c) <https://disruption.global.ntt/>

os agentes é essencial para que a tarefa seja realizada, quanto para o desenvolvimento de robôs sociais que têm por principal objetivo interagir de maneira natural com o ser humano, expressando emoções e personalidades, o desenvolvimento da comunicação entre humano e robô é de extrema importância. Além de desenvolver técnicas para que o robô obtenha informações do ser humano e se comunique de forma clara, é necessário definir quais meios de comunicação melhor se aplicam a cada contexto, focando tanto em métricas que avaliem a eficiência da comunicação quanto nas percepções do humano durante a interação com o robô, para que os objetivos da interação sejam cumpridos de maneira satisfatória e agradável para as pessoas envolvidas. Não há interação sem comunicação (Goodrich & Schultz, 2007), tornando esta uma importante área de estudo dentro da interação humano-robô.

Objetivos e contribuições

Para que seja possível desenvolver sistemas capazes de interagir com seres humanos em diferentes áreas de aplicação, é necessário estudar cada aspecto relacionado a esse tipo de interação. Especificamente, a motivação para este trabalho é investigar como o tipo de comunicação afeta a interação das pessoas com agentes virtuais.

O objetivo geral é comparar a utilização apenas de comunicações explícitas com a combinação de comunicações explícitas e implícitas e observar os efeitos em medidas relacionadas aos objetivos da interação e à percepção das pessoas. O trabalho se estende desde a implementação de sistemas de comunicação humano-robô, o planejamento e a realização de experimentos de interação, até a análise estatística dos resultados. Com isso, busca-se enriquecer a literatura e obter intuições sobre aspectos relacionados à comunicação humano-robô, incluindo os procedimentos experimentais e de análise de resultados.

Na implementação da infraestrutura de comunicação para os experimentos de interação, este trabalho propõe:

1. A identificação de gestos de apontar a partir da cadeia cinemática do humano;
2. A estimação da direção do olhar do humano utilizando atributos faciais na imagem do rosto e geometria projetiva;
3. A utilização de primitivas geométricas descritas em quatérnios duais para representação e interpretação de gestos de apontar e direção do olhar do humano;
4. Uma agente virtual com voz, expressões faciais e direção do olhar simulada com geometria projetiva.

Em relação aos experimentos, as contribuições deste trabalho são:

1. Protocolo experimental de um experimento piloto de interação entre humanos e agentes virtuais, para comparação de diferentes combinações de comunicações explícitas e implícitas;
2. Estimacão de parâmetros relacionados a medidas objetivas (tempo e número de erros na execução das tarefas) e subjetivas (aceitação, sociabilidade e transparência da agente virtual e percepção de eficiência da interação) na interação entre humanos e agentes virtuais. Os resultados podem ser utilizados como informações prévias em trabalhos futuros.

Por fim, este trabalho contribui para o estado da arte ao trazer para o contexto de interação humano-robô a estimacão de parâmetros Bayesiana como alternativa aos testes de hipóteses frequentistas, comumente encontrados na literatura.

Estrutura da dissertacão

O Capítulo 2 apresenta uma revisão da literatura sobre interação e comunicacão humano-robô, especialmente sobre tecnologias de comunicacão natural e experimentos de interação. O Capítulo 3 descreve a infraestrutura de comunicacão humano-robô desenvolvida para a realizacão deste trabalho. O planejamento experimental, incluindo as hipóteses de trabalho, a interação proposta e as ferramentas de análises estatísticas das medidas de interesse, é apresentado no Capítulo 4 e, em seguida, os resultados são discutidos no Capítulo 5. O Capítulo 6 apresenta as conclusões e trabalhos futuros. Nos apêndices, encontram-se avaliações de alguns dos sistemas de comunicacão implementados (Apêndice A) e detalhes sobre o experimento de interação (Apêndice B), além de algumas figuras de uma das análises estatísticas aplicadas às medidas subjetivas analisadas (Apêndice C).

2

REVISÃO BIBLIOGRÁFICA E
ESTADO DA ARTE

Goodrich & Schultz (2007) definem interação humano-robô (IHR) como a área responsável pelo estudo de todos os aspectos relacionados à interação entre humanos e sistemas robóticos, desde o desenvolvimento de tecnologias até a criação de técnicas de avaliação. De acordo com o trabalho, todo sistema robótico precisa de algum nível de interação com seres humanos — mesmo robôs autônomos necessitam de pelo menos supervisão e direcionamento para realizarem as tarefas a que se destinam — e, com o crescimento da área, as possibilidades de interação crescem para além de operação e supervisão e passam a incluir interações físicas e sociais e aspectos como autonomia, autoridade e troca de informações. Uma das características da área de IHR é sua multidisciplinaridade: o desenvolvimento de robôs capazes de interagir de maneira eficiente e natural com seres humanos requer esforços conjuntos de profissionais de engenharias, robótica, ciências cognitivas, psicologia, interação humano-computador, biomecânica, entre outros (Goodrich & Schultz, 2007; Bauer et al., 2008; Beckerle et al., 2017).

Humanos possuem diferentes mecanismos de interação, como o de ação conjunta, isto é, quando dois ou mais agentes coordenam suas ações para alcançar um objetivo em comum (Sebanz et al., 2006). De acordo com Sebanz et al. (2006), para que ações conjuntas sejam bem sucedidas é necessário que os humanos compartilhem a percepção da realidade e do ambiente, ou seja, um agente saiba o que o outro consegue ou não perceber e quais os focos de atenção; que possam prever as ações do outro, o que pode ser alcançado pela

observação das ações atuais ou pelo conhecimento dos objetivos; e, por fim, que coordenem suas ações, levando sempre em consideração as funções, capacidades e ações de todos os envolvidos. Esse tipo de comportamento pode ser observado em times de esportes coletivos, como vôlei e basquete, quando duas ou mais pessoas carregam um objeto pesado em um ambiente com obstáculos ou quando uma pessoa adulta ajuda uma criança a se vestir, por exemplo. Os mecanismos de interação entre humanos, como a ação conjunta, são muitas vezes inspiração para o desenvolvimento da interação entre humanos e robôs. Bauer et al. (2008) discutem as habilidades que precisam ser desenvolvidas para a aplicação de ações conjuntas em cenários de colaboração humano-robô: os agentes precisam ser capazes de estimar a intenção um do outro para que possam construir uma intenção conjunta, planejar as próximas ações e executá-las. A habilidade de se comunicar de maneira natural com seres humanos, através de gestos, olhar e emoções, por exemplo, é uma das características dos robôs socialmente interativos definidos por Fong et al. (2003). Os robôs precisam perceber e interpretar os humanos durante a interação, lendo as informações enviadas por diferentes tipos de comunicação, para serem capazes de estimar intenções, prever ações, obter retornos e se ajustar adequadamente, ao mesmo tempo em que devem se comunicar por meios que sejam facilmente percebidos e interpretados pelas pessoas com quem interagem, para que estas também possam entender seus comportamentos e prever suas ações (Fong et al., 2003). A comunicação é, portanto, de extrema importância para interações bem sucedidas.

2.1 Comunicação humano-robô

Mavridis (2015) apresenta duas motivações para se equipar robôs interativos com capacidades de comunicação natural, ou seja, que utilizem os meios de comunicação naturais dos humanos. A primeira justificativa é de que pode-se tirar proveito da capacidade que as pessoas têm de interagir e ensinar: os robôs podem aprender e se adaptar durante a interação, reduzindo a necessidade de profissionais para programá-los e reprogramá-los e possibilitando sua utilização em diferentes contextos. A segunda motivação trata das possíveis aplicações que podem ser beneficiadas pela utilização de comunicações naturais, como os robôs socialmente interativos. A meta-análise realizada por Roesler et al. (2021) investigou os efeitos do antropomorfismo de robôs em diferentes variáveis de IHR, incluindo como os robôs são percebidos, a postura das pessoas em relação a eles, reações emocionais, desempenho e comportamento social. Em aplicações sociais, os resultados mostraram efeitos positivos consistentes para todas as variáveis analisadas. Além disso, observou-se que implementações de características antropomórficas através de comunicações naturais também causaram efeitos positivos consistentes na maioria das categorias de variáveis investigadas.

Humanos possuem diversas maneiras de se comunicar, desde linguagens verbais e gestos

até comunicações mais sutis. Pistas sobre o estado interno de uma pessoa podem ser obtidas a partir da observação de suas expressões faciais e da entonação utilizada durante a fala, assim como suas intenções podem ser estimadas observando-se seus movimentos pelo ambiente e acompanhando a direção de seu olhar, por exemplo. Todos esses meios de comunicação podem ser explorados na interação humano-robô.

Os diferentes tipos de comunicação podem ser classificados como explícitos ou implícitos. Breazeal et al. (2005) definem essas duas classificações de acordo com a intenção: se a informação é transmitida de maneira deliberada, a comunicação é explícita; caso contrário, se a informação transmitida é inerente a um comportamento, a comunicação é implícita. A mesma definição é feita no trabalho de Bauer et al. (2008), que enfatizam a necessidade de maior trabalho de interpretação quando informações sobre as intenções dos agentes durante uma interação são transmitidas implicitamente. Já Knepper et al. (2017) consideram que a comunicação implícita acontece quando uma informação é incorporada a um comportamento ou ação para que seja transmitida implicitamente. A utilização desse tipo de comunicação considera os conhecimentos individuais de cada agente da interação, assim como os conhecimentos que são compartilhados, e depende de contexto. O trabalho de Che et al. (2020) segue a linha de classificação de Knepper et al. (2017) e define comunicações explícitas como aquelas transmitidas de maneira deliberada e sem ambiguidade, enquanto as implícitas precisam levar em consideração o contexto para serem interpretadas. Para exemplificar os tipos de comunicação sob cada definição, a Tabela 2.1 mostra comunicações explícitas e implícitas usadas pelos robôs nos trabalhos de Breazeal et al. (2005) e Che et al. (2020).

Table 2.1: Exemplos de comunicações explícitas e implícitas usadas nos trabalhos de Breazeal et al. (2005) e Che et al. (2020).

	Breazeal et al. (2005)	Che et al. (2020)
Explícitas	<ul style="list-style-type: none"> - gestos de apontar - gestos de cabeça (em resposta a perguntas de sim e não do humano) 	<ul style="list-style-type: none"> - sinais hápticos (vibrações com significados definidos previamente)
Implícitas	<ul style="list-style-type: none"> - expressões faciais (para informar estado interno) - direção do olhar (para informar foco de atenção) 	<ul style="list-style-type: none"> - movimento (mudança de direção e velocidade)

As classificações dos tipos de comunicação são úteis, porém não podem ser vistas com rigidez. Um exemplo de comunicação que é flexível em termos de classificação são as expressões faciais. Se considerarmos que expressões faciais são involuntárias, elas são

classificadas como implícitas, de acordo com a definição de Breazeal et al. (2005). Entretanto, as pessoas podem usar expressões faciais deliberadamente quando têm consciência de estarem sendo observadas, alterando a intensidade delas, por exemplo (Parkinson, 2005, apud Frith, 2009). Nesse caso, para a classificação baseada em intenção, a expressão facial seria uma comunicação explícita. Em seu trabalho, Che et al. (2020) mencionam que comunicações como sinais sonoros e visuais só são consideradas explícitas se tiverem significados predefinidos.

Neste trabalho, a classificação de comunicações leva em consideração o que as comunicações explícitas e implícitas dos trabalhos mostrados na Tabela 2.1 têm em comum:

Explícitas: comunicações mais diretamente interpretadas;

Implícitas: fornecem informações a partir de estimativas e inferências mais subjetivas.

2.1.1 Tecnologias de comunicação

Tradicionalmente, as habilidades de conversação de robôs consideram interações lideradas pelo humano, com o robô exercendo o papel de assistente (Mavridis, 2015). Esse tipo de interação restringe a comunicação de humano para robô a comandos simples, normalmente pouco flexíveis (alterar as palavras usadas pode prejudicar a interpretação) e não há muita necessidade de que o robô fale, a menos que alguma ambiguidade precise ser resolvida (Mavridis, 2015). Essa dinâmica nem sempre é adequada e essas habilidades não são suficientes em interações mais complexas como aquelas a que se destinam os robôs socialmente interativos, por exemplo. Para que a comunicação seja natural, os robôs ainda precisam levar em consideração a entonação, o volume, o foco e todas as variáveis que podem transmitir informações emocionais, assim como as implicações do que é dito (Fong et al., 2003; Mavridis, 2015). A implementação desse tipo de comunicação é, portanto, bastante desafiadora.

Além de linguagens verbais, gestos também são muito comuns na comunicação entre humanos. Corpo, cabeça e mãos são usados em gestos de apontar, de saudação, de concordância, de negação e de despedida e em línguas de sinais complexas como a Língua Brasileira de Sinais. O primeiro passo para que esse tipo de comunicação possa ser usado em interações humano-robô é equipar o robô com a capacidade de perceber os gestos executados pelas pessoas. Entre as tecnologias para sensoriamento, câmeras de profundidade têm sido cada vez mais utilizadas (Liu & Wang, 2018), o que pode ser associado ao lançamento dos sensores Microsoft Kinect (Zhang, 2012; Lun & Zhao, 2015). Os sensores Kinect são equipados com uma câmera RGB e um sensor de profundidade e permitem a obtenção de informações sobre a cadeia cinemática do humano, sendo atualmente utilizados em diversas aplicações, como jogos, educação, reabilitação e interação humano-robô (Lun & Zhao, 2015). Esse tipo de informação permite o desenvolvimento de métodos baseados em

modelos do corpo humano para a caracterização e diferenciação de gestos. Robôs podem ser equipados para reconhecerem sinais de “sim” e “não” e acenos com as mãos (Biswas & Basu, 2011), um dicionário de gestos pode ser criado para possibilitar o controle de robôs móveis (Gu et al., 2012; Kim et al., 2015) e gestos de apontar podem ser interpretados para indicação de regiões de interesse no ambiente (Tölgyessy et al., 2018; Azari et al., 2019).

Rastrear a cadeia cinemática do humano permite também o monitoramento geral de seus movimentos. Com isso, é possível, por exemplo, determinar qual a etapa atual em uma tarefa colaborativa a partir da localização e da postura do humano no ambiente ou estimar emoções a partir do estilo de caminhada para definir distâncias confortáveis entre humano e robô (Narayanan et al., 2020). Além disso, métodos baseados em sensores não acoplados ao corpo humano permitem movimentos sem interferências, o que é importante quando o objetivo é proporcionar uma interação natural entre humanos e robôs.

Direção do olhar também é uma fonte importante de informação sobre o humano. Com essa informação, é possível estimar intenções, resolver ambiguidades e prever ações (Bauer et al., 2008; Mutlu et al., 2009; Huang & Mutlu, 2016). Cazzato et al. (2020) fazem uma revisão de métodos para rastreamento da direção do olhar, um problema explorado na área de visão computacional. As abordagens podem ser baseadas em imagens que contenham o rosto da pessoa ou em análises da cena e utilizam diversos tipos de equipamentos, desde óculos e capacetes até câmeras de baixa resolução posicionadas no ambiente. Processamento de imagens e aprendizado de máquina são importantes tópicos dentro dessa área de pesquisa. No trabalho, discute-se que ainda não há um método de rastreamento do olhar que seja a melhor escolha em todos os contextos e que é difícil avaliar as abordagens utilizando uma métrica única. Por isso, é importante considerar as especificidades de cada aplicação, como a necessidade de resposta em tempo real e a possibilidade de se realizar etapas de calibração e de se usar equipamentos mais invasivos, para definir qual método melhor atende os requisitos (Cazzato et al., 2020).

Assim como movimento e características da fala, como entonação e velocidade, as expressões faciais e sinais fisiológicos também são fontes de informação sobre o estado interno das pessoas (Zoghbi et al., 2010; Mavridis, 2015). Na comunicação humana, expressões faciais são produzidas e percebidas tanto de maneira consciente quanto inconscientemente (Frith, 2009) e são importante fonte de informação para se estimar emoções. Algumas técnicas utilizadas no reconhecimento de expressões faciais de humanos são discutidas no trabalho de Huang et al. (2019). Já com sinais fisiológicos, é possível, por exemplo, estimar o nível de engajamento durante a interação utilizando a resposta galvânica e a temperatura da pele (Mower et al., 2007) ou detectar mentiras a partir de características da pupila (Pasquali et al., 2020). Sinais hápticos, obtidos através de sensores de força/pressão, podem ser utilizados para estimar intenções e prever ações em diversas aplicações colaborativas, como transporte de objetos e processos de montagem (Ajoudani et al., 2018).

Além das habilidades de percepção e interpretação, os robôs também são equipados com ferramentas para produção de comunicações naturais. Kismet (Figura 2.1a), um robô criado para interagir e aprender com humanos, é capaz de transmitir pistas sociais através de expressões faciais e direção do olhar (Breazeal, 2003); Valerie (Figura 2.1b) é uma recepcionista virtual que se comunica através de voz, gestos de cabeça, expressões faciais e direção do olhar e aceita entradas de voz e através de um teclado (Gockley et al., 2005); a aplicação *robot_face* (Figura 2.1c), desenvolvida para ser usada em aplicações de IHR, é uma animação não realista de um rosto e se comunica por voz, com movimentos labiais sincronizados, e expressões faciais (Seib et al., 2013); Kaspar (Figura 2.1d), que se assemelha a um brinquedo infantil, se comunica através de voz, gestos e expressões faciais, além de possuir sensores táteis que permitem o reconhecimento de interações por tato com humanos e um teclado para controle manual do robô (Robins et al., 2018).

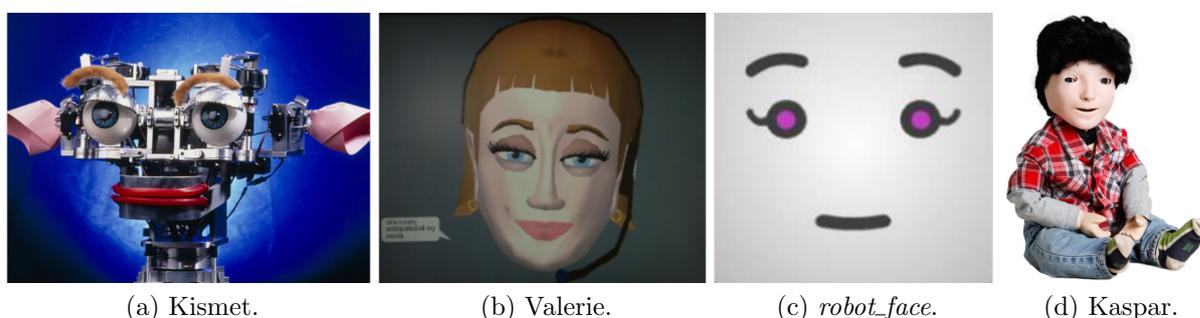


Figure 2.1: Exemplos de sistemas que utilizam comunicações naturais para interagir com humanos¹.

Para comunicação através de gestos, cabeças e rostos virtuais permitem alguns sinais explícitos com a cabeça, como movimentos de “sim” e “não”, que também podem ser executados por cabeças e torsos físicos. Usar agentes virtuais em conjunto com manipuladores robóticos, por exemplo, pode aumentar a expressividade do sistema final, possibilitando a execução de gestos que utilizam braços e mãos. Já robôs humanoides podem imitar uma maior variedade de gestos utilizados pelas pessoas. Direção do olhar também pode ser usada para transmitir informações sobre o foco de atenção do robô e para regular a tomada de vez durante a interação, *i.e.*, a dinâmica em que a palavra ou a ação é alternada entre os agentes, assim como feito com Kismet no trabalho de Breazeal (2003). Os movimentos dos robôs, físicos ou virtuais e antropomórficos ou não, também podem ser expressivos e transmitir informações sobre cansaço, estilo, urgência e personalidade aos humanos (Venture & Kulić, 2019).

Soluções para cada uma das modalidades de comunicação individuais podem ser exploradas, aprimorando-se as técnicas e melhorando o desempenho. Porém, muitas vezes o objetivo é a criação de plataformas que unam diferentes tipos de comunicação, tanto para a

¹Fontes: (a) <https://robots.ieee.org/robots/kismet/> (b) Gockley et al., 2005 (c) Seib et al., 2013 (d) <https://robots.ieee.org/robots/kaspar/>

criação de sistemas complexos quanto para a sua aplicação em experimentos de investigação de aspectos de IHR. Lenz et al. (2010) desenvolveram uma infraestrutura multimodal de comunicação para a investigação de aspectos físicos, cognitivos e comportamentais da interação entre humanos e robôs. A plataforma é composta pelo humanoide BERT2, que possui cabeça, torso, braços e mãos, é capaz de falar, produzir diferentes expressões faciais e movimentar a cabeça, além de reconhecer falas, gestos de apontar e direção do olhar do humano. Zhang et al. (2010), interessados em estudar variáveis de tempo nas interações entre humanos e agentes virtuais para que seja possível desenvolver agentes capazes de imitar padrões de comportamento, desenvolveram uma plataforma composta por quatro sistemas: um ambiente virtual, um agente virtual e sistemas de coleta e processamento de dados. O ambiente virtual para interação pode ser manipulado tanto pela pessoa quando pelo agente virtual, que simula comportamentos humanos. São coletados dados de movimentos dos olhos, orientação da cabeça, entradas manuais e falas do humano e olhar, ações manuais e expressões faciais do agente virtual. O processamento dos dados busca por padrões de comportamento que podem ser usados na criação de modelos e no levantamento de hipóteses para estudos sobre IHR. Já HIPOP, criada por Lazzeri et al. (2014), é uma plataforma modular de aquisição de dados desenvolvida para a utilização em experimentos de investigação de comportamentos emocionais em IHR. A plataforma é composta por diferentes módulos independentes que podem ser configurados e combinados de acordo com a aplicação. Entre os módulos estão: um sistema de aquisição de dados do estado de um robô androide capaz de falar, exibir diferentes expressões faciais e movimentar olhos e cabeça; uma camisa sensorial para monitoramento de sinais fisiológicos do humano; uma luva para aquisição de dados fisiológicos e reconhecimento de gestos com a mão e posturas; e um sistema de detecção da direção do olhar do humano e monitoramento de outros parâmetros, como tempo de fixação do olhar.

2.1.2 Experimentos de interação e estudos comparativos

Com todas essas possibilidades de comunicação surgem diversas perguntas. Quais tipos de comunicação devem ser utilizados em quais contextos? Adicionar meios de comunicação implícita nos robôs melhora a interação com as pessoas? Diferentes comunicações afetam tanto o desempenho quanto as percepções do humano sobre a interação e o robô? Como os sinais enviados por robôs são interpretados? Quais fatores influenciam essa interpretação? Experimentos de interação entre humanos e robôs buscam responder essas e outras perguntas. Em um cenário em que o robô faz perguntas a pessoas que passam e param para interagir com ele, Bruce et al. (2002) investigaram se equipar o robô com maior expressividade, mais especificamente a presença de um rosto com expressões faciais e movimento de cabeça para direcionamento do olhar, afetava o engajamento das pessoas e, em seus resultados, apenas o primeiro fator teve efeito na quantidade de pessoas que

pararam para interagir com o robô. Breazeal et al. (2005) exploraram a combinação de comunicações implícitas e explícitas em uma tarefa em que o humano primeiro deve ensinar o nome de botões ao robô e depois fazer com que ele os acione. O trabalho compara duas condições, uma com apenas comunicação explícita do robô e outra com a combinação de comunicações explícitas e implícitas. Na primeira, o robô comunica seu estado interno apenas quando requisitado pela pessoa; na segunda, o robô dá pistas sobre seu foco de atenção e seu estado interno através do direcionamento de seu olhar e de expressões faciais, além de transmitir vivacidade com comportamentos como piscar de olhos. Em ambos os casos, a pessoa se comunica com o robô apenas explicitamente, por voz e gestos. Os resultados observados por Breazeal et al. (2005) indicaram que participantes entenderam melhor e criaram modelos mentais melhores do robô na configuração em que as comunicações implícitas e explícitas eram combinadas. Nessa condição, o tempo de execução da tarefa também foi menor e a interação se mostrou mais robusta a erros, que foram mais rapidamente percebidos e melhor mitigados. Já Mutlu et al. (2009) buscaram entender como pistas vazadas através do olhar do robô são interpretadas e utilizadas pelas pessoas durante a interação. Para isso, propuseram um jogo em que a pessoa deveria adivinhar o objeto escolhido pelo robô, dentre algumas opções disponíveis, utilizando perguntas de “sim” ou “não” e compararam duas condições de comunicação para dois humanoides diferentes, um com características antropomórficas mais abstratas e outro um androide de aparência próxima à humana. Em uma das condições de comunicação os robôs olhavam de relance para o objeto escolhido antes de responder as perguntas. Participantes acertaram o objeto escolhido mais rapidamente e com menos perguntas quando o robô androide usava a comunicação implícita pelo olhar. A discussão no trabalho sugere que as pessoas podem não ter atribuído intencionalidade à comunicação do robô de aparência menos similar à humana e por isso não houve diferença significativa entre as duas condições nesse caso. Em uma interação de longo prazo, Tanaka et al. (2012) analisaram os efeitos de um robô comunicativo, que fala e acena com a cabeça, em comparação com um robô de controle que não se comunica, em algumas medidas fisiológicas e aspectos como apetite e sono de mulheres idosas que moram sozinhas. Os resultados sugerem que a companhia de um robô comunicativo pode causar benefícios em funções cognitivas e em outros aspectos do dia a dia dessa parcela da população.

Mais interessados na leitura de informações do humano, Huang & Mutlu (2016) investigaram os efeitos na interação quando o robô utiliza a direção do olhar da pessoa para antecipar seus comandos explícitos e agir de acordo com a previsão, observando que utilizar a previsão melhorava variáveis de desempenho da tarefa e fazia com que as pessoas percebessem o robô como mais consciente da interação. Iwasaki et al. (2019) observaram, em experimentos tanto em um ambiente real quanto em um ambiente controlado, que demonstrar ao humano que o robô percebia seus comportamentos encorajava a interação entre eles. Já em um contexto de navegação social, *i.e.*, quando a navegação de um agente

influência na navegação do outro, Che et al. (2020) observaram que as pessoas entendiam melhor as intenções do robô, confiavam mais nele e conseguiam navegar de maneira mais eficiente (usando caminhos mais curtos) quando o robô comunicava sua intenção explicitamente (enviando sinais hápticos a um equipamento usado pelo humano), quando necessário, e implicitamente (mudando de direção e velocidade) e previa os movimentos do humano para se adequar. Essa condição foi comparada com outras duas: quando o robô previa os movimentos e se comunicava apenas implicitamente e quando realizava somente evitamento de colisão sem previsão de movimentos.

Também é importante investigar como utilizar cada tipo de comunicação disponível no robô. Takayama & Pantofaru (2009) buscaram identificar fatores que influenciam no comportamento proxêmico das pessoas, *i.e.*, nas distâncias pessoais confortáveis, quando humano e robô se aproximam um do outro. Entre os fatores analisados pelas autoras, estão duas possibilidades de direção do olhar do robô, para as pernas ou para o rosto da pessoa, e observou-se que quando o robô olhava para o rosto da pessoa, a distância confortável era maior para mulheres do que para homens. Por outro lado, considerando um cenário de navegação, o trabalho de Fiore et al. (2013) sugere que o comportamento proxêmico do robô, *i.e.*, qual o nível de proximidade que o robô permite entre ele e o humano, é mais importante do que o comportamento do seu olhar na percepção que as pessoas têm do robô como um agente social e nos estados emocionais que atribuem a ele. Em uma interação que incluía um momento de contato físico entre humano e robô, Hirano et al. (2018) investigaram diferentes estilos de toque e comportamentos do olhar do robô durante o contato físico. Com os resultados obtidos, observou-se que manter o contato visual aumentava o conforto das pessoas durante a interação de toque e que ativamente tocar o robô parece ser melhor do que ser tocado por ele. Já o trabalho de Agrigoroaie et al. (2020) explorou a teoria de foco regulatório (Higgins, 1997, apud Agrigoroaie et al., 2020) e duas abordagens de comportamento, quando as pessoas guiam suas ações para alcançar os objetivos ou para evitar fracassos (Crowe & Higgins, 1997, apud Agrigoroaie et al., 2020), e concluiu que o tempo de interação aumentava quando humano e robô tinham o mesmo tipo de comportamento. As abordagens foram caracterizadas pela velocidade dos movimentos e da fala do robô e pela construção da mensagem que convidava a pessoa a interagir e formulários foram aplicados após a interação para classificar o comportamento de participantes. Belkaid et al. (2021) investigaram o efeito do olhar de um robô na tomada de decisões das pessoas durante um jogo, comparando olhar mútuo e desvio de olhar do robô. Seus resultados indicam que essa comunicação do robô é importante, tendo causado efeito, por exemplo, no tempo de resposta e na estratégia adotada por participantes. Esses trabalhos mostram como a maneira com que o robô se comunica precisa ser ajustada considerando diversos fatores, como o perfil das pessoas com quem ele interage e o contexto da aplicação.

Outros aspectos da interação e da utilização dos tipos de comunicação também podem

ser investigados. Gockley et al. (2006) analisaram os efeitos do humor demonstrado pela robô Valerie (Figura 2.1b) na interação e na percepção das pessoas, observando que na condição positiva (feliz), Valerie foi percebida como menos enigmática, ou seja, mais fácil de entender, do que nas condições neutra e negativa (triste). Guznov et al. (2019) investigaram como o nível de transparência nas informações fornecidas pelo robô e a maneira como o robô fala (se usa “eu” ou “nós”) influencia no desempenho, na confiança e na carga de trabalho do humano. Os resultados mostraram que mudanças na fala (no caso, o uso de “nós”) com a intenção de induzir comunicações mais confiáveis e de inclusão em times, podem não ser suficientes e até serem prejudiciais dependendo do contexto da interação. Também é necessário investigar se e quando utilizar as comunicações disponíveis nos robôs, para que não haja falta nem excesso de informações para as pessoas. Unhelkar et al. (2017) discutem os desafios de se implementar algoritmos para a tomada desse tipo de decisão, que incluem as dificuldades de se estimar custos e benefícios das comunicações e de se definir modelos para agentes humanos. Nessa linha, o planejamento para navegação socialmente consciente proposto por Che et al. (2020) inclui o planejamento para comunicação explícita do robô com o humano. Aspectos éticos na IHR, como o apego emocional a robôs e a possibilidade de que eles enganem as pessoas, também são investigados (Maris et al., 2020). A comunicação entre humanos e robôs tem diferentes objetivos e pode provocar diversas respostas e consequências, que devem ser levadas em consideração no seu desenvolvimento.

Por fim, também tenta-se identificar outros fatores que podem influenciar a IHR, como idade, gênero, familiaridade com robôs e até se as pessoas possuem ou não animais de estimação. Por exemplo, os resultados do trabalho de Rau et al. (2009) indicam que o contexto cultural pode influenciar a percepção das pessoas sobre o robô e a disposição em aceitar suas recomendações. Interações e comunicações bem sucedidas dependem de muitas variáveis, culturais, emocionais, cognitivas e sociais, por exemplo, por isso é importante uma extensa investigação das interações entre humanos e robôs, sob diferentes pontos de vista e em diferentes cenários, para que seja possível desenvolver robôs que possam cumprir suas funções de maneira eficiente e agradável para as pessoas com quem interagem.

2.2 Conclusão

Como a comunicação é essencial para a interação, o estudo da comunicação humano-robô é importante dentro da área de IHR. Usar a comunicação humana como inspiração para o desenvolvimento de comunicações entre humanos e robôs permite interações melhores e mais naturais, aproveitando habilidades que as pessoas já tem (Fong et al., 2003; Mavridis, 2015; Roesler et al., 2021). Além de desenvolver tecnologias de comunicação individuais, muitas vezes o objetivo é combinar múltiplas modalidades de comunicação em plataformas para serem usadas em experimentos de interação, como nos trabalhos de Lenz et al. (2010), Zhang et al. (2010) e Lazzeri et al. (2014). Neste trabalho, uma infraestrutura de

comunicação humano-robô também é implementada para a utilização em experimentos de interação.

Dentro da interação e da comunicação entre humanos e robôs, diversos aspectos podem ser investigados, como quais tipos de comunicações utilizar, como utilizá-las, se e quando é vantajoso se comunicar, levando em consideração custos e benefícios das comunicações, e quaisquer fatores que possam influenciar na interação, como cultura, gênero e idade das pessoas. Estudos como o de Breazeal et al. (2005), que investigam a comunicação de robô para humano, sugerem que medidas objetivas e subjetivas da interação melhoram quando o robô utiliza tanto comunicações explícitas, aquelas que tem interpretações mais diretas, quanto implícitas, das quais informações são retiradas a partir de estimativas e inferências mais subjetivas. Já trabalhos como o de Huang & Mutlu (2016) apontam que aproveitar as informações transmitidas implicitamente pelas pessoas também melhora a interação. Considerando essa literatura sobre comunicações explícitas e implícitas na interação entre humanos e robôs, este trabalho define hipóteses comparando a utilização apenas de comunicações explícitas com a combinação de comunicações explícitas e implícitas em uma interação entre humanos e agentes virtuais e as investiga através de experimentos de interação. Aspectos relacionados ao planejamento de experimentos de interação e à análise de resultados também são considerados, buscando-se obter intuições sobre procedimentos experimentais e enriquecendo a literatura de comunicação humano-robô.

3

INFRAESTRUTURA DE COMUNICAÇÃO HUMANO-ROBÔ

Este capítulo descreve o desenvolvimento de uma infraestrutura de comunicação humano-robô multimodal para a realização de experimentos de interação. O sistema completo é responsável pela percepção e produção de comunicações explícitas e implícitas selecionadas, sendo que o desempenho em conjunto é mais importante que o desempenho individual de cada subsistema. Entre as características desejadas estão baixo custo de equipamentos e a captura de informações do humano sem que sejam necessários trajes ou acessórios especiais, possibilitando interações mais naturais. Resultados parciais do desenvolvimento da infraestrutura de comunicação humano-robô foram publicados no *17th IEEE Latin American Robotics Symposium/8th Brazilian Symposium of Robotics* por Campos & Adorno (2020).

As comunicações explícitas do humano percebidas são gestos de apontar e entradas manuais, enquanto as comunicações implícitas reconhecidas são direção do olhar e posicionamento no ambiente. Para desempenhar o papel de robô nos experimentos, uma aplicação de uma agente virtual foi criada, reduzindo custo e complexidade se comparado à utilização de um sistema robótico físico. A agente virtual se comunica explicitamente por voz, sinais sonoros e informações em tela e implicitamente por expressões faciais, direção do olhar e levantar as sobrancelhas.

A próxima seção faz uma revisão sobre quatérnios e quatérnios duais, utilizados neste trabalho para a representação de movimentos rígidos em três dimensões e primitivas

geométricas. As seções seguintes detalham cada um dos subsistemas de comunicação.

3.1 Fundamentos matemáticos

3.1.1 Quatérnios

Quatérnios são números compostos por uma parte real e uma parte imaginária, são representados por $\mathbf{h} = h_1 + \hat{i}h_2 + \hat{j}h_3 + \hat{k}h_4$ e podem ser entendidos como extensões dos números complexos. O conjunto dos quatérnios é dado por (Hamilton, 1844, apud Adorno, 2017)

$$\mathbb{H} \triangleq \{h_1 + \hat{i}h_2 + \hat{j}h_3 + \hat{k}h_4 : h_1, h_2, h_3, h_4 \in \mathbb{R}\},$$

sendo que $\hat{i}^2 = \hat{j}^2 = \hat{k}^2 = \hat{i}\hat{j}\hat{k} = -1$.

As operações de soma e multiplicação de quatérnios são análogas às utilizadas para os números complexos, respeitando-se as propriedades das unidades imaginárias. O conjugado \mathbf{h}^* de um quatérnio dual é tal que $\mathbf{h}^* \triangleq \text{Re}(\mathbf{h}) - \text{Im}(\mathbf{h})$, sendo $\text{Re}(\cdot)$ e $\text{Im}(\cdot)$ operadores que extraem as partes real e imaginária, respectivamente, e sua norma é definida como $\|\mathbf{h}\| \triangleq \sqrt{\mathbf{h}\mathbf{h}^*} = \sqrt{\mathbf{h}^*\mathbf{h}}$.

Os quatérnios que possuem parte real nula são chamados de quatérnios puros e formam o conjunto $\mathbb{H}_p \triangleq \{\mathbf{h} \in \mathbb{H} : \text{Re}(\mathbf{h}) = 0\}$, enquanto os quatérnios unitários, aqueles com norma igual a 1, compõem o conjunto $\mathbb{S}^3 \triangleq \{\mathbf{h} \in \mathbb{H} : \|\mathbf{h}\| = 1\}$. Assim como os números complexos podem ser usados para representar movimentos rígidos em duas dimensões, quatérnios representam movimentos rígidos em três dimensões (Adorno, 2017). O conjunto \mathbb{H}_p é usado para representar translações, tendo forte relação com o conjunto \mathbb{R}^3 . As unidades imaginárias do quatérnio formam uma base ortogonal e um ponto no espaço é representado por $\mathbf{p} = \hat{i}p_x + \hat{j}p_y + \hat{k}p_z$, sendo p_x , p_y e p_z as coordenadas do ponto em cada um dos eixos ortogonais. As operações de produto interno e externo entre dois quatérnios puros, $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{H}_p$, são definidas, respectivamente, como (Adorno, 2017)

$$\langle \mathbf{p}_1, \mathbf{p}_2 \rangle \triangleq -\frac{(\mathbf{p}_1\mathbf{p}_2 + \mathbf{p}_2\mathbf{p}_1)}{2} \quad \text{e} \quad \mathbf{p}_1 \times \mathbf{p}_2 \triangleq \frac{(\mathbf{p}_1\mathbf{p}_2 - \mathbf{p}_2\mathbf{p}_1)}{2}.$$

A interpretação geométrica de tais operações é a mesma das operações entre vetores de três dimensões. A translação inversa é dada pelo conjugado $\mathbf{p}^* = -\mathbf{p}$, de modo que $\mathbf{p} + \mathbf{p}^* = \mathbf{p} - \mathbf{p} = 0$.

Por sua vez, os elementos do conjunto \mathbb{S}^3 são usados para representar rotações. Todo $\mathbf{r} \in \mathbb{S}^3$ pode ser escrito na forma $\mathbf{r} = \cos(\phi/2) + \mathbf{n} \sin(\phi/2)$, sendo $\mathbf{n} \in \mathbb{H}_p \cap \mathbb{S}^3$ (*i.e.*, um quatérnio puro de norma unitária) o eixo de rotação e $\phi \in \mathbb{R}$ o ângulo de rotação. A operação inversa da transformação \mathbf{r} é dada pelo seu conjugado \mathbf{r}^* ($\mathbf{r}\mathbf{r}^* = 1$), que representa uma rotação de $-\phi$ em torno do eixo \mathbf{n} .

Seja \mathbf{r}_b^a a rotação do sistema de coordenadas \mathcal{F}_b em relação a \mathcal{F}_a , o ponto \mathbf{p}^b , dado em

relação ao sistema de coordenadas \mathcal{F}_b , pode ser projetado e descrito em relação ao sistema de coordenadas \mathcal{F}_a . O ponto em relação a \mathcal{F}_a é dado por

$$\mathbf{p}^a = \mathbf{r}_b^a \mathbf{p}^b (\mathbf{r}_b^a)^* \quad (3.1)$$

A transformação do ponto \mathbf{p}^a para \mathbf{p}^b é dada por $\mathbf{p}^b = \mathbf{r}_a^b \mathbf{p}^a (\mathbf{r}_a^b)^*$, sendo $\mathbf{r}_a^b = (\mathbf{r}_b^a)^{-1} = (\mathbf{r}_b^a)^*$.

Desde que representadas a partir do mesmo sistema de coordenadas inercial, translações sucessivas são dadas pela soma de quatérnios de translação, enquanto rotações sucessivas são obtidas pela multiplicação de quatérnios de rotação. Usando o exemplo mostrado na Figura 3.1, obtém-se a transformação entre os sistemas de coordenadas \mathcal{F}_a e \mathcal{F}_c , descrita em relação à \mathcal{F}_a , através das operações

$$\mathbf{p}_{ac}^a = \mathbf{p}_{ab}^a + \mathbf{p}_{bc}^a = \mathbf{p}_{ab}^a + \mathbf{r}_b^a \mathbf{p}_{bc}^b (\mathbf{r}_b^a)^* \quad \text{e} \quad \mathbf{r}_c^a = \mathbf{r}_b^a \mathbf{r}_c^b \quad (3.2)$$

A ausência de movimento de rotação é representada pelo quatérnio $\mathbf{r} = \cos(0) + \mathbf{n} \sin(0) = 1$ e a ausência de translação por $\mathbf{p} = \hat{i}0 + \hat{j}0 + \hat{k}0 = 0$.

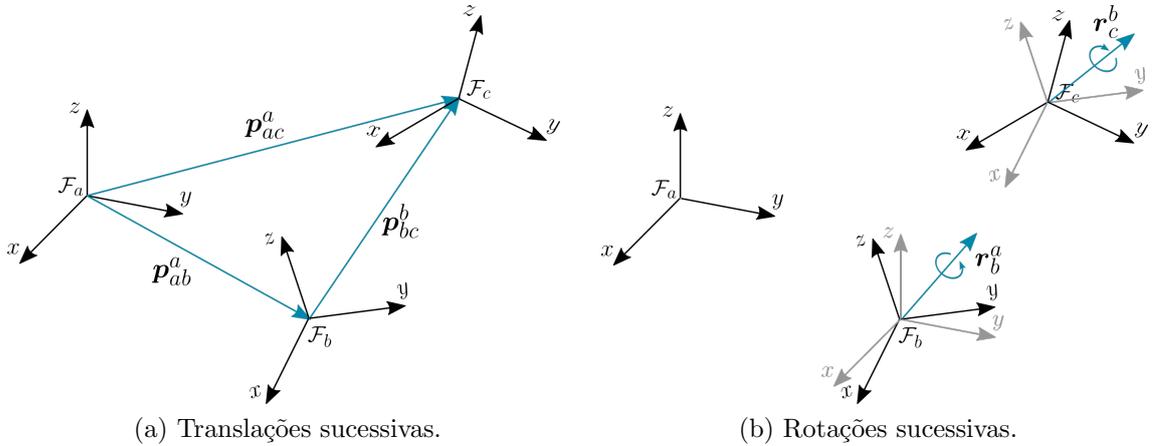


Figure 3.1: Transformações sucessivas entre sistemas de coordenadas.

3.1.2 Quatérnios duais e primitivas geométricas

A unidade dual ε é uma unidade algébrica que respeita as propriedades $\varepsilon \neq 0$ e $\varepsilon^2 = 0$ (Clifford, 1871, apud Adorno, 2017) e está presente nos chamados números duais, formados por uma parte primária e uma parte dual, sendo esta a que acompanha a unidade ε . Números duais compostos por quatérnios são chamados quatérnios duais e formam o conjunto

$$\mathcal{H} \triangleq \{\mathbf{h} + \varepsilon \mathbf{h}' : \mathbf{h}, \mathbf{h}' \in \mathbb{H}, \varepsilon \neq 0, \varepsilon^2 = 0\}.$$

Quatérnios duais possuem fortes propriedades algébricas e podem ser usados na representação de movimentos rígidos e de primitivas geométricas de maneira direta e compacta (Adorno, 2017).

Primitivas geométricas como linhas e planos são úteis na definição de restrições em movimentos e na delimitação de áreas de interesse no espaço, por exemplo. Seja $\mathbf{l} \in \mathbb{H}_p \cap \mathbb{S}^3$ a direção da linha e $\mathbf{p}_l \in \mathbb{H}_p$ um ponto arbitrário nela (ver Figura 3.2a), uma linha de Plücker em quatérnios duais é representada por

$$\underline{\mathbf{l}} = \mathbf{l} + \varepsilon(\mathbf{p}_l \times \mathbf{l}) = \mathbf{l} + \varepsilon\mathbf{m}, \quad (3.3)$$

onde \mathbf{m} é o momento da linha e indica a posição espacial de $\underline{\mathbf{l}}$, independente do ponto \mathbf{p}_l (Daniilidis, 1999, apud Adorno, 2017).

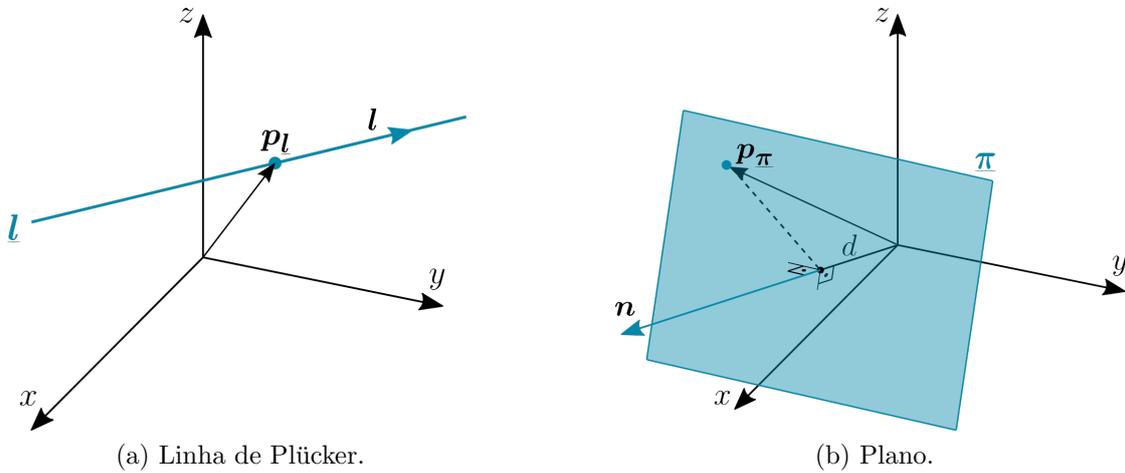


Figure 3.2: Primitivas geométricas.

Um plano pode ser completamente representado a partir da direção de sua normal, $\mathbf{n} \in \mathbb{H}_p \cap \mathbb{S}^3$, e de um ponto arbitrário $\mathbf{p}_\pi \in \mathbb{H}_p$ sobre ele. Seja $d = \langle \mathbf{p}_\pi, \mathbf{n} \rangle$ a distância perpendicular do plano em relação ao sistema de coordenadas de referência (ver Figura 3.2b), em quatérnios duais o plano $\underline{\pi}$ é dado por (Adorno, 2017)

$$\underline{\pi} = \mathbf{n} + \varepsilon d. \quad (3.4)$$

3.2 Gestos de apontar

Gestos de apontar são um tipo de comunicação explícita que pode ser usada para indicar objetos ou regiões no ambiente. Uma câmera RGB-D (Microsoft Kinect) e a aplicação `openni_tracker`¹ para ROS (*Robot Operating System*) são usadas para rastrear a posição

¹http://wiki.ros.org/openni_tracker

de 15 juntas da cadeia cinemática do humano, mostradas na Figura 3.3, incluindo quadris, cotovelos e mãos.

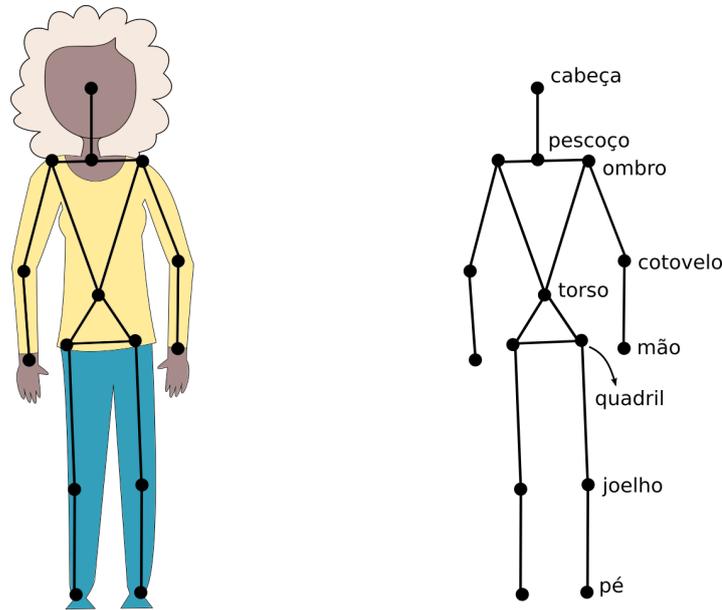


Figure 3.3: Juntas da cadeia cinemática do humano rastreadas. Cada ponto representa uma junta.

Neste trabalho, o gesto de apontar é caracterizado pela distância entre as juntas de quadril e mão. As Figuras 3.4a e 3.4b ilustram a diferença na distância D_{qm} entre quadril e mão quando a pessoa está em posição de repouso e quando aponta. A distância D_{qm} , calculada a cada instante, é dada por

$$D_{qm} = \|\mathbf{p}_{\text{quadril}} - \mathbf{p}_{\text{mão}}\|,$$

sendo $\mathbf{p}_{\text{quadril}}$ e $\mathbf{p}_{\text{mão}}$ as posições das juntas do quadril e da mão, respectivamente, descritas em relação ao sistema de coordenadas da câmera RGB-D.

As distâncias são obtidas para as juntas esquerdas e direitas e são comparadas com limites definidos previamente: se ambas as distâncias são menores que seus limites, considera-se que a pessoa está em repouso; se pelo menos uma delas é maior que o limite, então a pessoa está apontando e a maior das distâncias indica o braço utilizado para o gesto. O gesto pode ser feito com qualquer um dos braços, mas assume-se que apenas um é usado por vez.

O gesto de apontar é então representado a partir das posições de cotovelo e mão do braço que aponta, $\mathbf{p}_{\text{cotovelo}}$ e $\mathbf{p}_{\text{mão}}$, como ilustrado na Figura 3.4c. Seja

$$\mathbf{l}_{\text{gesto}} = \frac{\mathbf{p}_{\text{mão}} - \mathbf{p}_{\text{cotovelo}}}{\|\mathbf{p}_{\text{mão}} - \mathbf{p}_{\text{cotovelo}}\|}$$

a direção para onde o braço aponta e $\mathbf{p}_l \in \{\mathbf{p}_{\text{mão}}, \mathbf{p}_{\text{cotovelo}}\}$, então define-se a linha de

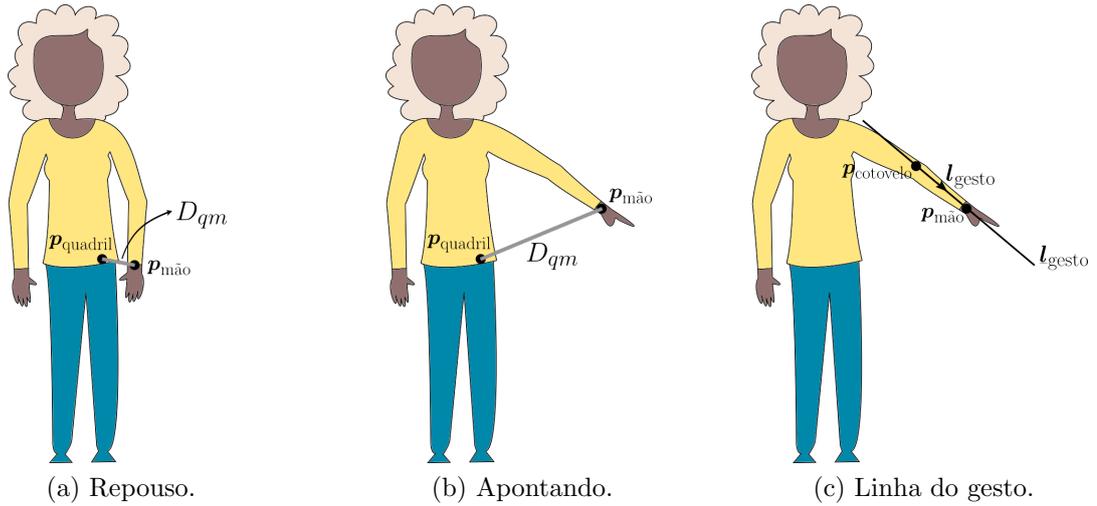


Figure 3.4: A distância D_{qm} entre quadril e mão é usada para determinar se a pessoa está em repouso ou apontando. A linha de Plücker \underline{l}_{gesto} , para representar o gesto, é definida a partir dos pontos de cotovelo e mão, $\mathbf{p}_{cotovelo}$ e $\mathbf{p}_{mão}$.

Plücker $\underline{l}_{gesto} = \mathbf{l}_{gesto} + \varepsilon(\mathbf{p}_l \times \mathbf{l}_{gesto})$ para representar o gesto de apontar.

Objetos ou regiões no ambiente são representados por caixas compostas por seis planos, paralelos ou perpendiculares entre si. Considerando posição e dimensões dos objetos conhecidas, os seis planos são definidos de forma que todos tenham normal direcionada para o interior do objeto, como ilustrado na Figura 3.5. Os planos de cada objeto $o \in \mathcal{O}$, sendo $\mathcal{O} = \{1, 2, \dots, n_{objetos}\}$, são dados por

$$\underline{\pi}_{o,i} = \mathbf{n}_{o,i} + \varepsilon d_{o,i}, \forall i \in \{1, \dots, 6\}.$$

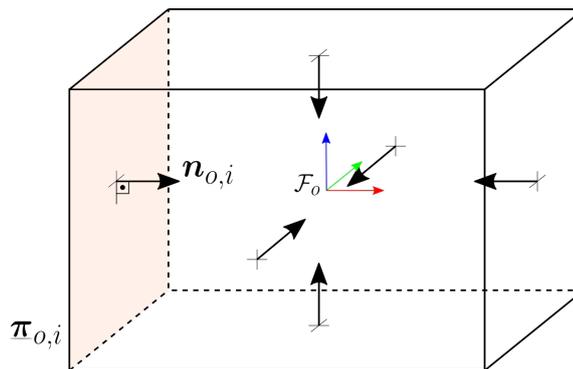


Figure 3.5: Objetos são modelados por caixas compostas por seis planos $\underline{\pi}_{o,i}$ com normais $\mathbf{n}_{o,i}$ que apontam para seu interior.

A linha de Plücker \underline{l}_{gesto} e os planos dos objetos são descritos com respeito a um sistema de coordenadas fixo de referência e são usados para a interpretação do gesto de apontar, como ilustrado na Figura 3.6. Partindo de um ponto \mathbf{p}_l na linha, o ponto de cruzamento

entre a linha do gesto e o plano $\underline{\pi}_i = \mathbf{n}_i + \varepsilon d_i$ é dado por

$$\mathbf{p}_c = \mathbf{p}_l + D_c \mathbf{l}_{\text{gesto}}, \quad (3.5)$$

sendo D_c a distância de \mathbf{p}_l até \mathbf{p}_c medida sobre a linha $\mathbf{l}_{\text{gesto}}$. Como o ponto \mathbf{p}_c pertence ao plano, deve satisfazer a equação $d_i = \langle \mathbf{p}_c, \mathbf{n}_i \rangle$. Considerando a Equação (3.5), obtém-se

$$\begin{aligned} d_i &= \langle \mathbf{p}_c, \mathbf{n}_i \rangle \\ &= \langle \mathbf{p}_l + D_c \mathbf{l}_{\text{gesto}}, \mathbf{n}_i \rangle \\ &= \langle \mathbf{p}_l, \mathbf{n}_i \rangle + \langle D_c \mathbf{l}_{\text{gesto}}, \mathbf{n}_i \rangle \\ &= \langle \mathbf{p}_l, \mathbf{n}_i \rangle + D_c \langle \mathbf{l}_{\text{gesto}}, \mathbf{n}_i \rangle. \end{aligned} \quad (3.6)$$

Reorganizando os termos em (3.6), obtém-se a distância D_c , dada por

$$D_c = \frac{d_i - \langle \mathbf{p}_l, \mathbf{n}_i \rangle}{\langle \mathbf{l}_{\text{gesto}}, \mathbf{n}_i \rangle}, \quad (3.7)$$

definida sempre que linha e plano não são paralelos (nesse caso, o denominador em (3.7) seria igual a zero).

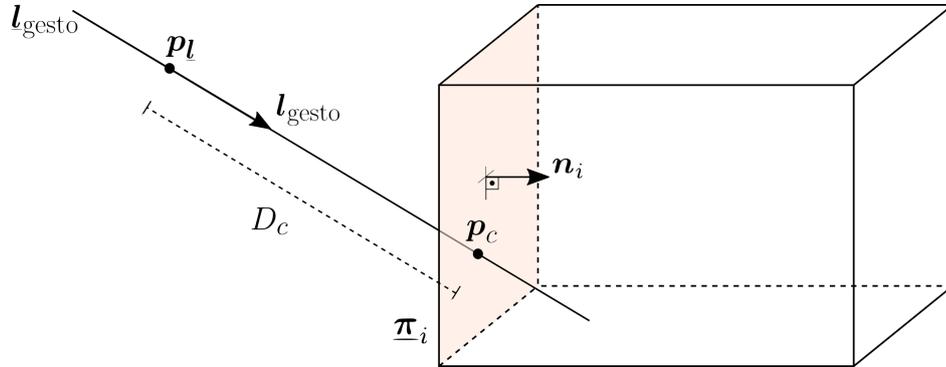


Figure 3.6: Cruzamento entre a linha do gesto $\mathbf{l}_{\text{gesto}}$ e o plano $\underline{\pi}_i$ do objeto. A distância entre \mathbf{p}_l e o ponto de cruzamento \mathbf{p}_c medida sobre a linha é dada por D_c .

Para reduzir o número de cálculos necessários, são considerados apenas objetos que estejam próximos da linha do gesto. Um objeto é considerado próximo se a distância entre seu centro geométrico e a linha é menor ou igual à maior dimensão da caixa que o representa. A distância entre linha e ponto, definida por Marinho et al. (2018), é dada por

$$\text{dist}(\mathbf{l}_{\text{gesto}}, \mathbf{p}_o) = \left\| \mathbf{p}_o \times \mathbf{l}_{\text{gesto}} - \mathbf{m}_l \right\| = \left\| \mathbf{p}_o \times \mathbf{l}_{\text{gesto}} - (\mathbf{p}_l \times \mathbf{l}_{\text{gesto}}) \right\|, \quad (3.8)$$

sendo \mathbf{p}_o o centro geométrico do objeto. Para objetos próximos da linha, os pontos de cruzamento $\mathbf{p}_{c_o,i}$, $i \in \{1, \dots, 6\}$, entre a linha e os seis planos do objeto são obtidos. A

função $\text{dist}(\mathbf{p}_{c_{o,i}}, \underline{\boldsymbol{\pi}}_{o,j}) : \mathbb{H}_p \times \mathcal{H} \rightarrow \mathbb{R}$, tal que

$$\text{dist}(\mathbf{p}_{c_{o,i}}, \underline{\boldsymbol{\pi}}_{o,j}) = \langle \mathbf{p}_{c_{o,i}}, \mathbf{n}_{o,j} \rangle - d_{o,j}, \quad (3.9)$$

calcula a distância sinalizada entre os planos $\underline{\boldsymbol{\pi}}_{o,j} = \mathbf{n}_{o,j} + \varepsilon d_{o,j}$ e o ponto $\mathbf{p}_{c_{o,i}}$, para todo $i, j \in \{1, \dots, 6\}$, do ponto de vista dos planos (Marinho et al., 2018). Se as distâncias entre o ponto de cruzamento e todos os planos forem não-negativas, *i.e.*, o ponto está sobre um ou mais planos ($\text{dist}(\mathbf{p}_{c_{o,i}}, \underline{\boldsymbol{\pi}}_{o,j}) = 0$) e na direção positiva dos planos restantes ($\text{dist}(\mathbf{p}_{c_{o,i}}, \underline{\boldsymbol{\pi}}_{o,j}) > 0$), então o ponto está dentro da caixa que delimita o objeto. No exemplo da Figura 3.7, o ponto \mathbf{p} está sobre o plano em azul e as distâncias de todos os outros planos, indicadas pelas setas, são positivas.

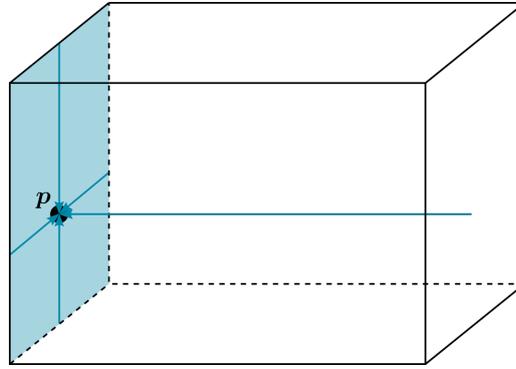


Figure 3.7: O ponto \mathbf{p} está sobre o plano azul (distância zero) e a distâncias positivas de todos os outros planos, indicadas pelas setas.

Objetos cruzados pela linha são aqueles que possuem pelo menos um ponto $\mathbf{p}_{c_{o,i}}$ em seu interior, sendo este sempre associado a uma distância $D_{c_{o,i}}$. Esses objetos são considerados possibilidades de indicação e, dentre os pontos de cruzamento que pertencem ao interior de cada um, a menor distância $D_{c_{o,i}}$ observada é armazenada. Cada objeto cruzado está, portanto, vinculado a uma distância $D_{c_{o,i}}$ mínima. O objeto cruzado mais próximo do ponto \mathbf{p}_l no braço que aponta, que possui a menor distância $D_{c_{o,i}}$ armazenada, é considerado como a indicação da pessoa, visto que os objetos mais longe podem estar ocultos pelo primeiro.

A Figura 3.8 exemplifica duas situações: uma em que a linha do gesto não cruza nenhum objeto (nenhuma indicação é reconhecida) e outra em que a linha cruza dois objetos ao mesmo tempo, $o = 1$ e $o = 2$. No segundo caso, os dois pontos de cruzamento que pertencem ao interior do objeto 1 são $\mathbf{p}_{c_{1,2}}$ e $\mathbf{p}_{c_{1,4}}$ e a menor distância associada ao objeto é $D_{c_{1,4}}$; já o objeto 2 tem os pontos $\mathbf{p}_{c_{2,1}}$ e $\mathbf{p}_{c_{2,3}}$ em seu interior e menor distância $D_{c_{2,1}}$. Os dois objetos são considerados possibilidades de indicação, já que ambos são cruzados pela linha do gesto. Como $D_{c_{2,1}}$ é menor que $D_{c_{1,4}}$, indicando que o objeto 2 está mais próximo do ponto \mathbf{p}_l que o objeto 1, considera-se que a pessoa indica o objeto 2.

A implementação do sistema foi feita em Python, utilizando a biblioteca DQ Robotics

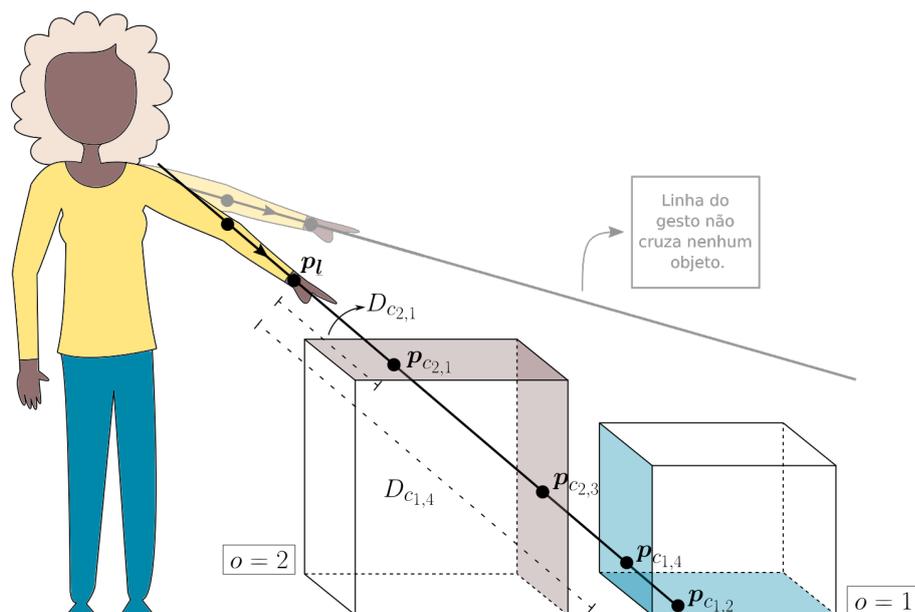


Figure 3.8: Quando a linha do gesto não cruza nenhum objeto, considera-se que não há indicação. Quando a linha cruza mais de um objeto, o escolhido como indicação é o que apresenta menor distância D_c associada.

(Adorno & Marinho, 2021) para as manipulações matemáticas. O algoritmo de reconhecimento e interpretação de gestos foi avaliado em um experimento (Campos & Adorno, 2019; Campos & Adorno, 2020), descrito no Apêndice A.1. A taxa média de acertos foi de 77,8% (desvio padrão de 16,4%), considerado um bom desempenho para a aplicação pretendida.

3.3 Direção do olhar

A direção do olhar é um tipo de comunicação implícita que pode informar ao robô qual é o foco de atenção da pessoa, seja para antecipação de comandos explícitos ou para determinação do foco de atenção geral durante a interação. A abordagem utilizada neste trabalho divide o problema de obtenção de informação a partir da direção do olhar do humano em três etapas: detecção de pontos no rosto, determinação da direção do olhar e interpretação. Os pontos de interesse no rosto são os centros das íris dos olhos e a boca, obtidos a partir de metodologias já existentes na literatura. Com os pontos conhecidos, a determinação da direção do olhar é feita a partir de análises geométricas e a metodologia para interpretação é a mesma utilizada para interpretar os gestos de apontar, descrita na Seção 3.2.

3.3.1 Detecção de pontos no rosto

OpenCV² é uma biblioteca de visão computacional que conta com diversas funções de manipulação, processamento e transformação de imagens e é utilizada em diferentes etapas deste trabalho. Um dos algoritmos já implementados na biblioteca é o de Viola & Jones (2001) para detecção de atributos faciais como rosto, olhos e boca, retornando regiões retangulares para cada um³, como ilustrado na Figura 3.9. Neste trabalho considera-se que apenas um rosto é visto por vez e os atributos faciais de interesse são buscados em cada quadro de imagem da câmera. O ponto definido para a boca é o centro do retângulo associado a ela e as regiões retangulares obtidas para os olhos são parâmetros para o algoritmo de estimativa dos centros das íris. O algoritmo utilizado é o de Timm & Barth (2011), que apresenta metodologia simples e de fácil implementação e seu desempenho se mostrou bom o suficiente para as aplicações necessárias neste trabalho. Seguindo o trabalho original, a notação vetorial é utilizada na descrição da metodologia.

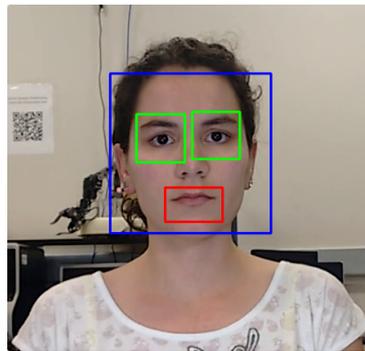


Figure 3.9: Exemplo da detecção de rosto (retângulo azul), olhos (retângulos verdes) e boca (retângulo vermelho) utilizando funções disponíveis na biblioteca OpenCV.

Timm & Barth (2011) estimam a localização dos centros das íris utilizando gradientes de cor. Entre as vantagens citadas no trabalho estão a possibilidade de utilização de câmeras de baixa resolução, a baixa complexidade computacional, quando comparado a outros métodos existentes, a robustez diante de problemas como reflexos em óculos e fios de cabelo presentes na imagem do rosto e a invariância a mudanças lineares na iluminação. A Figura 3.10a ilustra o campo vetorial dos gradientes de cor de uma imagem de um objeto circular escuro sobre um fundo claro, representando o que normalmente acontece entre a íris (escura) e a esclera (clara) no olho. Na borda do círculo, os vetores gradiente assumem valores máximos, devido à transição abrupta de cor. A Figura 3.10b mostra um exemplo dos pontos de gradientes altos na imagem de um olho real. Utilizando o conhecimento desse campo vetorial, estima-se o centro do círculo escuro.

Seja $\mathbf{x}_k \in \mathbb{R}^2$ um ponto na imagem, dado em pixels, e $\mathbf{g}_k \in \mathbb{R}^2$ o vetor gradiente relacionado a ele. Dado o possível centro $\mathbf{c} \in \mathbb{R}^2$ a ser testado, o vetor de translação entre

²<https://opencv.org/>

³https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html

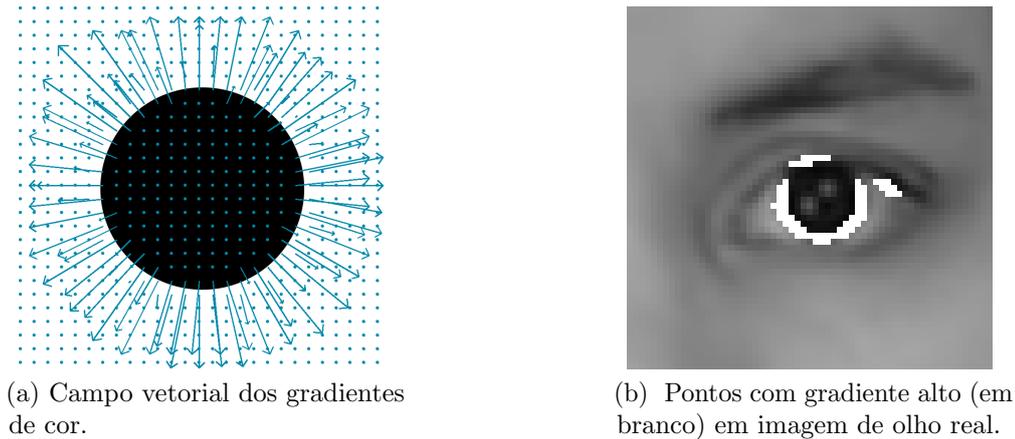


Figure 3.10: Os gradientes de cor são utilizados para estimar o centro do círculo.

\mathbf{c} e o ponto \mathbf{x}_k é representado por $\mathbf{v}_k = \mathbf{x}_k - \mathbf{c}$. Para que todos os pontos tenham o mesmo peso, os vetores \mathbf{v}_k são normalizados, obtendo-se $\hat{\mathbf{v}}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$. Os vetores gradiente são também normalizados para que o algoritmo seja mais robusto em relação a mudanças na iluminação e no contraste da imagem, de modo que $\hat{\mathbf{g}}_k = \mathbf{g}_k / \|\mathbf{g}_k\|$. Como ilustrado na Figura 3.11, o produto interno entre $\hat{\mathbf{v}}_k$ e $\hat{\mathbf{g}}_k$ é máximo quando \mathbf{c} coincide com o centro real e apresenta valores menores para pontos que não estão na linha que conecta o centro real ao ponto \mathbf{x}_k (linha pontilhada na Figura 3.11a). O centro real é aquele que gera produtos internos $\langle \hat{\mathbf{v}}_k, \hat{\mathbf{g}}_k \rangle$ máximos para todos os pontos \mathbf{x}_k da imagem ao mesmo tempo.

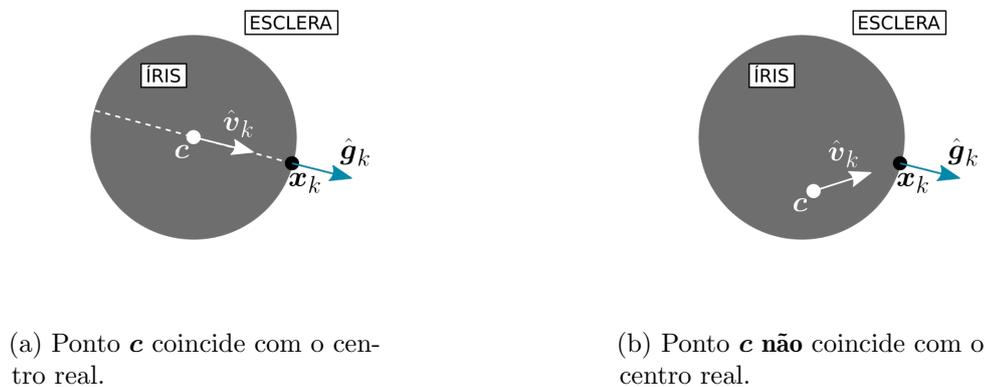


Figure 3.11: Diferença do produto interno entre $\hat{\mathbf{v}}_k$ e $\hat{\mathbf{g}}_k$.

Timm & Barth (2011) sugerem que seja dada prioridade a possíveis centros escuros, com a utilização de um peso w_c , dado pelo valor cinza do pixel de cada centro testado na imagem do olho filtrada e invertida. A imagem é filtrada para evitar problemas como reflexos de óculos e a invertida para que os valores cinza sejam maiores dentro da região da íris, como ilustrado na Figura 3.12. Além disso, possíveis centros nas bordas da imagem do olho não são analisados e um limite é utilizado para desconsiderar gradientes em regiões homogêneas, ou seja, que possuem gradientes baixos ou iguais a zero (ver Figura 3.10a), reduzindo a complexidade computacional. Neste trabalho, são considerados apenas pontos

com gradiente maior que 90% da magnitude do maior gradiente observado. Os índices dos pontos não descartados formam o conjunto $\mathcal{N} \subseteq \{1, \dots, N\}$, sendo N o número total de pixels da imagem do olho. A partir de tais considerações, o centro real é estimado por

$$\mathbf{c}_{\text{estimado}} \in \arg \max_{\mathbf{c}} \left(\frac{1}{N} \sum_{k \in \mathcal{N}} w_c \langle \hat{\mathbf{v}}_k, \hat{\mathbf{g}}_k \rangle^2 \right), \quad (3.10)$$

para todo \mathbf{g}_k tal que $\|\mathbf{g}_k\| > 0,9 \|\mathbf{g}_{\max}\|$, sendo \mathbf{g}_{\max} o gradiente de maior magnitude observado. A obtenção dos vetores gradiente e a filtragem da imagem do olho são feitas utilizando funções de OpenCV. A implementação neste trabalho considera a média das últimas cinco estimativas feitas pelo algoritmo de Timm & Barth (2011).

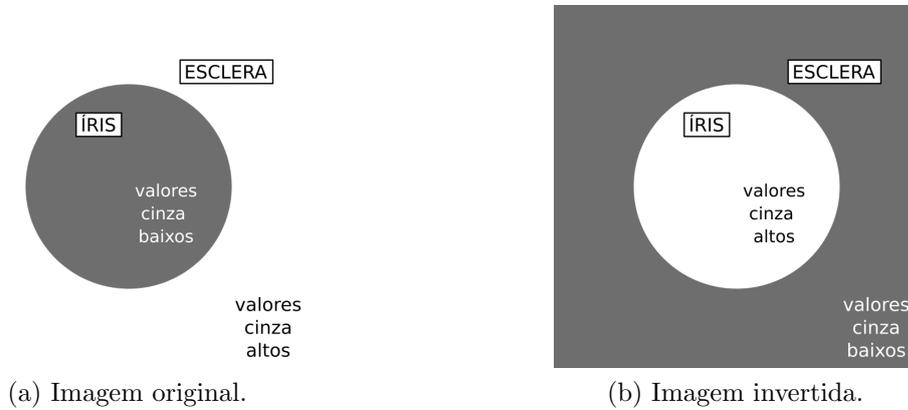


Figure 3.12: Diferença nos valores cinza na representação da íris e da esclera.

Mesmo que o humano se mantenha sempre na área de visão da câmera, nem sempre é possível detectar seus atributos faciais; muitas vezes, a movimentação da cabeça causa oclusões ou dificulta o reconhecimento. Para lidar com esse problema, filtros de Kalman com modelo de velocidade constante são aplicados aos centros das íris e aos centros dos retângulos de olhos e boca. Seja $\mathbf{a} \in \mathbb{R}^2$ o ponto rastreado e $\dot{\mathbf{a}}$ a sua velocidade, o vetor de estados é dado por $\mathbf{y} = [\mathbf{a}^T \ \dot{\mathbf{a}}^T]^T$. No instante t , a entrada e a saída do sistema são dadas, respectivamente, por

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0}_{2 \times 2} & \mathbf{I}_2 \end{bmatrix} \mathbf{y}_{t-1} \text{ e } \mathbf{z}_t = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_{2 \times 2} \end{bmatrix} \mathbf{y}_t,$$

onde T é o período de amostragem da câmera que capta a imagem do rosto, $\mathbf{I}_2 \in \mathbb{R}^{2 \times 2}$ é a matriz identidade e $\mathbf{0}_2 \in \mathbb{R}^{2 \times 2}$ é uma matriz de zeros. As predições dos filtros dos olhos e boca são usadas para definir regiões menores para a busca dos atributos faciais. Após algumas iterações sem detecção, os filtros são reinicializados e os atributos são buscados em regiões maiores: dois terços mais à esquerda ou mais à direita do retângulo do rosto para os olhos e a metade inferior para a boca. As matrizes de covariância \mathbf{R} e \mathbf{Q} dos filtros foram definidas manualmente, por tentativa e erro, e a implementação de filtro de Kalman

disponível em OpenCV é utilizada.

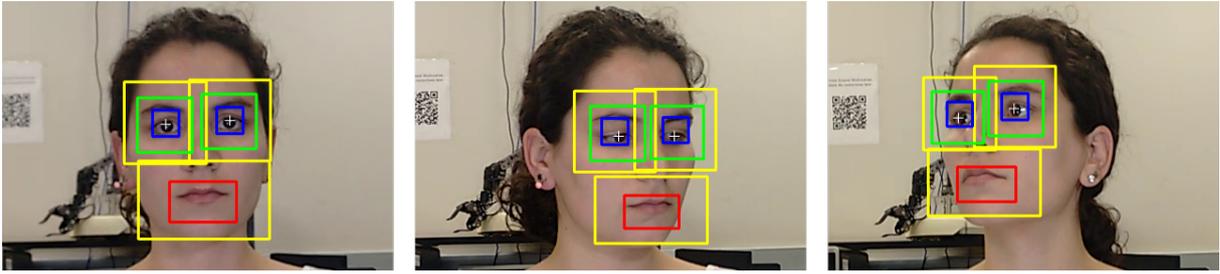


Figure 3.13: Exemplo de funcionamento do algoritmo de detecção dos pontos no rosto. Retângulos amarelos e azuis representam regiões de busca, retângulos verdes e vermelhos são as detecções de olhos e boca, respectivamente, e os centros das íris obtidos são indicados pelas cruzes brancas.

A Figura 3.13 ilustra o funcionamento do algoritmo de detecção dos pontos de interesse, que foi implementado em C++. Os retângulos amarelos representam as regiões de busca de olhos e boca, definidas a partir das predições dos filtros. Olhos detectados são representados em verde e a boca em vermelho. As predições dos filtros dos olhos também são usadas para definir as regiões representadas em azul, nas quais o algoritmo de detecção dos centros das íris é aplicado. As cruzes brancas na figura indicam os pontos dos centros das íris obtidos.

O algoritmo final de detecção dos atributos faciais, utilizado para a obtenção de distâncias de interesse descritas na próxima seção, passou por uma avaliação (Campos & Adorno, 2020), detalhada no Apêndice A.2, e o desempenho observado (média de erros, em centímetros, para as distâncias calculadas) foi considerado aceitável para a aplicação neste trabalho.

3.3.2 Determinação da direção do olhar e interpretação

Com os pontos dos centros das íris e o ponto central da região da boca, a direção do olhar do humano é estimada a partir de análises geométricas. Considera-se que a pessoa começa olhando diretamente para a câmera, de frente, como ilustrado na imagem mais à esquerda da Figura 3.14. Nessa posição, os pontos de centros das íris e da boca são armazenados para serem usados como referência e os filtros de Kalman são inicializados considerando velocidades iniciais nulas, ou seja, $\mathbf{y}_0 = [\mathbf{a}_0^T \ \mathbf{0}_2^T]$, sendo $\mathbf{0}_n \in \mathbb{R}^n$ um vetor coluna de zeros.

Seja \mathcal{F}_c o sistema de coordenadas da câmera que captura a imagem do rosto, com eixo z_c apontando para frente e eixo x_c para a sua direita. A Figura 3.14 ilustra como as distâncias horizontal entre os olhos (medida sobre o eixo x_c de \mathcal{F}_c) e vertical entre olhos e boca (medida sobre o eixo y_c de \mathcal{F}_c) variam quando a pessoa movimentar a cabeça e olhar para diferentes direções, assim como a posição da boca em relação à referência (primeira imagem).

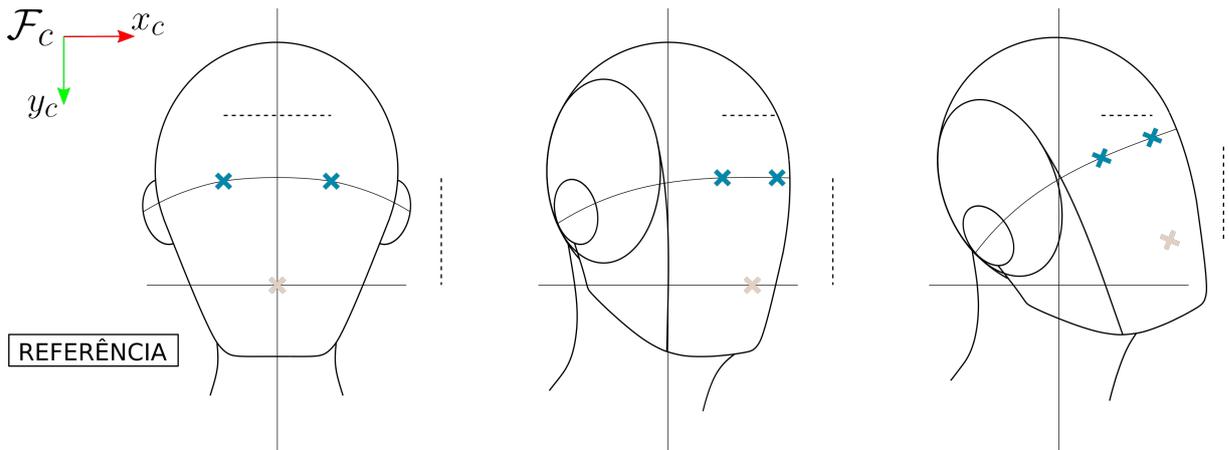


Figure 3.14: Distâncias horizontal entre os olhos e vertical entre olhos e boca, medidas com respeito a \mathcal{F}_c , variam com o movimento da cabeça, assim como a posição da boca.

Considerando o sistema de coordenadas \mathcal{F}_h localizado na cabeça do humano, com eixo z_h apontando para frente e eixo x_h apontando para a sua direita, quando a cabeça é rotacionada ao redor de seu eixo y_h (olhando para a direita ou para a esquerda), a distância horizontal entre os olhos projetada na imagem captada pela câmera é reduzida em relação à referência, como mostrado na Figura 3.15. Seja $D_{ox,ref}$ a distância absoluta horizontal de referência entre os olhos e D_{ox} a distância absoluta atual, obtida a cada instante, o ângulo de rotação θ em torno do eixo y_h de \mathcal{F}_h é dado por

$$\theta = \arccos\left(\frac{D_{ox}}{D_{ox,ref}}\right). \quad (3.11)$$

Para determinar se o movimento é para a direita ou para a esquerda, o deslocamento horizontal da boca é utilizado: se a translação da boca, partindo da referência para a posição atual, é no sentido positivo de x_c de \mathcal{F}_c , então a pessoa se virou para a sua esquerda, ou seja, o movimento foi na direção negativa de x_h ; se o movimento da boca foi para a esquerda da câmera (negativo de x_c), então a pessoa se virou para a sua direita (positivo de x_h). Sejam d_{bx} e d_{by} os deslocamentos nos eixos x_c e y_c de \mathcal{F}_c do ponto de referência da boca $\mathbf{p}_{b,ref}$ para o ponto atual \mathbf{p}_b , a rotação em torno do eixo y_h é dada por

$$\mathbf{r}_y = \begin{cases} \cos\left(\frac{-\theta}{2}\right) + \hat{j} \sin\left(\frac{-\theta}{2}\right), & \text{se } d_{bx} > 0 \\ \cos\left(\frac{\theta}{2}\right) + \hat{j} \sin\left(\frac{\theta}{2}\right), & \text{se } d_{bx} < 0 \\ 1, & \text{se } d_{bx} = 0. \end{cases} \quad (3.12)$$

Para a rotação ao redor do eixo x_h , a mesma abordagem é aplicada à distância vertical entre olhos e boca na imagem do rosto. A média das coordenadas verticais dos centros das íris é utilizada como coordenada vertical dos olhos para se calcular as distâncias absolutas $D_{ob,ref}$ e D_{ob} , de referência e atual, respectivamente, entre olhos e boca, medidas ao longo

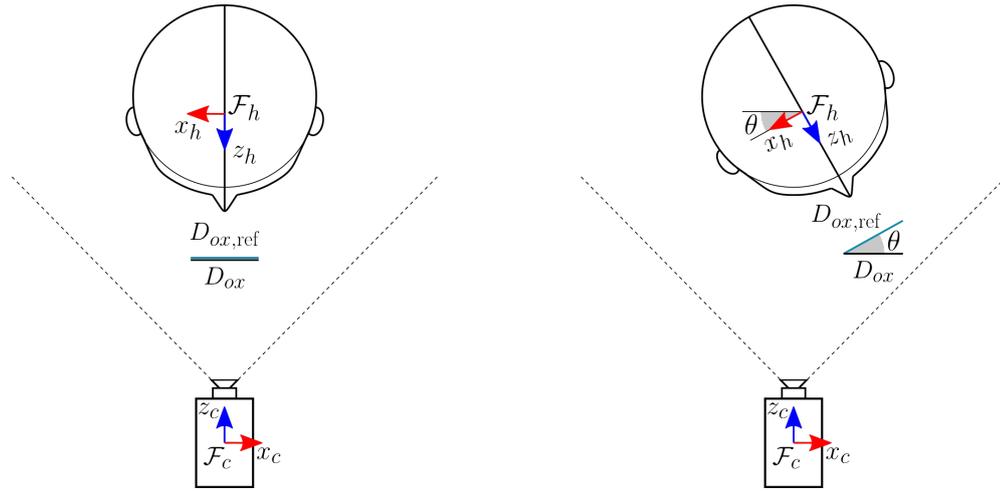


Figure 3.15: Diferença na distância D_{ox} entre os olhos medida no eixo x_c de \mathcal{F}_c quando o humano rotaciona a cabeça ao redor de seu eixo y_h .

de y_c , como ilustrado na Figura 3.16. O ângulo de rotação ϕ da cabeça ao redor de x_h de \mathcal{F}_h é dado por

$$\phi = \arccos\left(\frac{D_{ob}}{D_{ob,ref}}\right). \quad (3.13)$$

O deslocamento vertical da boca é utilizado para determinar a direção do movimento, para cima ou para baixo, de modo que a rotação \mathbf{r}_x da cabeça ao redor do eixo x_h é

$$\mathbf{r}_x = \begin{cases} \cos\left(\frac{-\phi}{2}\right) + \hat{j} \sin\left(\frac{-\phi}{2}\right), & \text{se } d_{by} > 0 \\ \cos\left(\frac{\phi}{2}\right) + \hat{j} \sin\left(\frac{\phi}{2}\right), & \text{se } d_{by} < 0 \\ 1, & \text{se } d_{by} = 0. \end{cases} \quad (3.14)$$

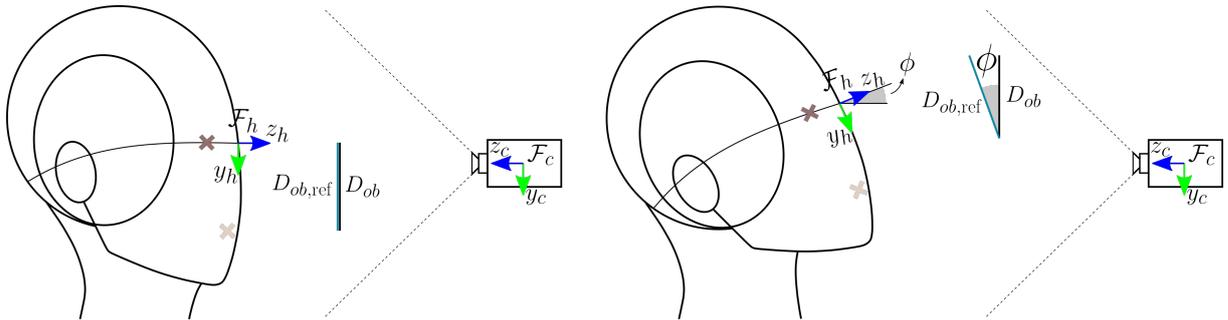


Figure 3.16: Diferença na distância D_{ob} entre olhos e boca medida no eixo y_c de \mathcal{F}_c quando o humano rotaciona sua cabeça ao redor de seu eixo x_h .

O sistema de coordenadas de referência \mathcal{F}_{ref} , definido quando a pessoa olha de frente para a câmera, é rotacionado de \mathbf{r}_y seguido de \mathbf{r}_x , utilizando-se o sistema de coordenadas móvel, para se obter a orientação de \mathcal{F}_h . As posições de \mathcal{F}_{ref} e \mathcal{F}_h são dadas pela posição da cabeça $\mathbf{p}_{cabeça}$, conhecida a todo instante. A direção do olhar é, portanto, paralela à direção positiva do eixo z_h , ou seja, a direção da linha do olhar é $\mathbf{l}_{olhar}^h = \hat{k}$, do ponto de

vista de \mathcal{F}_h . Caso a transformação entre o ponto da cabeça $\mathbf{p}_{\text{cabeça}}$ e o ponto entre os olhos da pessoa $\mathbf{p}_{\text{olhos}}$ seja conhecida, o ponto $\mathbf{p}_{\text{olhos}}$ pode ser usado para definir a linha do olhar. Para a aplicação neste trabalho, que considera objetos e regiões grandes e não muito próximos do rosto da pessoa, a aproximação $\mathbf{p}_{\text{olhos}} = \mathbf{p}_{\text{cabeça}}$ é utilizada. Com $\mathbf{p}_{\text{olhos}}$ e $\mathbf{l}_{\text{olhar}}$ descritos em relação a um mesmo referencial, a linha do olhar é representada por $\mathbf{l}_{\text{olhar}} = \mathbf{l}_{\text{olhar}} + \varepsilon (\mathbf{p}_{\text{olhos}} \times \mathbf{l}_{\text{olhar}})$.

As análises geométricas para determinação da direção do olhar foram implementadas em Python e a mesma metodologia utilizada na interpretação dos gestos de apontar, descrita na Seção 3.2, é usada para a interpretação do foco de atenção da pessoa, a partir da linha $\mathbf{l}_{\text{olhar}}$ e dos objetos e regiões conhecidas.

Não são consideradas rotações em torno de z_h , porque tal movimento causaria mudanças nas distâncias D_{ox} e D_{ob} não relacionadas a rotações ao redor de x_h e y_h , já que, por simplificação, D_{ox} e D_{ob} são medidas no sistema de coordenadas da câmera e são relacionadas às projeções dos segmentos de linha nos eixos x_c e y_c , respectivamente. Assume-se também que as rotações da cabeça são de baixa amplitude, para evitar oclusão dos atributos faciais, e que não há translação da cabeça em relação aos eixos x_c e y_c , para que seja possível usar o deslocamento da boca para determinar a direção das rotações. Pequenos movimentos de translação da pessoa ao longo de z_c são permitidos, isto é, se afastar ou se aproximar da câmera. Para isso, as distâncias de referência, $D_{ox,\text{ref}}$ e $D_{ob,\text{ref}}$, são atualizadas a todo instante, utilizando a distância perpendicular entre cabeça e câmera no instante que as primeiras referências são obtidas ($D_{z,\text{inicial}}$) e no instante atual ($D_{z,\text{atual}}$), obtendo-se as distâncias de referência atuais $D_{\text{ref},\text{atual}} = D_{\text{ref},\text{inicial}}(D_{z,\text{inicial}}/D_{z,\text{atual}})$. Por fim, o algoritmo não detecta movimentos apenas dos olhos, então os experimentos foram planejados de modo a induzir as pessoas a movimentarem a cabeça ao olhar para pontos de interesse.

3.4 Agente virtual

O sistema desenvolvido para interagir com o humano se apresenta na forma de uma agente virtual. Se comparado ao desenvolvimento de um sistema robótico físico, a utilização de agentes virtuais tem custo e complexidade reduzidos, características relevantes nas escolhas das tecnologias de comunicação implementadas para este trabalho. A agente virtual é equipada com voz, para fornecer informações pertinentes durante a interação, é capaz de transmitir emoções através de expressões faciais e direciona seu olhar para pontos específicos do ambiente para informar à pessoa sobre o seu foco de atenção. O desenvolvimento de cada uma dessas funcionalidades é detalhado a seguir.

3.4.1 Voz

A comunicação explícita da agente virtual com a pessoa durante a interação é através de voz, para se apresentar, dar instruções ou informar sobre problemas, por exemplo. Como os experimentos envolvem interação de curta duração e com objetivos específicos, as ações da pessoa e da agente virtual são consideradas, de maneira geral, previsíveis, possibilitando a definição de um conjunto limitado de frases disponíveis para serem usadas nos momentos adequados.

A solução escolhida para a criação da voz foi a gravação em voz humana de frases selecionadas. Para evitar o reconhecimento da voz durante o experimento, o que poderia causar estranhamento ou a criação de um modelo mental impreciso (por exemplo, supor que a agente virtual seja teleoperada), os áudios foram modificados. As imagens criadas para o rosto da agente virtual incluem simulações de movimentos da boca durante a fala.

3.4.2 Rosto e expressões faciais

Para representar a agente virtual e transmitir, de maneira implícita, emoções e informações sobre seu estado interno, foram criadas imagens de um rosto para serem exibidas na tela de um computador. A aplicação da agente virtual foi implementada em Python, utilizando as bibliotecas OpenCV e Pillow⁴ para manipulação das imagens e o pacote `tkinter`⁵ para exibição na tela.

Kalegina et al. (2018) realizaram um estudo sobre a percepção de humanos em relação a diferentes rostos virtuais de robôs, variando desde um rosto composto por apenas um círculo até representações mais realistas, com muitos detalhes. As autoras investigaram como cada atributo facial influencia na percepção. De acordo com seus resultados, rostos sem boca são considerados menos confiáveis e menos amigáveis e a presença de sobrancelhas não afeta significativamente esses aspectos. A agente virtual criada para este trabalho tem dois olhos, sobrancelhas e boca, com formas simples e pouco realistas, como mostrado na Figura 3.17. A representação com poucos detalhes foi utilizada para simplificar a implementação e evitar a possível influência do chamado vale da estranheza (em inglês, *uncanny valley*) — quando um robô tem características muito semelhantes a de humanos e isso causa estranhamento e desconforto durante a interação. Seus olhos são circulares e com íris coloridas e suas expressões faciais são caracterizadas a partir de variações nas sobrancelhas e na boca. Apesar dos resultados de Kalegina et al. (2018) indicarem que olhos sem pupila foram significativamente percebidos como menos confiáveis e amigáveis, a opção sem pupila utilizada (círculos brancos sobre um fundo preto) é bem diferente da opção com íris e sem pupila desenvolvida para este trabalho.

As expressões criadas, mostradas na Figura 3.17, são: neutra, feliz, satisfeita, triste,

⁴<https://pillow.readthedocs.io/en/stable/>

⁵<https://docs.python.org/3/library/tkinter.html>

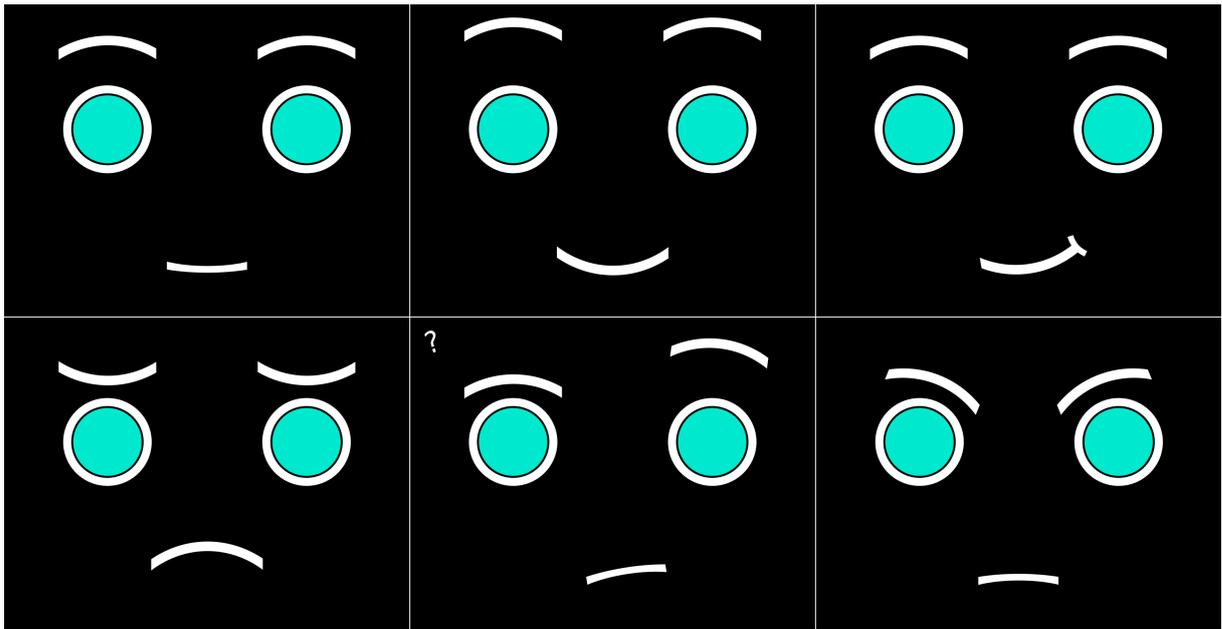


Figure 3.17: Expressões faciais da agente virtual. Em sentido *horário*: expressões neutra, feliz, satisfeita, com raiva, confusa e triste.

confusa e com raiva. O reconhecimento das expressões foi avaliado através de um formulário, aplicado a um grupo com interação com a agente virtual e a outro sem interação, e os resultados são mostrados no Apêndice A.3. Além das expressões, há uma imagem que simula um levantar de sobrancelhas, que pode ser usado como uma comunicação implícita em alguns momentos, por exemplo para chamar a atenção da pessoa durante a interação. A transição entre as expressões é abrupta, mas, para transmitir vivacidade, os olhos da agente virtual piscam e, quando ela fala, sua boca se move. Para simular o movimento da boca durante a fala foram criadas três opções de bocas abertas que são selecionadas aleatoriamente durante a execução do áudio de voz (ver Figura 3.18). Não houve preocupação em criar formatos de boca específicos e associá-los adequadamente a cada sílaba, imitando os movimentos feitos por uma pessoa ao pronunciar cada palavra, já que não há intenção de que a agente virtual seja altamente realista.

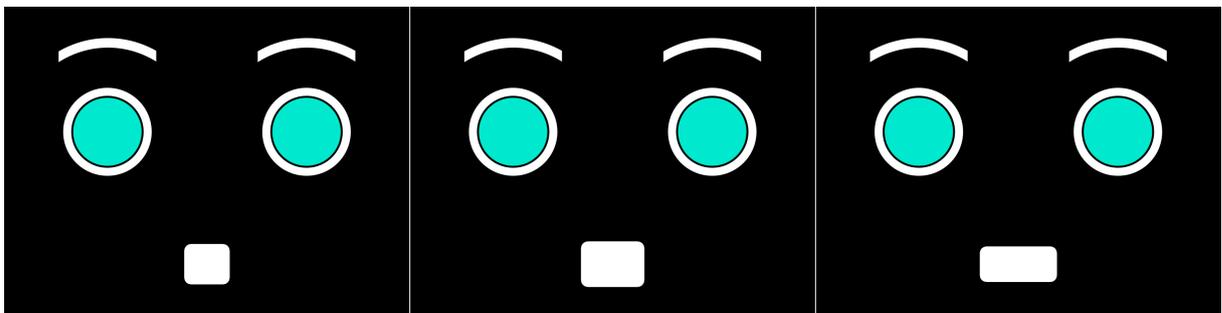


Figure 3.18: Diferentes formatos de boca para quando a agente virtual fala.

3.4.3 Direção do olhar

A agente virtual é capaz de direcionar seu olhar para pontos específicos, como o rosto da pessoa ou objetos da cena, para implicitamente transmitir informações sobre seu foco de atenção. O movimento é virtual e a simulação é feita com geometria projetiva, utilizando a biblioteca DQ Robotics para a manipulação matemática e funções de transformações geométricas da biblioteca OpenCV para transformar a imagem do rosto.

Seja $\mathcal{F}_{\text{olhos}}$ um sistema de coordenadas posicionado no ponto médio entre os olhos do rosto virtual, como mostrado na Figura 3.19, e π_{ref} (marrom claro na figura) o plano da tela que exibe a imagem do rosto (delimitada pelo retângulo de contorno marrom escuro na figura). Um ponto \mathbf{p}_{ref} , descrito em relação a $\mathcal{F}_{\text{olhos}}$, à frente dos olhos (direção negativa de z_{olhos}) é definido como referência, isto é, se a agente virtual olha para esse ponto, a imagem do rosto é a original, como uma das mostradas na Figura 3.17. O ponto para onde a agente virtual deve olhar é representado por $\mathbf{p}_{\text{final}}$, também descrito em relação a $\mathcal{F}_{\text{olhos}}$. Os deslocamentos de $\mathbf{p}_{\text{ref}} = z_r \hat{k}$ para $\mathbf{p}_{\text{final}} = x_f \hat{i} + y_f \hat{j} + z_f \hat{k}$ nos eixos x_{olhos} , y_{olhos} e z_{olhos} são dados por $d_{px} = x_f$, $d_{py} = y_f$ e $d_{pz} = z_f - z_r$, respectivamente. As distâncias absolutas entre o ponto médio dos olhos e os pontos de referência e final são dadas, respectivamente, por D_{ref} e D_{final} . O sistema de coordenadas original $\mathcal{F}_{\text{olhos}}$ é rotacionado para que o negativo de seu eixo z_{olhos} aponte para $\mathbf{p}_{\text{final}}$. Os ângulos de rotação, mostrados na Figura 3.19, são

$$\theta = \arctan\left(\frac{-d_{px}}{D_{\text{ref}} - d_{pz}}\right) \text{ e } \phi = \arcsin\left(\frac{d_{py}}{D_{\text{final}}}\right),$$

em torno de y_{olhos} e x_{olhos} móveis, respectivamente. As rotações são, portanto, dadas por $\mathbf{r}_y = \cos(\theta/2) + \hat{j} \sin(\theta/2)$ e $\mathbf{r}_x = \cos(\phi/2) + \hat{i} \sin(\phi/2)$.

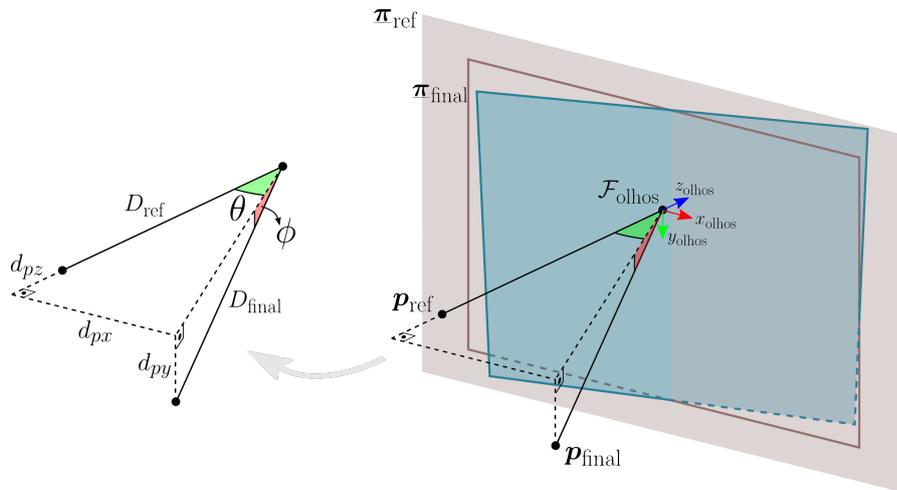


Figure 3.19: O sistema de coordenadas $\mathcal{F}_{\text{olhos}}$ é rotacionado de θ ao redor do eixo y e de ϕ ao redor do eixo x para que o negativo do eixo z do sistema de coordenadas rotacionado aponte para $\mathbf{p}_{\text{final}}$.

A normal do plano π_{ref} está na direção negativa de z_{olhos} , enquanto π_{final} (em azul na Figura 3.19), que representa o plano da tela se houvesse um movimento físico real, tem normal na direção negativa do eixo z do sistema de coordenadas rotacionado. Os deslocamentos entre o ponto médio dos olhos e cada um dos cantos da imagem original do rosto (vértices do retângulo marrom escuro na Figura 3.20) são conhecidos e esses mesmos vértices são obtidos no plano rotacionado (pontos azuis na Figura 3.20). Linhas que conectam o ponto \mathbf{p}_{ref} aos vértices no plano π_{final} são definidas e os pontos de cruzamento entre elas e o plano da tela π_{ref} são obtidos (pontos marrons na Figura 3.20). Esse pontos podem estar dentro ou fora das dimensões da imagem (retângulo marrom escuro na Figura 3.20) e são entradas para as funções de OpenCV que transformam a imagem original do rosto. A Figura 3.21 mostra um exemplo de uma imagem antes e depois de uma transformação, com os pontos marrons indicando os cruzamentos das linhas com o plano da tela.

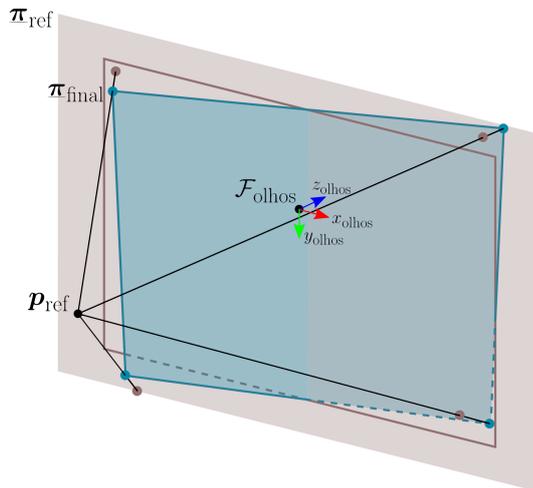
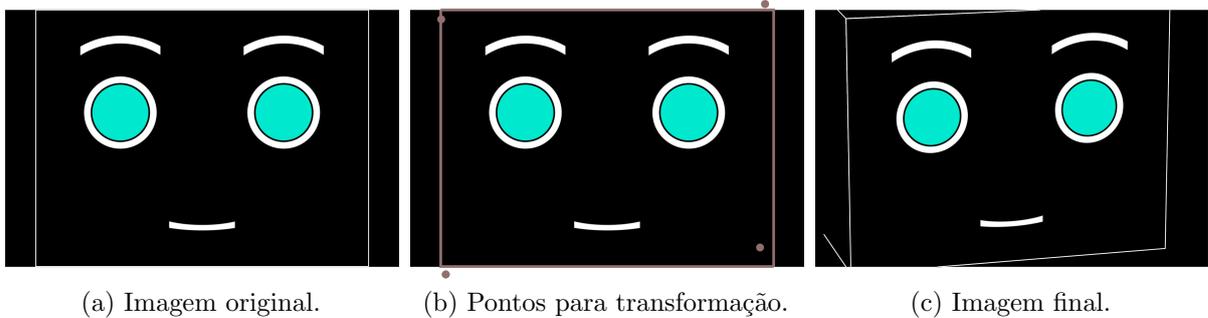


Figure 3.20: As linhas que conectam \mathbf{p}_{ref} e os vértices do retângulo azul no plano π_{final} cruzam o plano π_{ref} nos pontos marrons. O retângulo original, que delimita a imagem original do rosto, está representado pelo contorno *marrom*, enquanto o retângulo rotacionado possui contorno *azul*.



(a) Imagem original.

(b) Pontos para transformação.

(c) Imagem final.

Figure 3.21: Exemplo da transformação de uma imagem para simular a direção do olhar da agente virtual. Os pontos marrons representam os pontos de cruzamento entre linhas e o plano da tela e são onde os vértices da imagem devem estar após a transformação.

A transformação das imagens utilizando geometria projetiva é adequada apenas para rotações de ângulos menores que 90° . Para aumentar o alcance do olhar da agente virtual, foram criadas algumas imagens que simulam o movimento apenas dos olhos. Tendo em mente os experimentos de interação deste trabalho, quatro opções de imagens desse tipo foram criadas e são mostradas na Figura 3.22. Com as imagens usadas em conjunto com as transformações de geometria projetiva, pode-se simular o olhar para pontos mais próximos, sobre ou até mesmo atrás do plano da tela. O algoritmo descrito anteriormente não considera a utilização dessas imagens, mas elas podem ser selecionadas manualmente, de acordo com as necessidades da aplicação.

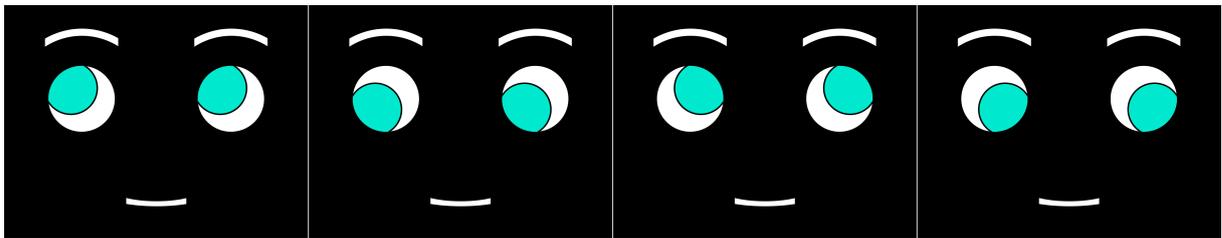


Figure 3.22: Diferentes opções de olhos para aumentar o alcance do olhar da agente virtual.

3.5 Outras comunicações

Além das comunicações descritas nas seções anteriores, outras modalidades foram incluídas no sistema, de acordo com a interação proposta. Para que a pessoa possa inserir informações explícitas e para que a agente virtual possa atualizar o andamento de tarefas, foram criadas aplicações em tela, com entradas manuais através de mouse e teclado. O sistema, representado pela agente virtual, também se comunica explicitamente por sinais sonoros, para indicar acertos e erros. O posicionamento da pessoa pelo ambiente também pode ser fonte de informação implícita, para indicar, por exemplo, qual a etapa atual da interação, quais os próximos movimentos ou se instruções dadas foram seguidas. Como as juntas da cadeia cinemática do humano estão disponíveis (ver Seção 3.2), essa informação poderia ser usada, junto com um mapa do ambiente, para a determinação da localização da pessoa durante a interação. Outra opção, porém, é a utilização de marcadores espalhados pelo ambiente. Se colocados em posições adequadas, pode-se definir determinados marcadores que, quando não detectados pelas câmeras de supervisão, indicam que a pessoa está em uma localização específica, que causa a oclusão desses marcadores selecionados. Como o ambiente dos experimentos de interação já contava com infraestrutura que facilitava a sua implementação, a solução com marcadores é utilizada.

3.6 Conclusão

Este capítulo descreveu o desenvolvimento de uma infraestrutura de comunicação multimodal para a realização de experimentos de interação entre humano e agente virtual. As metodologias de comunicação explícita e implícita deveriam ser acessíveis e de simples implementação, além de não necessitarem que as pessoas usem equipamentos e trajes especiais. A Tabela 3.1 lista todos os tipos de comunicação disponíveis.

Table 3.1: Tipos de comunicações disponíveis para humano e agente virtual.

	Humano	Agente virtual
Explícitas	gestos entradas manuais	VOZ sinais sonoros informações em tela
Implícitas	direção do olhar localização no ambiente	expressões faciais levantar de sobrancelhas direção do olhar

Os subsistemas foram implementados para serem integrados via ROS e se comunicam entre si por tópicos. Módulos extras, particulares de cada aplicação, também podem ser conectados à rede para se comunicarem com os subsistemas. O sistema completo foi montado em um dos laboratórios vinculados ao grupo MACRO (*Mechatronics, Control and Robotics*) da UFMG, ambiente destinado à realização dos experimentos de interação.

4

PLANEJAMENTO EXPERIMENTAL

A infraestrutura de comunicação descrita no capítulo anterior foi desenvolvida para ser utilizada em experimentos de interação entre humano e agente virtual. O objetivo dos experimentos é investigar os efeitos do tipo de comunicação na interação, tanto em fatores relacionados à tarefa quanto na percepção das pessoas sobre a interação e a agente virtual. A literatura sobre interação humano-robô, particularmente os trabalhos descritos na Seção 2.1.2, sugere que a combinação de comunicações explícitas e implícitas traz benefícios à interação. São então definidas duas configurações de comunicação:

EX: apenas comunicações explícitas de humano e agente virtual;

EXIM: comunicações explícitas e implícitas de humano e agente virtual.

Este capítulo apresenta as hipóteses específicas deste trabalho, o experimento de interação proposto para avaliá-las e as análises de dados destinadas a cada variável.

4.1 Hipóteses

Se a combinação de comunicações explícitas e implícitas torna as atitudes dos agentes mais transparentes e, portanto, mais previsíveis, o que é importante para uma ação conjunta e, conseqüentemente, uma tarefa bem sucedida, espera-se que tarefas sejam executadas em menor tempo e com menor número de erros:

H1: O tempo de execução da tarefa é menor na configuração EXIM do que na configuração EX.

H2: O número de erros cometidos pela pessoa é menor na configuração EXIM do que na configuração EX.

Se a agente virtual utiliza comunicações implícitas em conjunto com comunicações explícitas, assemelhando-se a agentes humanos, e aparenta ser mais consciente da interação por utilizar as informações implícitas das pessoas, espera-se que elas a entendam melhor e a percebam mais como uma agente social, tornando a interação mais natural e mais agradável:

H3: A agente virtual é mais aceita na configuração EXIM do que na configuração EX.

H4: A agente virtual é percebida como mais sociável na configuração EXIM do que na configuração EX.

H5: A agente virtual é percebida como mais transparente na configuração EXIM do que na configuração EX.

Por fim, espera-se que com a maior sociabilidade da agente virtual e o melhor desempenho na tarefa com a combinação de comunicações explícitas e implícitas, a interação seja percebida como mais eficiente:

H6: A interação é percebida como mais eficiente na configuração EXIM do que na configuração EX.

4.2 Interação proposta

Para a avaliação das hipóteses definidas, propõe-se uma atividade semelhante a um jogo. Para a escolha da atividade, levou-se em consideração a infraestrutura de comunicação disponível e a duração total da interação, que não deveria ser muito longa para evitar que as pessoas se cansassem. Além disso, desejava-se que as tarefas pudessem ser repetidas com pequenas variações, para que participantes realizassem a mesma atividade nas duas configurações de comunicação, e que na configuração EXIM as comunicações implícitas tanto de humano quanto de agente virtual pudessem fornecer informações que ajudassem na execução das tarefas.

A atividade proposta é dividida em duas fases. Na primeira fase, após se apresentar e dar instruções, a agente virtual mostra uma sequência de quatro cores e a pessoa deve apontar para objetos coloridos espalhados no ambiente na ordem correta. Quando uma indicação correta é reconhecida, um sinal sonoro de acerto é dado e a cor é atualizada numa aplicação em tela; caso a indicação esteja errada, um sinal sonoro de erro é emitido. A sequência de cores funciona como uma senha para se avançar para a segunda fase, em que a pessoa deve contar quantas vezes determinados objetos aparecem em imagens de objetos embaralhados. São quatro imagens de objetos embaralhados fixadas no ambiente e a pessoa é instruída a preencher os campos na tela após a contagem de cada um dos

objetos indicados. Em ambas as fases, se um tempo limite é alcançado (quatro minutos para cada fase), a agente virtual termina a tarefa, preenchendo a senha ou inserindo os valores de contagem. O Apêndice B apresenta mais detalhes sobre a interação.

Cada configuração de comunicação é representada por uma agente virtual e no início da execução o sistema define aleatoriamente qual será utilizada para cada uma. São duas agentes virtuais, Luna e Sofia, mostradas na Figura 4.1, que são diferenciadas pelo nome, pela cor dos olhos e pelo tom de voz. Participantes interagem com as duas agentes virtuais (e as duas configurações de comunicação) e a ordem também é definida aleatoriamente pelo sistema no início da participação.

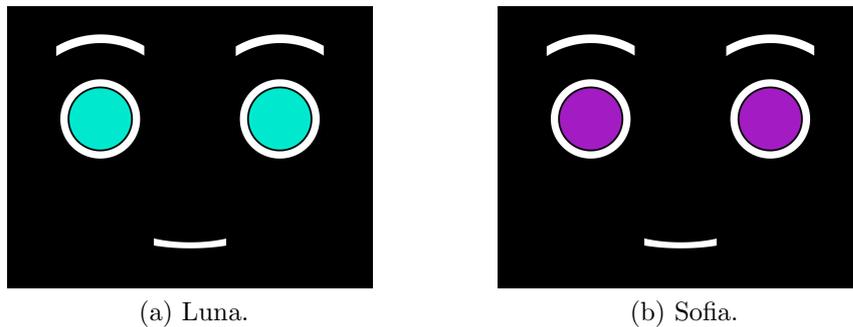


Figure 4.1: Duas agentes virtuais criadas para interagirem com as pessoas durante os experimentos. Cada agente virtual é usada em uma configuração de comunicação e essa correspondência é definida aleatoriamente pelo sistema para cada participante.

Na configuração EX, a agente virtual usa apenas voz, sinais sonoros e informações em tela para se comunicar, mantendo sempre a expressão neutra e o olhar fixo à frente, e utiliza as informações explícitas da pessoa, através dos gestos de apontar na fase 1 e das entradas manuais na fase 2. Na configuração EXIM, além das comunicações explícitas, a agente virtual tem expressões faciais adequadas durante a interação (por exemplo, fica feliz quando a pessoa acerta ou termina, triste quando há algum erro e confusa quando percebe algum problema) e direciona seu olhar pelo ambiente. Na fase 1 de EXIM, a agente virtual sempre olha para o objeto que deve ser indicado naquele momento e, caso a pessoa cometa um erro ou a senha seja repetida, a agente virtual olha para ela e ergue as sobrancelhas, chamando sua atenção, antes de olhar novamente para o objeto correto. Além disso, a agente virtual antecipa o comando de início da fase 2 quando percebe que a pessoa se posicionou no local indicado. Durante a fase 2 da configuração EXIM, se a direção do olhar da pessoa é identificada enquanto ela realiza a contagem em uma das imagens, a agente virtual segue o olhar e diz o valor correto em forma de sugestão (*e.g.*, “*Eu acho que aí são cinco.*”) e sempre que a pessoa abre um campo de preenchimento na tela, a agente virtual olha para ele. Durante toda a interação na configuração EXIM, a agente virtual olha para a pessoa enquanto fala com ela, exceto quando está dando uma dica na fase 2 (nesse caso, o olhar da agente virtual é usado para indicar a qual objeto da contagem ela se refere). Em ambas as configurações de comunicação, a agente virtual move a boca

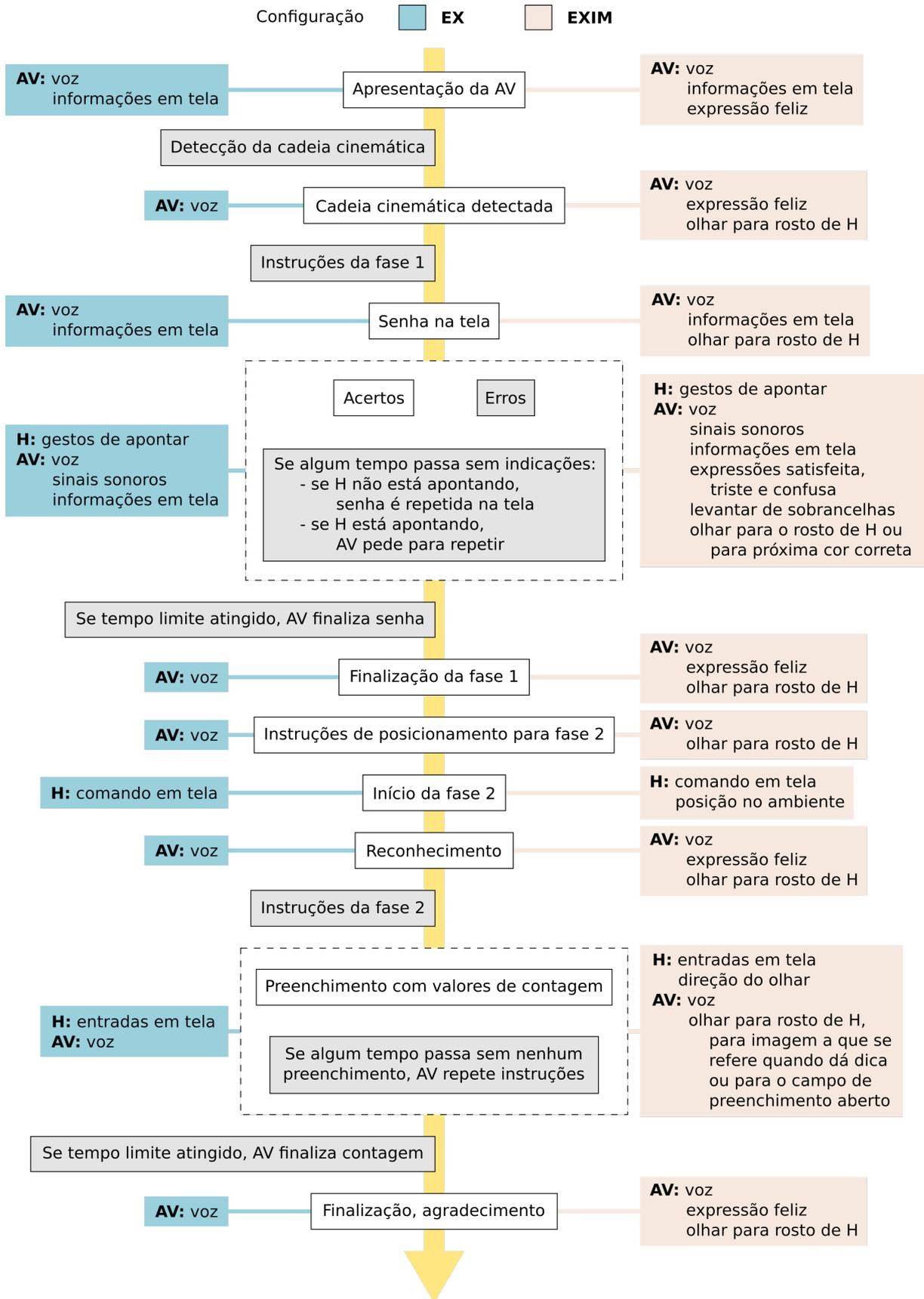


Figure 4.2: Fluxo da interação entre humano (H) e uma das agentes virtuais (AV). Caixas em cinza indicam etapas que não são necessariamente executadas. Em azul, são mostradas as comunicações usadas em cada etapa na configuração EX e em rosa, na configuração EXIM.

ao falar e pisca os olhos, para transmitir vivacidade. O significado das comunicações da agente virtual não foi informado previamente a participantes. A Figura 4.2 resume as etapas da interação e as comunicações usadas em cada uma das configurações.

Os objetos para a fase 1 são caixas de quatro cores (azul, amarelo, branco e preto). A pose desses objetos é obtida utilizando-se marcadores e um sistema de câmeras existente no laboratório destinado aos experimentos. Os objetos são grandes e os seus modelos virtuais ainda consideram dimensões um pouco maiores, para facilitar o reconhecimento dos gestos durante a interação. As imagens de contagem para a fase 2, fixadas em uma das paredes do ambiente, também são regiões de interesse de dimensões grandes e suas poses fixas são determinadas manualmente. Considerando a posição que a pessoa é instruída a ocupar para a fase 2, as imagens de contagem foram posicionadas de modo a incentivar que se movimente a cabeça ao olhar para elas. O ambiente dos experimentos é mostrado no Apêndice B.

4.3 Medidas e instrumentos de medição

As medidas objetivas da interação, relacionadas com as hipóteses H1 e H2, são tempo e número de erros. O tempo para realização da tarefa, incluindo as duas fases, é registrado pelo sistema, excluindo-se tempos de execução de áudios de voz da agente virtual, para evitar influência das leves diferenças entre as falas de cada agente virtual na comparação entre as configurações (ver Apêndice B), assim como os tempos necessários para instruções da fase, que são dadas apenas na primeira execução da tarefa. Tempos relacionados a inicialização ou solução de problemas dos sistemas de comunicação (detecção da cadeia cinemática do humano e obtenção dos pontos de referência no rosto para o rastreamento da direção do olhar) também não são considerados. Os erros contabilizados na fase 1 são de indicações de cores incorretas e na fase 2 são valores errados inseridos nos campos de resposta na aplicação em tela. Eventuais erros causados por falhas nos sistemas (por exemplo, identificação incorreta de gesto de apontar) são descartados. Após a realização dos experimentos, definiu-se que sempre que o tempo limite da fase fosse atingido antes da finalização da tarefa, um erro seria somado para cada cor não adicionada na fase 1 e para cada campo não preenchido na fase 2.

Já as medidas subjetivas, relacionadas com as hipóteses H3, H4, H5 e H6, são divididas em quatro variáveis: aceitação, transparência e sociabilidade das agentes virtuais e percepção da eficiência das interações. Cada uma delas é medida através de uma escala Likert (Likert, 1932), isto é, um conjunto de *itens* (afirmações) para os quais participantes devem responder com uma dentre as seguintes opções: (1) discordo totalmente, (2) discordo, (3) não sei, (4) concordo e (5) concordo totalmente. São, portanto, cinco *níveis* de resposta, que são mostrados nos questionários sempre na mesma ordem e, neste trabalho, sem estarem associados aos números. A Tabela 4.1 mostra todos os itens que formam as escalas.

Table 4.1: Itens das escalas Likert para medição de cada uma das variáveis de percepção. Os itens marcados com (R) são tratados em escala reversa. O termo AV é substituído pelo nome da agente virtual.

Aceitação da agente virtual	<ol style="list-style-type: none"> 1. Eu achei a AV intimidadora. (R) 2. Eu achei a AV amigável. 3. Eu me senti confortável enquanto interagia com a AV. 4. Eu gostei de interagir com a AV. 5. Eu achei desagradável interagir com a AV. (R) 6. Eu gostaria de interagir mais com a AV.
Sociabilidade da agente virtual	<ol style="list-style-type: none"> 1. Eu senti que a AV entendia o que eu estava fazendo. 2. Ao interagir com a AV, eu senti como se estivesse com uma pessoa real. 3. Às vezes eu sentia que a AV estava realmente olhando para mim. 4. Às vezes a AV parecia ter sentimentos reais. 5. O comportamento da AV é similar ao de uma pessoa.
Transparência da agente virtual	<ol style="list-style-type: none"> 1. Eu entendia a AV. 2. Eu conseguia saber o que a AV estava “pensando”. 3. Eu sabia quando a AV estava prestando atenção em mim. 4. Durante a interação, as intenções da AV eram claras para mim.
Percepção de eficiência da interação	<ol style="list-style-type: none"> 1. Eu podia contar com a AV para me ajudar durante a tarefa. 2. AV me ajudou a executar a tarefa. 3. AV me atrapalhou. (R) 4. AV não fez diferença no meu desempenho. (R)

Os itens foram definidos considerando o contexto específico e foram inspirados e adaptados de trabalhos como os de Heerink et al. (2010) e Iwasaki et al. (2019). Na maioria dos casos, o nível mais baixo de resposta (“discordo totalmente”) indica percepção mais negativa da variável, por exemplo, discordar do item 1 da escala de transparência indica que a agente virtual foi considerada pouco transparente. Por outro lado, para o item 1 da escala de aceitação, uma resposta em níveis mais baixos indica que a agente virtual foi mais aceita. Esses casos de escala reversa, em que níveis mais baixos indicam respostas mais positivas, são indicados por um (R) junto às sentenças na Tabela 4.1.

Um questionário de avaliação contendo todos os 19 itens é aplicado após a interação com cada uma das agentes virtuais e o termo AV (agente virtual) nas sentenças listadas é substituído pelo nome da agente virtual que a pessoa indica ter acabado de interagir (Luna ou Sofia). Os questionários são sempre preenchidos virtualmente e os itens são apresentados em ordem aleatória para cada participante sem a separação entre as variáveis e sem a indicação de escala reversa. Após as duas interações e o preenchimento dos dois questionários de avaliação, um questionário final é aplicado. Nele, participantes respondem com qual agente virtual preferiram interagir e podem adicionar observações e comentários sobre o experimento e as interações com as agentes virtuais. Idade, gênero (opções *Feminino*, *Masculino*, *Prefiro não dizer* e *Outros*, sendo a última com campo para preenchimento), escolaridade (opções de *Ensino fundamental incompleto* a *Doutorado completo*) e familiaridade com agentes virtuais como Luna e Sofia (opções *Sim*, *Não* e *Não sei responder*) também são coletados através do questionário final, para permitir a discussão da influência desses fatores no desempenho e nas percepções das pessoas nas diferentes configurações de comunicação. O questionário final também contém uma seção para avaliação do reconhecimento das expressões faciais criadas para as agentes virtuais (ver Apêndice A.3).

4.4 Ferramentas de análises estatísticas

A literatura de diferentes áreas de estudo está repleta de resultados embasados em valores- p , obtidos através de testes de significância de hipótese nula. Segundo Wagenmakers et al. (2018) e Kelter (2020), esse método é muito utilizado em pesquisas médicas e em psicologia, por exemplo, além de também ser comum nos estudos de interação humano-robô (todos os estudos comparativos citados na Seção 2.1.2 reportaram valores- p). O procedimento faz parte da abordagem *frequentista*, que se baseia em distribuições amostrais e interpreta probabilidades como frequências relativas (Montgomery & Runger, 2011; Kruschke, 2015).

O método de testes de significância de hipótese nula (do inglês, *Null Hypothesis Significance Testing* - NHST) utiliza um conjunto de dados aleatório, isto é, uma *amostra* aleatória, para testar afirmações sobre um parâmetro que descreve uma variável de interesse. Uma hipótese nula define um valor específico para o parâmetro, que indique a ausência

do efeito que se deseja medir ou ausência de novidade sobre algum conhecimento. Por exemplo, ao se comparar se uma nova versão de um sistema tem desempenho melhor que a anterior, a hipótese nula descreve a ausência de diferença no desempenho entre as versões; ao investigar se o valor de um parâmetro sofreu alteração em relação a um valor de referência conhecido, a hipótese nula define que o valor do parâmetro não mudou. Calcula-se então uma estatística de teste, *i.e.*, uma função do conjunto de dados utilizada em testes de hipóteses, cuja ideia é ser um resumo da amostra (Montgomery & Runger, 2011; Kruschke & Liddell, 2018). Diferentes amostras geram diferentes estatísticas de teste, que é, portanto, uma variável aleatória. O método NHST então se pergunta qual a probabilidade de se obter uma estatística de teste pelo menos tão extrema, ou seja, tão distante do valor considerado mais provável caso a hipótese nula seja verdadeira, quanto a calculada com a amostra disponível. Se essa probabilidade for considerada alta, em relação a um limite predeterminado, isso significa que a estatística de teste obtida é bastante provável sob a hipótese nula, que, portanto, não deve ser rejeitada; por outro lado, se a probabilidade for considerada baixa, conclui-se que a hipótese nula deve ser rejeitada, porque seria pouco provável obter-se a estatística de teste calculada com a amostra se o valor do parâmetro fosse igual ao da hipótese nula. A probabilidade de se obter um valor tão ou mais extremo quanto o obtido com a amostra disponível é dada pelo *valor-p*.

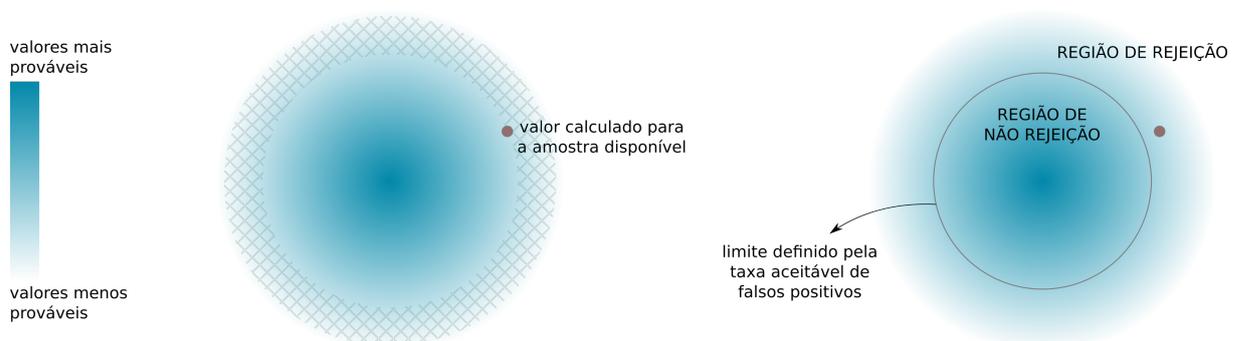


Figure 4.3: Nuvem de possibilidades da estatística de teste, assumindo hipótese nula verdadeira. A área hachurada corresponde a valores tão ou mais extremos que o valor observado na amostra disponível e a taxa aceitável de falsos positivos define regiões de rejeição e não rejeição da hipótese nula.

Por exemplo, ao se jogar uma moeda, tem-se 50% de chances de sair cara e 50% de sair coroa, desde que a moeda seja justa (hipótese nula). Após algumas jogadas (amostra), se a quantidade de caras observadas (estatística de teste) for próxima da metade do total de jogadas, a moeda é considerada justa (hipótese nula não rejeitada), enquanto uma proporção de caras muito distante de 50% é considerada uma indicação de que a moeda tem algum viés (hipótese nula rejeitada). Para ilustrar o procedimento, a Figura 4.3 mostra uma nuvem de possibilidades para a estatística de teste, assumindo que a hipótese nula seja verdadeira. Valores mais próximos do centro da nuvem são mais prováveis (regiões mais escuras), enquanto valores mais distantes são menos prováveis (regiões mais claras).

O ponto marrom representa o valor obtido a partir do conjunto de dados disponível e a área hachurada na nuvem à esquerda abrange valores tão ou mais extremos que ele. Considerando toda a nuvem de possibilidades, a proporção dela que é hachurada representa o valor- p . Sob a hipótese nula, valores bem distantes do centro, apesar de menos prováveis, ainda são possíveis. A rejeição da hipótese nula (indicando presença do efeito de interesse) quando a hipótese nula é verdadeira (ausência de efeito) caracteriza um erro de falso positivo, também chamado de erro tipo I. A taxa considerada aceitável de falsos positivos é o que determina o limite com que o valor- p será comparado para se definir se a hipótese nula é rejeitada ou não, criando as regiões mostradas à direita na Figura 4.3. Na situação ilustrada, o valor- p obtido é menor que o limite e, portanto, a hipótese nula seria rejeitada.

A nuvem de possibilidades representa uma distribuição de probabilidade para possíveis amostras, ou seja, a amostra real disponível é julgada como mais ou menos provável em relação ao conjunto de amostras hipotéticas que poderiam ter sido obtidas. Esse conjunto de resultados possíveis muda com a maneira com que os dados são coletados e se múltiplos testes serão realizados. Por exemplo, se observações são coletadas até que um tamanho amostral predeterminado seja alcançado, então o conjunto de possibilidades de amostras é definido com todas elas tendo o mesmo tamanho; se, por outro lado, as observações são coletadas por um período de tempo específico, o tamanho amostral é um número aleatório e o conjunto de amostras possíveis (e, conseqüentemente, o valor- p calculado) é diferente do primeiro caso (Kruschke, 2015). Se pretende-se realizar múltiplos testes, é preciso levar em consideração a probabilidade de se obter em qualquer um deles valores tão ou mais extremos que o considerado, ou seja, a nuvem de resultados possíveis e o valor- p mudam, causando a necessidade de se ajustar a região de rejeição da hipótese nula para manter a taxa de falsos positivos geral dentro do limite aceitável (Kruschke, 2015). O cálculo do valor- p depende, portanto, das intenções de coleta de dados e dos testes que se pretende realizar (Kruschke, 2015; Kruschke & Liddell, 2018; Wagenmakers et al., 2018). Desse modo, duas amostras iguais podem levar a diferentes conclusões acerca de um parâmetro se o método de coleta de dados e as análises pretendidas forem diferentes. Essa é uma das desvantagens do valor- p .

Outra limitação do método é que ele só permite que a hipótese nula seja rejeitada ou não; hipóteses alternativas não podem ser aceitas, já que o valor- p é obtido olhando-se apenas para a distribuição sob a hipótese nula. Mesmo que a hipótese nula seja verdadeira, o aumento indefinido do tamanho amostral leva inevitavelmente a um valor- p menor que o limite fixado e à decisão de rejeição da hipótese nula, independente se a diferença detectada tem relevância prática ou não (Montgomery & Runger, 2011; Kruschke, 2013). Com o aumento do tamanho amostral, a nuvem de possibilidades ilustrada na Figura 4.3 se torna cada vez mais compacta e a região de rejeição cada vez maior, até que mesmo estatísticas de teste bem próximas do valor sob a hipótese nula levem a valores- p considerados baixos e a hipótese nula seja rejeitada. O método responde à pergunta sobre a probabilidade de

se obter um dado amostral pelo menos tão extremo quanto o obtido caso determinada hipótese seja verdadeira e essa é a única informação que o valor- p fornece. Discussões sobre as limitações do método e as interpretações muitas vezes inadequadas feitas sobre o valor- p existem há bastante tempo e têm crescido ainda mais recentemente (Baker, 2016; Matthews et al., 2017; Wagenmakers et al., 2018; Kruschke & Liddell, 2018; Kelter, 2020).

Dentro da abordagem frequentista, uma alternativa para o método NHST com valor- p é a estimação de parâmetros utilizando a estimativa de máxima verossimilhança e sua incerteza descrita pelo intervalo de confiança. A estimativa de máxima verossimilhança fornece o valor do efeito ou do parâmetro para o qual os dados obtidos têm a maior probabilidade de ocorrer e a incerteza dessa estimativa é dada pelo intervalo de confiança (quanto menor, mais precisa). Um intervalo de confiança de $c\%$ ($c \in [0, 100]$) é um intervalo obtido por um procedimento que, se repetido para diferentes amostras hipotéticas, conterá o valor real do parâmetro em $c\%$ das repetições (Montgomery & Runger, 2011; Wagenmakers et al., 2018). Enquanto o NHST se pergunta se um determinado valor do parâmetro pode ou não ser rejeitado, o intervalo de confiança fornece todos os valores de parâmetros que não seriam rejeitados para aquela amostra, dado um determinado limite de falsos positivos (Wagenmakers et al., 2018; Kruschke & Liddell, 2018). Seguindo o exemplo dado em Kruschke (2015), seja N o número de vezes que uma moeda é jogada e z o número de caras observadas. A hipótese nula determina que a moeda é justa, ou seja, que a proporção de vezes que a saída é cara é 0,5. Uma amostra de $z = 7$ e $N = 24$ (coletada até que o tamanho amostral $N = 24$ fosse atingido) é obtida e usada para testar a hipótese. Se, na verdade, a moeda é enviesada e apenas 12,6% das vezes cai cara, a probabilidade de que uma amostra de 24 jogadas dessa moeda tenha pelo menos 7 caras, ou seja, que a proporção de caras seja maior ou igual à obtida com a amostra disponível ($z/N = 0,292$), é de aproximadamente 2,5%. Se o valor real da proporção é 0,511, a probabilidade de uma amostra com proporção de caras menor ou igual a 0,292 é também de 2,5%. O intervalo $[0,126, 0,511]$ é o intervalo de confiança de 95% para uma amostra de $z = 7$ e $N = 24$, coletada até se alcançar o tamanho amostral N , e contém todos os valores de parâmetros que não seriam rejeitados com um taxa aceitável de falsos positivos de 5% (Kruschke, 2015). O intervalo de confiança, assim como o valor- p , depende das intenções de coleta de dados e, além disso, não fornece informações sobre as probabilidades de cada valor dentro dele, não sendo possível dizer quais deles são mais ou menos prováveis (Montgomery & Runger, 2011; Kruschke & Liddell, 2018).

Já a abordagem Bayesiana se contrapõe à abordagem frequentista e se fundamenta no teorema de Bayes para realocar credibilidade sobre possibilidades a partir de informações coletadas. A ideia é que se tem uma crença inicial sobre o valor de um parâmetro ou um conjunto deles e, com a observação dos dados coletados, essas crenças são atualizadas levando-se em consideração as informações prévias e as obtidas com a amostra atual. As crenças são representadas por distribuições de probabilidade. Seja $P(\lambda)$ a probabilidade

a priori da combinação de valores de parâmetros λ sem considerar os dados (crença inicial), $P(\mathcal{D}|\lambda)$ a verossimilhança ou a probabilidade de se obter o conjunto de dados \mathcal{D} considerando determinados valores de parâmetros λ e $P(\mathcal{D})$ a probabilidade dos dados de acordo com o modelo considerado, o teorema de Bayes determina que a probabilidade *a posteriori*, ou seja, a probabilidade de um ou mais valores de parâmetros considerando o conjunto de dados disponível é dada por

$$P(\lambda|\mathcal{D}) = P(\mathcal{D}|\lambda) \frac{P(\lambda)}{P(\mathcal{D})}. \quad (4.1)$$

Para comparar duas hipóteses ou dois modelos, como muitas vezes é a intenção ao se aplicar o método NHST, o teste de hipóteses com o fator de Bayes é uma alternativa que pode ser mais informativa. Nele, define-se duas distribuições *a priori*, uma delas relacionada com a hipótese nula (probabilidade consideravelmente alta no valor nulo e baixa nos outros valores) e outra relacionada com a hipótese alternativa (com probabilidades mais bem distribuídas entre diferentes valores do parâmetro), e usa-se o teorema de Bayes e a probabilidade relativa entre as duas hipóteses (Wagenmakers et al., 2018; Kruschke & Liddell, 2018; Kelter, 2020). O grau de mudança da credibilidade relativa *a priori* para a credibilidade relativa *a posteriori* é expresso pelo fator de Bayes, que, além de indicar qual das hipóteses é favorecida pelos dados, quantifica essa mudança. O teste de hipóteses Bayesiano é inerentemente comparativo, levando em consideração as duas hipóteses (ou modelos) e os dados, e permite que se quantifique evidência a favor tanto da hipótese nula quanto da hipótese alternativa. Entretanto, o fator de Bayes é bastante sensível à escolha de distribuições *a priori* (Kruschke, 2015), o que requer cautela na sua utilização.

Outra opção de análise comparativa é utilizar a razão entre verossimilhanças de duas hipóteses (ou modelos) concorrentes, dada por

$$\mathcal{L} = \frac{P(\mathcal{D}|\lambda_1)}{P(\mathcal{D}|\lambda_2)},$$

sendo λ_1 e λ_2 os parâmetros ou conjuntos de parâmetros que maximizam a verossimilhança sob as hipóteses 1 e 2, respectivamente, como evidência quantitativa a favor ou contra hipóteses (Glover, 2018; Perneger, 2021). Nessa abordagem, as duas hipóteses também são levadas em consideração e não apenas a hipótese nula, como nos testes frequentistas. Entretanto, diferente do fator de Bayes, que ajusta a razão de verossimilhança a partir das crenças iniciais sobre os parâmetros (Glover, 2018), ao usar apenas essa razão na análise, as hipóteses são avaliadas com base somente na evidência fornecida pelo conjunto de dados.

Quando o objetivo não é testar hipóteses e sim estimar a magnitude de um efeito ou o valor de um parâmetro, a abordagem Bayesiana fornece informações mais completas que a estimação frequentista com intervalo de confiança. As crenças iniciais são representadas pelas distribuições *a priori* para os parâmetros e a escolha dessas distribuições é um tópico

importante na análise Bayesiana. Se, por exemplo, estudos anteriores sugerem com alguma confiança um valor específico para um parâmetro, a distribuição *a priori* pode refletir essa crença, incorporando informações prévias à análise. Quando não se tem essas informações anteriores, normalmente são usadas distribuições *a priori* vagas e pouco informativas, para que não tenham forte influência nas distribuições *a posteriori* (Kruschke & Liddell, 2018; Kelter, 2020). A Figura 4.4 ilustra uma distribuição *a priori* uniforme, com probabilidades iguais para todos os valores de um parâmetro dentro de um intervalo. Após o procedimento, uma distribuição *a posteriori* é obtida e as probabilidades de cada valor do parâmetro são atualizadas. O que se obtém é uma distribuição de probabilidade, ou seja, a credibilidade de cada valor do parâmetro de interesse é indicada pela própria distribuição. Para resumir a distribuição *a posteriori* pode-se usar medidas como média, moda e mediana para indicar os valores mais prováveis e definir intervalos como o 95% HDI (do inglês, *Highest Density Interval*), que engloba 95% da distribuição, permitindo que se afirme que valores para o parâmetro dentro do HDI são mais prováveis que valores fora dele (Kruschke, 2015). O HDI contém os valores de parâmetros mais prováveis e sua largura indica a incerteza da estimativa: quanto menor, mais precisa a estimativa e mais certeza se tem sobre o valor do parâmetro.

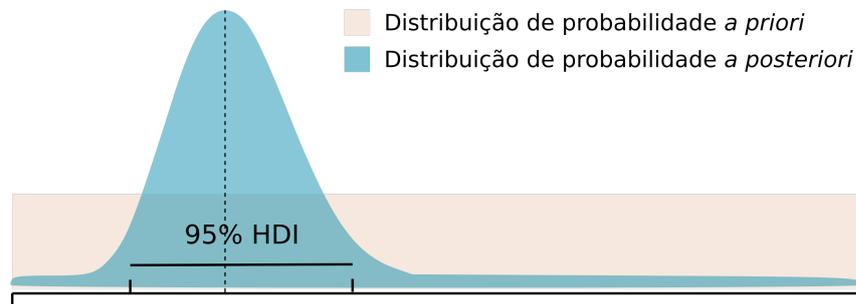


Figure 4.4: Distribuições de probabilidade *a priori* e *a posteriori* para valores de parâmetros na estimação Bayesiana. O 95% HDI engloba 95% da distribuição posterior e contém os valores mais prováveis para o parâmetro.

Para obter a distribuição *a posteriori*, é necessário calcular o denominador na Equação 4.1. O termo $P(\mathcal{D})$ é uma constante dada por (Kruschke, 2015)

$$P(\mathcal{D}) = \int P(\mathcal{D}|\Lambda)P(\Lambda)d\Lambda,$$

em que Λ representa todas as possibilidades de parâmetros e não apenas uma combinação específica deles. Essa integral pode ser impossível de se resolver analiticamente, especialmente quando se tem múltiplos parâmetros sendo estimados, criando um espaço de parâmetros que envolve distribuições conjuntas de todas as combinações de parâmetros possíveis. Quando esse é o caso, pode-se gerar uma amostra grande e representativa da distribuição *a posteriori* para aproximá-la, sem que seja necessário calcular $P(\mathcal{D})$. Algoritmos frequentemente utilizados para se obter essa amostra da distribuição *a posteriori* fazem

parte da classe de métodos de Monte Carlo via cadeias de Markov (do inglês, *Markov Chain Monte Carlo* - MCMC)¹.

Com a distribuição *a posteriori* também é possível avaliar a credibilidade de valores específicos para o parâmetro, seja um valor nulo, indicando ausência de um efeito, ou outro valor de interesse. Um procedimento usado para fazer esse tipo de análise é com a definição de uma região de equivalência prática (do inglês, *Region Of Practical Equivalence* - ROPE) ao redor do valor de interesse, isto é, determinar um conjunto de valores que, para fins práticos, são considerados equivalentes ao valor especificado (Kruschke, 2013). Caso seja necessário se decidir sobre aceitar ou rejeitar um valor específico para o parâmetro, Kruschke (2015) propõe uma regra de decisão utilizando a ROPE e o 95% HDI da distribuição *a posteriori*: se todo o 95% HDI está dentro da ROPE ao redor do parâmetro, ele é considerado aceito para fins práticos; se toda a ROPE está fora do 95% HDI, o valor do parâmetro em questão é considerado não provável e é rejeitado; se nenhuma das situações anteriores acontece, não se tem um resultado claro, indicando que o conjunto de dados disponível não é suficiente para se tomar uma decisão. A Figura 4.5 ilustra algumas possíveis situações e as decisões que seriam tomadas segundo o critério de Kruschke (2015). A ROPE deve ser definida de acordo com a aplicação e o contexto do parâmetro, para que realmente reflita equivalência prática. Com uma ROPE adequada, um valor de interesse só é aceito quando se tem uma estimativa suficientemente precisa para o parâmetro, o que se reflete na distribuição *a posteriori* através de um 95% HDI estreito.

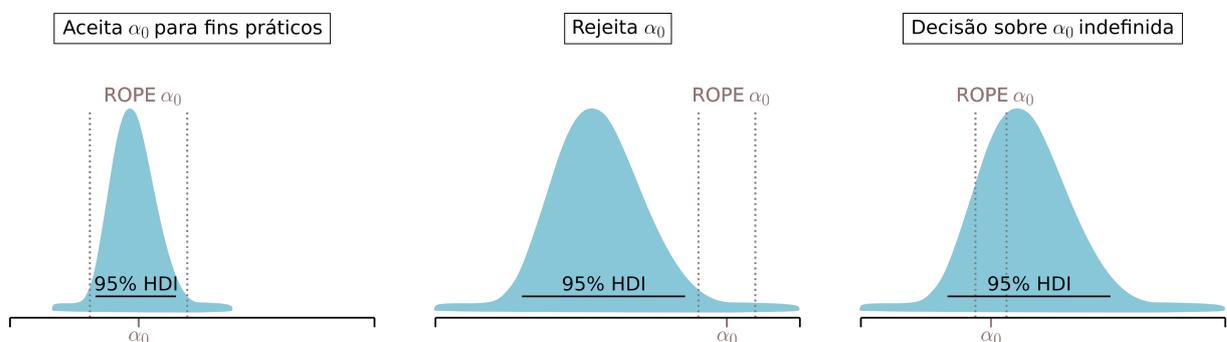


Figure 4.5: Exemplos de possíveis relações entre o 95% HDI da distribuição *a posteriori* e a ROPE ao redor do valor de interesse α_0 para o parâmetro. Se possível, uma decisão sobre α_0 é tomada para cada caso, segundo o critério proposto por Kruschke (2015).

Considerando todos os pontos apresentados sobre diferentes métodos disponíveis para análise, a estimação de parâmetros Bayesiana é usada neste trabalho. A abordagem Bayesiana fornece informações mais ricas e a estimação de parâmetros com incerteza permite discussões além da dicotomia aceitar/rejeitar hipóteses ou de uma escolha entre dois modelos ou hipóteses. Entretanto, por serem populares, diversas implementações dos métodos NHST estão disponíveis e os testes são facilmente aplicados. Para que se possa realizar uma comparação qualitativa, mesmo que superficial, entre os resultados

¹Mais informações sobre os métodos MCMC podem ser encontradas em Kruschke (2015), capítulo 7.

obtidos com cada método, testes de hipóteses frequentistas comumente encontrados na literatura também são aplicados aos dados obtidos neste trabalho. A seguir, são detalhadas as análises frequentista e Bayesiana para as variáveis de interesse listadas na seção anterior.

4.4.1 Testes de significância de hipótese nula

Para as análises sob a abordagem frequentista, testes de hipóteses comumente utilizados na literatura são aplicados, de acordo com o tipo de cada variável.

4.4.1.1 Medidas objetivas

Tempo e número de erros são tratadas como variáveis em escala métrica, isto é, variáveis quantitativas, de intervalo ou de razão. Um teste frequentemente aplicado para esse tipo de variável é o teste t para uma ou duas amostras. O teste é sobre a média de uma população ou a diferença entre as médias de duas populações. No experimento proposto, cada participante passa pelas duas configurações de comunicação (EX e EXIM), ou seja, as medidas são feitas em pares. Nesses casos, uma maneira de cancelar as variações individuais é fazer a análise usando a diferença entre as condições para cada participante, resultando em um único conjunto de dados. Para cada participante, a diferença entre as observações em cada configuração de comunicação, $x_{EX} - x_{EXIM}$, nesse caso de tempo ou de número de erros, é tomada.

O teste t sobre a média μ de uma população com variância desconhecida determina que, se a hipótese nula, $\mu = \mu_0$, sendo μ_0 o valor nulo de interesse, é verdadeira, a estatística de teste

$$T = \frac{(\bar{x} - \mu_0)}{(s/\sqrt{N})} \quad (4.2)$$

tem uma distribuição t com $N - 1$ graus de liberdade, sendo \bar{x} a média e s o desvio padrão de uma amostra aleatória de tamanho N retirada de uma distribuição normal com média μ (Montgomery & Runger, 2011). Além da hipótese nula, que indica ausência de efeito, uma hipótese alternativa é definida para refletir a hipótese de trabalho e o efeito de interesse. Por exemplo, se acredita-se que a média pode ser maior que o valor nulo, a hipótese alternativa é definida como $\mu > \mu_0$ e, dessa forma, apenas estatísticas de teste maiores que determinado limite poderão levar à rejeição da hipótese nula; se a hipótese alternativa é $\mu < \mu_0$, apenas estatísticas de teste menores que dado limite são consideradas extremas; se efeitos positivos ou negativos são de interesse, estatísticas de teste distantes de zero em ambas as direções poderão levar à rejeição da hipótese nula.

Supondo que deseja-se comparar o desempenho de dois robôs na execução de tarefas. Para isso, o tempo de cada robô para realizar três tarefas distintas é registrado. O robô A demora 1,5 minuto na tarefa 1, 2,3 minutos na tarefa 2 e 1,2 minuto na tarefa 3, enquanto o robô B gasta 1,2, 2,4 e 0,8 minutos nas tarefas 1, 2 e 3, respectivamente. Como as tarefas

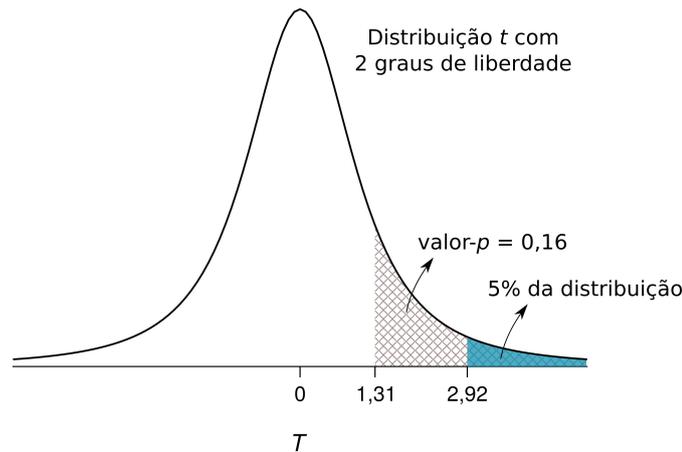


Figure 4.6: Distribuição da estatística de teste (4.2) sob a hipótese nula para $N = 3$. A hipótese alternativa do teste é $\mu > 0$ e a taxa aceitável de falsos positivos é 5%, definindo a região de rejeição mostrada em *azul*. Para uma amostra $\{0,3, -0,1, 0,4\}$, a estatística de teste é $T = 1,31$ e o valor- p é representado pela área hachurada, correspondendo a 16% da distribuição. Nesse caso, a hipótese nula, $\mu = 0$, não é rejeitada.

são as mesmas para ambos os robôs, a diferença de tempo $x_A - x_B$ na execução de cada uma é tomada e o conjunto de dados resultante é $\{0,3, -0,1, 0,4\}$. Por causa de uma nova tecnologia que foi utilizada, acredita-se que o robô B seja mais rápido, portanto, a hipótese nula sobre a média das diferenças é $\mu = 0$, enquanto a hipótese alternativa reflete o que se supõe sobre os robôs e determina que $\mu > 0$. A taxa aceitável de falsos positivos para o teste é de 5%. A Figura 4.6 mostra a distribuição da estatística de teste (4.2) sob a hipótese nula, com $N - 1 = 2$ graus de liberdade. Como apenas efeitos positivos são de interesse, a região de rejeição da hipótese nula abrange 5% da distribuição na sua cauda positiva (área sob a curva) e corresponde a valores de T acima de 2,92, como mostrado em azul na Figura 4.6. Para o conjunto de dados disponível, $\bar{x} = 0,2$, $s = 0,26$ e a estatística de teste é $T = 1,31$ e o valor- p obtido é 0,16 (área hachurada na Figura 4.6), ou seja, com uma amostra de $N = 3$, a probabilidade de se obter valores de $T \geq 1,31$ quando a hipótese nula é verdadeira é de 16%. Como $0,16 > 0,05$, a hipótese nula não é rejeitada.

Para a diferença entre as configurações de comunicação EX e EXIM, a hipótese nula para o teste t propõe que $\mu = 0$ (valor nulo $\mu_0 = 0$), enquanto a hipótese alternativa é dada por $\mu > 0$, expressando as hipóteses de trabalho descritas na Seção 4.1. O valor- p é dado pela probabilidade de que T assuma valores maiores que o obtido com a amostra disponível. Neste trabalho, a taxa de falsos positivos aceitável, com a qual o valor- p obtido com o teste é comparado, é de 5%, valor comumente encontrado na literatura. O teste é aplicado utilizando a função `t.test`² do pacote `stats` (versão 4.0.4) para linguagem R.

O procedimento do teste t , entretanto, se baseia na premissa de que a distribuição das médias amostrais é pelo menos aproximadamente normal. A normalidade é então verificada a partir de análise gráfica e de um teste de hipóteses com essa finalidade. Se

²<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/t.test.html>

as análises indicarem desvio significativo de normalidade, o teste de Wilcoxon de posto sinalizado (Wilcoxon, 1945) é aplicado. Para testar a hipótese nula $\tilde{\mu} = 0$ sobre a mediana $\tilde{\mu}$ da população, a amostra é ordenada de forma crescente considerando o valor absoluto das observações, sendo R_i o posto, *i.e.*, o número atribuído à posição, de cada observação i após o ordenamento. Para cada observação, a variável ψ_i assume valor igual a 1 se a observação é maior que o valor nulo ou $\psi_i = 0$ se a observação é menor que o valor nulo. A estatística de teste é então dada por

$$T = \sum_{i=1}^N R_i \psi_i, \quad (4.3)$$

que é a soma dos postos positivamente sinalizados (Hollander et al., 2014). Voltando ao exemplo do desempenho dos robôs A e B, a Tabela 4.2 mostra todas as oito possibilidades para a estatística de teste (4.3) quando $N = 3$. A coluna do meio na tabela mostra o conjunto dos postos positivamente sinalizados em cada caso, $\mathcal{P} \triangleq \{R_i \psi_i : R_i \psi_i > 0, i \in \{1, 2, 3\}\}$. No conjunto de dados, as diferenças de tempo para cada tarefa são $x_{\text{dif}_1} = 0,3$, $x_{\text{dif}_2} = -0,1$ e $x_{\text{dif}_3} = 0,4$ e, com isso, os postos após o ordenamento são $R_1 = 2$, $R_2 = 1$ e $R_3 = 3$. A estatística de teste, calculada com (4.3), é $T = 2 \cdot 1 + 1 \cdot 0 + 3 \cdot 1 = 5$. A hipótese alternativa é de que $\tilde{\mu} > 0$ e as linhas destacadas na Tabela 4.2 indicam as duas das oito possibilidades para as quais $T \geq 5$, ou seja, em quais opções a estatística de teste é pelo menos tão alta quanto a obtida com o conjunto de dados disponível. O valor- p obtido é $2/8 = 0,25$, que é maior que o limite de falsos positivos definido (5%), e, novamente, a hipótese nula não é rejeitada.

Table 4.2: Possibilidades para a estatística de teste (4.3) sob a hipótese nula para $N = 3$. Para cada uma, o conjunto \mathcal{P} dos postos positivamente sinalizados é mostrado. Para uma amostra $x_{\text{dif}_1} = 0,3$, $x_{\text{dif}_2} = -0,1$ e $x_{\text{dif}_3} = 0,4$, a estatística de teste é $T = 5$ e as linhas destacadas mostram as duas opções para as quais $T \geq 5$. O valor- p obtido para o teste com hipótese alternativa $\tilde{\mu} > 0$ é, portanto, $2/8 = 0,25$, que é maior que a taxa aceitável de falsos positivos (5%), e a hipótese nula não é rejeitada.

	\mathcal{P}	T
1	{}	0
2	{1}	1
3	{2}	2
4	{3}	3
5	{1, 2}	3
6	{1, 3}	4
7	{2, 3}	5
8	{1, 2, 3}	6

Quando há empate ao se ordenar os valores absolutos das observações, uma possível solução é atribuir a cada R_i das observações empatadas o valor médio dos postos associados ao grupo de empate. Quando observações iguais ao valor nulo estão presentes, uma opção é

descartá-las e atualizar o tamanho amostral. Existem outras opções para se tratar empates e observações nulas, mas, a partir da documentação e de testes feitos com a função em R utilizada neste trabalho, essas são as soluções implementadas. Para a diferença entre as configurações de comunicação EX e EXIM, o teste é feito com a função `wilcox.exact`³ do pacote `exactRankTests` (versão 0.8.32) para linguagem R com as hipóteses nula e alternativa sobre a mediana dadas por $\tilde{\mu} = 0$ e $\tilde{\mu} > 0$, respectivamente, e o mesmo limite de falsos positivos definido para o teste t . Uma premissa para a utilização do teste de Wilcoxon de posto sinalizado é de que a amostra vem de uma população simétrica. De acordo com Hollander et al. (2014) e Kloke & Mckean (2015), essa premissa já é satisfeita para observações pareadas.

4.4.1.2 Medidas subjetivas

As variáveis subjetivas são medidas através de escalas Likert, isto é, um conjunto de itens com escala de resposta ordinal, como descrito na Seção 4.3. Seguindo a proposta de Likert (1932), uma maneira frequentemente utilizada para lidar com esses dados é calcular a média ou a soma das pontuações de cada item da escala e tratar esse valor como a observação relacionada a cada participante. Há uma discussão na literatura sobre como esse conjunto de dados, gerado com a soma ou a média dos pontos da escala, deve ser tratado: se as medidas são ordinais ou intervalares e quais testes podem ser aplicados. Em seu trabalho, Nanna & Sawilowsky (1998) compararam o teste t e o teste de Wilcoxon para duas amostras independentes de dados em escala Likert e observaram que a potência (a sensibilidade do teste para se detectar diferenças) do teste de Wilcoxon foi quase sempre maior. Já os resultados de Meek et al. (2007) para os testes com uma única amostra indicaram que o teste t parece ser preferível exceto quando se deseja controlar fortemente a probabilidade de falsos positivos. Carifio & Perla (2008) argumentam que evidências empíricas suportam a ideia de que escalas Likert produzem dados intervalares e que testes paramétricos (baseados em distribuições específicas, como a normal) como o teste t podem ser usados, sendo esta também a recomendação de Harpe (2015), desde que as premissas dos testes sejam respeitadas. Já Liddell & Kruschke (2018) mostram que tratar dados ordinais de um único item como se fossem métricos leva a erros sistemáticos de falsos positivos, falhas na detecção de efeitos e até inversão de efeitos e que usar a média, apesar de fazer com que os dados se assemelhem mais a dados contínuos, não resolve os problemas.

Como não há consenso sobre o tipo de dados e se testes paramétricos como o teste t podem ou não ser aplicados, neste trabalho opta-se por uma análise mais conservadora, utilizando o teste de Wilcoxon de posto sinalizado. A média de pontos nos itens de cada escala é calculada e a diferença $x_{EXIM} - x_{EX}$ é tomada para cada participante, de modo que observações positivas são favoráveis às hipóteses de trabalho (ver Seção 4.1). As hipóteses

³<https://www.rdocumentation.org/packages/exactRankTests/versions/0.8-32/topics/wilcox.exact>

nula e alternativa para os testes são como as definidas anteriormente, $\tilde{\mu} = 0$ e $\tilde{\mu} > 0$, respectivamente, e a taxa de falsos positivos aceitável é novamente de 5%.

4.4.2 Estimação de parâmetros Bayesiana

De acordo com Kruschke (2015), a análise de dados Bayesiana segue os seguintes passos:

1. Identificação do tipo de dados, definindo qual a escala de medição (ordinal, métrica, etc.).
2. Definição de um modelo para descrever os dados. O modelo matemático deve ser adequado para representar a amostra disponível e ter parâmetros que sejam significativos e possam ser interpretados.
3. Especificação de distribuições *a priori* para os parâmetros que se deseja estimar. A escolha das distribuições *a priori* deve ser justificada.
4. Interpretação das distribuições *a posteriori* dos parâmetros, obtidas a partir da inferência Bayesiana.
5. Checagem posterior, comparando o conjunto de dados com previsões feitas a partir de valores posteriores dos parâmetros. Se as previsões não imitarem razoavelmente os dados, pode ser necessário mudar o modelo utilizado.

A principal referência utilizada em todas as etapas da análise Bayesiana foi o livro *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, de John K. Kruschke (2015). As distribuições *a posteriori* dos parâmetros são aproximadas usando métodos MCMC. O sistema JAGS (*Just Another Gibbs Sampler*), criado por Plummer (2003, apud Kruschke, 2015), que recebe a descrição do modelo e retorna amostras MCMC representativas das distribuições *a posteriori* dos parâmetros, é utilizado. Todos os códigos foram escritos em R, baseados nos exemplos e usando as funções que acompanham os trabalhos de Kruschke (2015) e Liddell & Kruschke (2018). A seguir, são apresentadas as três primeiras etapas da estimação de parâmetros Bayesiana para as variáveis analisadas neste trabalho.

4.4.2.1 Medidas objetivas

Um modelo comum para dados métricos é a distribuição normal, especificada através de dois parâmetros: média e desvio padrão. Para descrever dados com valores atípicos (*outliers*, em inglês), uma alternativa frequentemente utilizada é a distribuição *t*, que possui caudas mais pesadas (com probabilidades mais altas) que podem acomodar os dados melhor que a distribuição normal (Kruschke, 2015). Além dos parâmetros de tendência central e de escala, representados por μ e σ , respectivamente, a distribuição *t* possui

um terceiro parâmetro, o parâmetro de normalidade⁴ ν , que assume valores de 1 a ∞ e determina o peso nas suas caudas, como ilustrado na Figura 4.7. As caudas são mais pesadas quando o parâmetro é igual a 1 e, quanto maior, mais próxima a distribuição t se torna da distribuição normal. Assim como para a distribuição normal, o parâmetro de tendência central μ da distribuição t é a média. Entretanto, de acordo com Kruschke (2015), o parâmetro de escala σ não é o desvio padrão da distribuição t . Por causa das caudas mais pesadas que as da normal, o desvio padrão (a raiz quadrada da variância, que é dada por

$$\int_{-\infty}^{\infty} P(x)(x - \mu)^2 dx$$

para valores de x contínuos gerados com probabilidade $P(x)$) da distribuição t é infinito para ν menor que 2. Apesar disso, o parâmetro continua tendo relação com o espalhamento dos dados: para $\nu = \infty$, o intervalo $\pm\sigma$ de uma distribuição t de média $\mu = 0$ abrange 68% da distribuição, assim como para uma normal; já para $\nu = 1$, o intervalo $\pm\sigma$ abrange 50% da distribuição (Kruschke, 2015). A Figura 4.7 mostra o intervalo entre $\mu - \sigma$ e $\mu + \sigma$ e a porcentagem que ele abrange das distribuições com $\nu = 1$ e $\nu = \infty$.

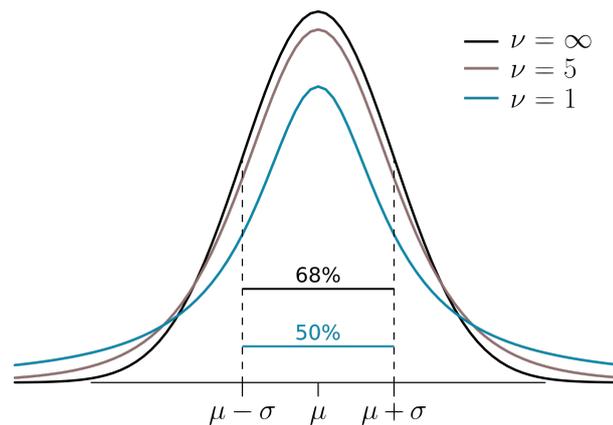


Figure 4.7: Exemplos de distribuições t de média μ , parâmetro de escala σ e diferentes parâmetros de normalidade ν . Quanto maior ν , mais próxima a distribuição é de uma distribuição normal. O parâmetro σ está relacionado com o espalhamento dos dados e abrange 50% da distribuição t com $\nu = 1$ e 68% da distribuição com $\nu = \infty$.

Para uma estimação mais robusta a possíveis valores atípicos, a distribuição t é usada no modelo para as variáveis métricas, isto é, considera-se que as observações (diferença de tempo ou número de erros) y vêm de uma distribuição t e o objetivo da inferência Bayesiana é estimar seus parâmetros μ , σ e ν . Estimações precisas de ν só são possíveis a partir de valores presentes nas caudas da distribuição, que são naturalmente menos prováveis. Portanto, não se espera que a distribuição *a posteriori* forneça informações

⁴O parâmetro ν também é conhecido como “graus de liberdade”, termo frequentemente encontrado no contexto de NHST quando a distribuição amostral é representada pela distribuição t , mas, assim como em Kruschke (2015), será chamado de “parâmetro de normalidade” no contexto da abordagem Bayesiana deste trabalho, descrevendo o efeito que causa na distribuição.

precisas sobre ν , o que não prejudica a análise, já que a distribuição t é usada apenas para dar mais flexibilidade ao modelo (Kruschke, 2015).

O próximo passo é a definição de distribuições *a priori* para os parâmetros a serem estimados. Seguindo as sugestões de Kruschke (2015), a distribuição *a priori* de μ é normal, a de σ é uniforme e a de ν é exponencial. Como não se tem informações prévias sobre os parâmetros, distribuições *a priori* largas e pouco informativas são desejadas, para que tenham mínima influência nas distribuições *a posteriori*. As distribuições *a priori* para μ e σ são dadas, respectivamente, por uma distribuição normal com média \bar{x} e desvio padrão alto, de $100s$, e por uma distribuição uniforme de $s/1000$ a $1000s$. Dessa maneira, independente dos dados, as distribuições *a priori* são largas e expressam incerteza sobre os valores de parâmetros. Já para ν , para que valores baixos para o parâmetro, mais adequados para dados com valores atípicos, sejam considerados prováveis na distribuição *a posteriori*, é preciso que os dados contenham uma quantidade razoável de valores atípicos (que são, por definição, raros) ou que a distribuição *a priori* dê considerável credibilidade a valores baixos. Além disso, valores de ν acima de aproximadamente 30 já fazem da distribuição t essencialmente uma distribuição normal, ou seja, a maior variação na distribuição acontece para valores baixos de ν (Kruschke, 2015). A distribuição exponencial como distribuição *a priori* é escolhida justamente para que se possa dar considerável credibilidade a valores baixos de ν , enquanto ainda se permite valores altos. A distribuição é deslocada para assumir valores de 1 a ∞ e sua média é 30. A Figura 4.8 mostra um diagrama da estimação Bayesiana das variáveis métricas, indicando a distribuição t para descrição dos dados e as distribuições *a priori* para seus parâmetros.

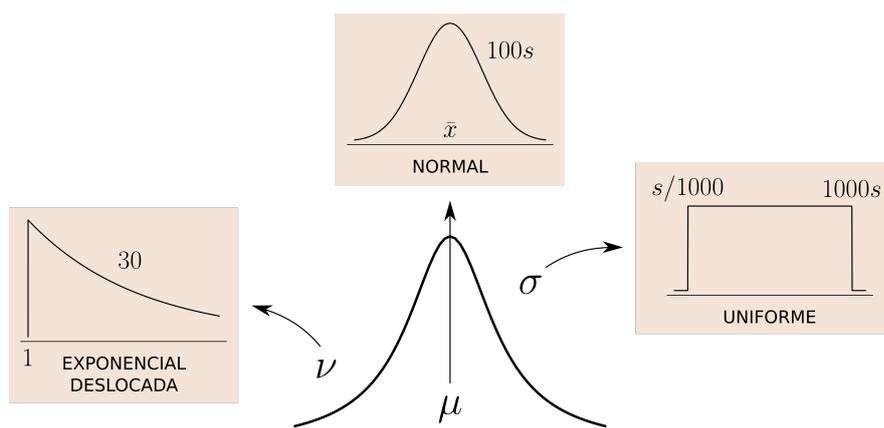


Figure 4.8: Diagrama da estimação Bayesiana das variáveis métricas. Uma distribuição t descreve os dados e seus parâmetros μ , σ e ν são estimados. As distribuições *a priori* para cada um dos parâmetros são mostradas, sendo \bar{x} e s a média e o desvio padrão amostrais.

Novamente as diferenças entre as configurações de comunicação, $x_{EX} - x_{EXIM}$, para tempo e número de erros são tomadas e o modelo é aplicado a um único grupo para cada variável, sendo que diferenças positivas são favoráveis às hipóteses deste trabalho.

4.4.2.2 Medidas subjetivas

Ao discutirem o tratamento de dados ordinais como métricos, Liddell & Kruschke (2018) mostram que usar a média dos pontos nos itens de uma escala Likert não resolve os problemas de falsos positivos, falhas na detecção e inversão de efeitos discutidos em seu trabalho e sugerem a utilização de um modelo com normal acumulada. Para múltiplos itens correlacionados, o modelo apresentado no trabalho de Liddell & Kruschke (2018) é uma extensão da análise Bayesiana de dados ordinais para um item discutida em Kruschke (2015), tratando os itens em conjunto, mas sem agregar as pontuações em uma única medida.

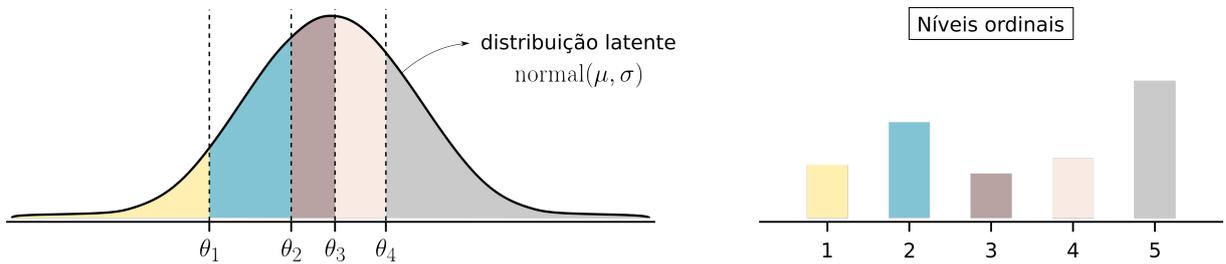


Figure 4.9: A probabilidade da resposta ordinal $y = k$ é dada pela probabilidade acumulada entre os intervalos θ_{k-1} e θ_k na distribuição normal latente de média μ e desvio padrão σ .

A ideia do modelo utilizado em Kruschke (2015) e Liddell & Kruschke (2018) é que a variável que se quer medir está em uma escala métrica contínua, mas que não pode ser acessada diretamente, ou seja, é uma variável latente. A escala Likert é uma maneira de acessar essa variável latente através de uma escala de respostas discreta e ordinal, fazendo com que as pessoas criem alguns limites nessa distribuição latente para traduzi-la nas categorias de resposta. Para $K \in \mathbb{N}$ níveis de resposta, $K - 1$ limites dividem a distribuição latente em K intervalos. O modelo ordinal considera que a probabilidade de cada resposta ordinal é a probabilidade acumulada de cada intervalo, ou seja, a área sob a curva da distribuição latente entre os limites de cada região, como ilustrado na Figura 4.9 (para $K = 5$). Assumindo que a distribuição latente é uma normal de média μ e desvio padrão σ e os limites são dados por θ_k , $k \in \{1, 2, \dots, K - 1\}$, a probabilidade da resposta ordinal $y = k$ é

$$P(y = k | \mu, \sigma, \theta_1, \dots, \theta_{K-1}) = \Phi\left(\frac{\theta_k - \mu}{\sigma}\right) - \Phi\left(\frac{\theta_{k-1} - \mu}{\sigma}\right), \quad (4.4)$$

onde $\Phi(\cdot)$ é a função normal padrão acumulada, ou seja,

$$\Phi(a) = P(Z \leq a) = \int_{-\infty}^a f(u) du,$$

sendo Z uma variável aleatória normal padrão e $f(\cdot)$ a função de densidade de probabilidade normal padrão. Esse modelo é conhecido também como *probit* ordenado ou *probit* ordinal. A Equação 4.4 se aplica inclusive aos níveis ordinais das extremidades ($y = 1$ e $y = K$)

se considerarmos dois limites virtuais em $\theta_0 = -\infty$ e $\theta_K = \infty$, sendo que $\Phi(-\infty) = 0$ e $\Phi(\infty) = 1$.

Ao todo, o modelo tem $K + 1$ parâmetros: os parâmetros μ e σ da distribuição latente e os $K - 1$ limites que a mapeiam nas respostas ordinais. Existem infinitas possibilidades de combinação para esses parâmetros que resultam nas mesmas probabilidades ordinais. Todo o conjunto de parâmetros mostrado na Figura 4.9 pode ser “arrastado”, “comprimido” ou “expandido” em relação ao eixo horizontal mantendo-se as probabilidades de cada nível ordinal. É necessário, portanto, fixar essa distribuição. Kruschke (2015) sugere que os limites mais extremos θ_1 e θ_{K-1} sejam fixados em valores significativos em relação à escala de respostas, especificamente em $\theta_1 = 1,5$ e $\theta_{K-1} = K - 0,5$, de modo que todos os outros parâmetros estimados estarão ancorados nesses valores significativos. Com isso, a interpretação das estimativas dos parâmetros pode ser feita a partir dos significados das opções de resposta. Supondo que se peça que as pessoas respondam um item de uma escala Likert que afirme “Eu gosto de robôs.” em uma escala de cinco níveis de resposta que vão de “discordo totalmente” a “concordo totalmente”, passando pela opção neutra “não sei”, a Figura 4.10 ilustra histogramas de três possíveis amostras e as distribuições latentes que se adequariam a cada uma delas, todas com os limites mais extremos fixados em 1,5 e 4,5. Para a amostra 1, as respostas se concentram no meio da escala ordinal então a média da distribuição latente seria algo próximo a 3, sugerindo que, em média, as pessoas não estão muito certas se gostam ou não de robôs. Já para a amostra 2, respostas negativas são mais frequentes e a média latente teria valor mais baixo, indicando que, em média, as pessoas não gostam de robôs. A amostra 3 ilustra o caso contrário, em que o resultado sugeriria que as pessoas gostam de robôs, com respostas em níveis mais altos sendo mais frequentes.

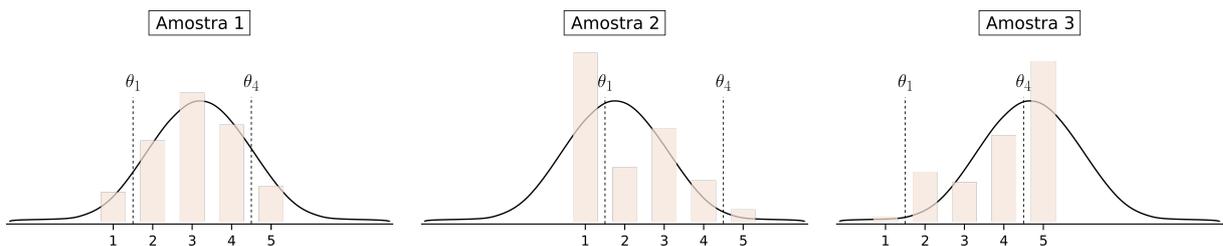


Figure 4.10: Exemplos de histogramas para respostas ordinais (cinco níveis) de três amostras e distribuições latentes que se adequariam a cada uma. Os limites extremos do modelo são fixados em $\theta_1 = 1,5$ e $\theta_4 = 4,5$.

Quando se tem dois ou mais grupos para os quais se deseja medir a mesma variável latente, a partir de um mesmo questionário, assume-se que a variável latente para os dois grupos tem a mesma forma, porém com média e desvio padrão diferentes. Já os limites, por estarem mais relacionados com a maneira de se medir a variável, ou seja, com a afirmação presente no item da escala Likert, são os mesmos para todos os grupos. Com isso, o que varia entre os grupos é o quanto se concorda ou discorda com a afirmação e a variância

desse sentimento. Já quando se utiliza múltiplos itens, como proposto por Likert (1932), o que se tem são diferentes maneiras de acessar uma mesma variável latente. Dentro de um único grupo, portanto, assume-se que os valores de μ e σ do grupo são comuns a todos os itens, mas os limites para cada um são diferentes (Liddell & Kruschke, 2018). Nesse caso, Liddell & Kruschke (2018) sugerem que os limites extremos apenas do primeiro item sejam fixados em $\theta_1 = 1,5$ e $\theta_{K-1} = K - 0,5$ e que todos os restantes (os intermediários do item 1 e todos os limites dos demais itens) sejam estimados. Adaptando a Equação 4.4 para o caso geral de múltiplos itens i e múltiplos grupos g , a probabilidade de cada resposta ordinal é dada por (Liddell & Kruschke, 2018)

$$P(y_g^{[i]} = k | \mu_g, \sigma_g, \theta_1^{[i]}, \dots, \theta_{K-1}^{[i]}) = \Phi\left(\frac{\theta_k^{[i]} - \mu_g}{\sigma_g}\right) - \Phi\left(\frac{\theta_{k-1}^{[i]} - \mu_g}{\sigma_g}\right). \quad (4.5)$$

Assim como no modelo para variáveis métricas, a distribuição normal latente pode ser substituída pela distribuição t , tornando a estimação mais robusta à presença de valores atípicos. Para isso, substitui-se $\Phi(\cdot)$ na Equação 4.5 pela função t acumulada, dada por

$$\Psi(a) = P(V \leq a) = \int_{-\infty}^a g(u) du,$$

sendo V uma variável aleatória de distribuição t e $g(\cdot)$ sua função densidade de probabilidade. O modelo determina que a resposta ordinal y vem de uma distribuição categórica com probabilidades $P(y = k | \mu_g, \sigma_g, \nu_g, \theta_1^{[i]}, \dots, \theta_{K-1}^{[i]})$ e o objetivo da inferência Bayesiana é estimar os parâmetros μ_g , σ_g e ν_g da distribuição latente de cada grupo e os limites que não tiverem sido fixados.

A partir das sugestões de Liddell & Kruschke (2018), a distribuição *a priori* para μ é novamente uma distribuição normal, dessa vez centrada no meio da escala ordinal e com desvio padrão igual ao número de categorias ordinais K , e a distribuição *a priori* de σ é uniforme com limite inferior de 0,01 e superior de $10K$, mantendo μ e σ na vizinhança dos dados, que só podem assumir valores entre 1 e K . Para ν , a distribuição *a priori* continua sendo uma exponencial deslocada com média 30 e para os limites θ_k não fixados, as distribuições *a priori* são normais com médias iguais a $k + 0,5$ e desvios padrão iguais a 2, permitindo variação considerável. A Figura 4.11 mostra um diagrama da análise Bayesiana das variáveis ordinais.

Limitações do modelo e ajustes

No modelo ordinal, é importante notar que não há nada que especifique que os limites $\theta_1, \dots, \theta_{K-1}$ sejam ordenados de maneira crescente, ou seja, que $\theta_1 < \theta_2 < \dots < \theta_{K-1}$. Entretanto, se os limites forem invertidos, de modo que $\theta_{k-1} > \theta_k$, a probabilidade calculada para o nível ordinal k é negativa, o que viola o primeiro axioma na teoria das probabilidades.

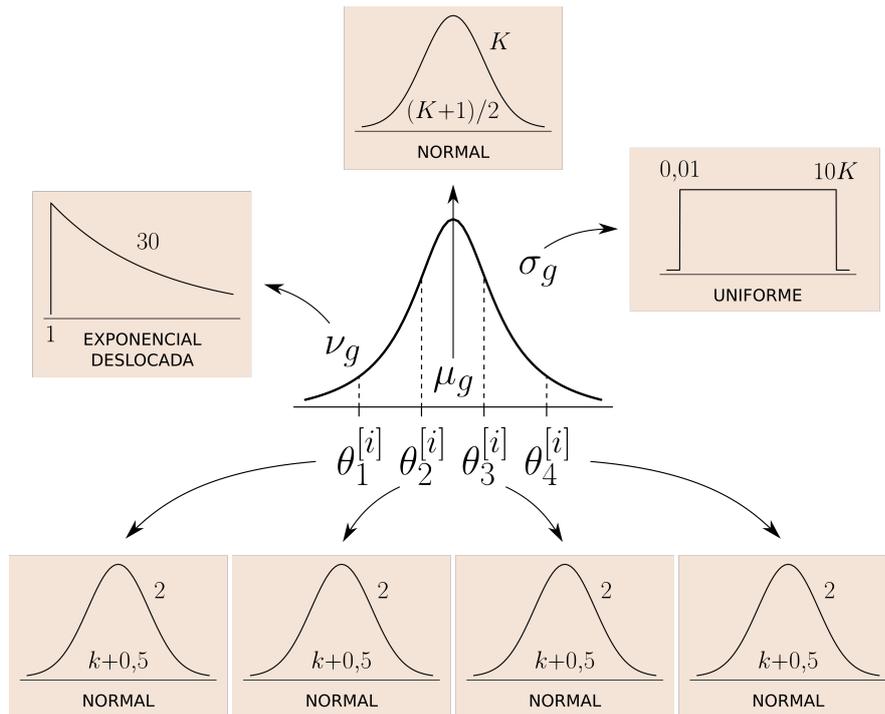


Figure 4.11: Diagrama da estimação Bayesiana das variáveis ordinais. Uma distribuição t descreve a variável latente e os seus parâmetros μ , σ e ν são estimados, assim como os limites $\theta_k^{[i]}$, $k \in \{1, 2, 3, 4\}$, que traduzem a variável latente nas respostas ordinais de cada item i das escalas Likert. As distribuições *a priori* para cada um dos parâmetros são mostradas, sendo K o número de níveis de respostas ordinais.

A geração da amostra MCMC das distribuições *a posteriori* dos parâmetros é feita da seguinte maneira: parte-se de uma posição inicial no espaço de parâmetros; escolhe-se então um dos parâmetros e um novo valor para ele é aleatoriamente selecionado considerando a sua distribuição de probabilidade condicionada aos dados e aos outros parâmetros; esse novo conjunto de parâmetros é mais uma observação da amostra MCMC, que vai crescendo à medida que se repete esse procedimento (Kruschke, 2015). Cada novo passo na construção da amostra MCMC leva em consideração os dados e as distribuições *a priori* dos parâmetros, mas, como novos valores para os parâmetros são gerados aleatoriamente, é possível que se selecione um novo limite que cause uma situação em que $\theta_{k-1} > \theta_k$. Para contornar essa limitação, no modelo enviado ao JAGS, o sistema que constrói a amostra MCMC, Kruschke (2015) determina que $P(y = k)$ é o valor máximo entre zero e a probabilidade calculada com a função acumulada. Desse modo, se a probabilidade calculada for negativa, o limite escolhido levaria a $P(y = k) = 0$ e, desde que os dados contenham pelo menos uma resposta no nível ordinal k , esse limite candidato seria rejeitado. Quando algum nível de resposta não está presente no conjunto de dados, não há nada (nem no modelo teórico nem no modelo enviado ao JAGS) que impeça limites invertidos e, consequentemente, probabilidades negativas.

Essa limitação do modelo é especialmente problemática quando o conjunto de dados é

pequeno, o que aumenta as chances de existirem níveis de resposta vazios. A escolha de Kruschke (2015) para lidar com níveis vazios é comprimir os dados e eliminar esses níveis. Se, por exemplo, inicialmente são considerados cinco níveis (1-2-3-4-5) mas nos dados o nível 2 não está presente (1-3-4-5), a implementação de Kruschke (2015) altera os dados e os mapeia considerando apenas quatro níveis (1-2-3-4). Esses novos dados, agora sem níveis vazios, são os utilizados para a inferência Bayesiana. Entretanto, é fácil perceber que esse mapeamento pode prejudicar a estimação dos parâmetros: se os dados sem o nível 2 tiverem sido gerados, por exemplo, por uma distribuição latente de média igual a 4, a compressão fará com que a média estimada seja deslocada para algum valor menor que 4.⁵

Table 4.3: Vantagens e desvantagens de cada método considerado para análise de dados ordinais com níveis vazios.

	Vantagens	Desvantagens
Manter níveis vazios	Os dados não são alterados	Possíveis probabilidades negativas
Comprimir níveis vazios	Sem probabilidades negativas	Dados são alterados Envies a estimação
Distribuições <i>a priori</i> de θ_k restritas	Os dados não são alterados Sem probabilidades negativas	Envies a estimação
Adicionar dados extras	Sem probabilidades negativas	Dados são alterados Estimativa de σ inflada

Além da possibilidade de se enviesar as estimativas, como os níveis ordinais estão relacionados com as opções de respostas nos questionários, comprimir os níveis vazios não foi considerada uma solução adequada para este trabalho. Algumas alternativas foram propostas e testadas com conjuntos de dados simulados. A Tabela 4.3 resume os métodos considerados e as vantagens e desvantagens de cada um. Definir distribuições *a priori* mais restritas para os limites, mais especificamente com desvios padrão menores, resolve o problema das probabilidades negativas, porém envies a estimação, já que distribuições *a priori* estreitas indicam forte crença inicial sobre o valor do parâmetro, o que não se tem no contexto deste trabalho. Não se tem informações prévias para centrar as distribuições *a priori* dos limites em valores diferentes para cada item e centrá-las de maneira restrita em valores iguais confronta a ideia do modelo de que os itens acessam a variável latente a partir de limites diferentes. Outra solução é a adição de dados extras para eliminar níveis vazios. Três maneiras de se fazer isso foram consideradas: adicionar uma resposta em

⁵Essa questão foi discutida inclusive com o autor do livro e do artigo usados como referência para o modelo ordinal, John Kruschke. Segundo ele, para que o número de níveis originais seja mantido mesmo quando existir níveis vazios nos dados, seria necessário mudar o modelo (usando distribuições *a priori* mais restritas para os limites, por exemplo) e o mecanismo para seleção de limites candidatos durante a geração da amostra MCMC.

todos os níveis, sejam eles vazios ou não, tentando evitar deslocamento da estimativa de μ ; adicionar o mínimo de respostas possível, ou seja, uma única resposta extra apenas em níveis vazios; e adicionar uma resposta em cada nível vazio e adicionar em níveis não vazios de modo a manter as probabilidades de cada nível o mais próximo das originais possível, tendo como base um tamanho amostral final com K (número de níveis) dados a mais. Porém, sempre que dados extras foram adicionados, a estimativa de σ foi inflada, com maior credibilidade sendo dada a valores maiores que o real.

A escolha de como lidar com níveis vazios levou em consideração dois pontos principais: os resultados deveriam ser matematicamente coerentes, isto é, sem probabilidades negativas, e os ajustes no modelo não poderiam favorecer as hipóteses do trabalho. A partir de testes feitos com diferentes amostras simuladas, adicionar dados extras foi a abordagem que apresentou os melhores resultados. Com ela, não se tem probabilidades negativas e apenas estimativas de σ foram prejudicadas de maneira mais significativa, o que torna mais difícil a validação das hipóteses deste trabalho (o parâmetro σ aparece no denominador do tamanho de efeito, calculado posteriormente com as amostras MCMC). As três propostas para se adicionar dados extras tiveram desempenhos muito parecidos na estimação dos parâmetros e a adição de dados extras apenas em níveis vazios foi escolhida por apresentar bons resultados de maneira consistente durante os testes, enquanto outros métodos variavam um pouco mais e, em alguns casos, geravam estimativas que favoreciam a observação de efeitos.

Outro ponto importante sobre o modelo é a sua forte dependência aos valores de limites fixados, já que fixar alguns deles condiciona todos os outros. Fixar média e/ou desvio padrão dificultaria a interpretação dos resultados e fixar os limites extremos em valores relacionados à escala de respostas nos questionários, como proposto por Kruschke (2015), parece uma solução melhor. Apesar de Liddell & Kruschke (2018) sugerirem que os limites extremos apenas do primeiro item sejam fixados, pode-se dizer que a ideia do modelo é que os limites extremos de *todos* os itens são fixos em 1,5 e $K - 0,5$, por causa da escala de respostas, e que apenas os níveis intermediários variam para cada um, porque, caso contrário, os resultados são diferentes dependendo de qual item é tratado como o primeiro na análise e uma discussão maior sobre isso seria necessária. Mesmo assim, neste trabalho segue-se a sugestão de Liddell & Kruschke (2018) para dar mais liberdade para a análise, ao invés de, por exemplo, fixar os limites extremos de todos os itens. Escolhendo de maneira arbitrária, o item tratado como o primeiro é o primeiro dos apresentados na Tabela 4.1 para cada escala. Entretanto, os valores escolhidos estão relacionados com a escala de respostas que é apresentada nos questionários e são por ela justificados. Nas análises frequentistas e na análise Bayesiana de tempo e erros, os dados são pareados e a diferença entre as configurações é utilizada. Para as medidas subjetivas tratadas com o modelo ordinal, se for tomada a diferença nas pontuações dos itens separadamente, considerando as cinco opções de resposta nos questionários deste trabalho (ver Seção 4.3), os dados resultantes

poderiam assumir nove níveis ordinais (de -4 a 4). Se, por exemplo, na condição EX a resposta a um item é *(5) concordo totalmente* e na condição EXIM é *(1) discordo totalmente*, a observação pareada seria $x_{\text{EXIM}} - x_{\text{EX}} = 1 - 5 = -4$, nível mais baixo da nova escala ordinal; se o contrário acontece, a observação final seria $x_{\text{EXIM}} - x_{\text{EX}} = 5 - 1 = 4$, caracterizando o nível mais alto da nova escala ordinal; se a resposta é a mesma em ambas as condições, a observação pareada seria igual a 0 , que corresponderia ao nível central da escala ordinal para análise pareada. Esses nove níveis não têm a relação direta com as opções de resposta que os cinco níveis originais têm, o que enfraquece a justificativa para os limites extremos fixos em $1,5$ e $K - 0,5$. Além disso, parear os dados aumenta o número de níveis e, conseqüentemente, as possibilidades de níveis vazios. Considerando a abordagem escolhida para lidar com níveis vazios, isso pode significar mais dados extras adicionados, interferindo cada vez mais no conjunto de dados e nas estimações. Por esses motivos, as análises Bayesianas das variáveis de percepção neste trabalho são feitas considerando dois grupos ao invés de apenas um com a diferença entre as condições.

4.5 Conclusão

Este capítulo apresentou as hipóteses de trabalho, de que a combinação de comunicações explícitas e implícitas melhora variáveis de desempenho e de percepção na interação entre humanos e agentes virtuais. As medidas objetivas são tempo e número de erros na execução da tarefa, registradas pelo sistema, e as medidas subjetivas são aceitação, sociabilidade e transparência da agente virtual e percepção de eficiência da interação, acessadas a partir de questionários com escalas Likert.

No experimento de interação proposto, semelhante a um jogo, participantes interagem com agentes virtuais em duas configurações de comunicação: uma com apenas comunicações explícitas e outra com a combinação de explícitas e implícitas. Resultados são analisados com NHST frequentista e estimação de parâmetros Bayesianas, como resumido na Tabela 4.4.

Table 4.4: Resumo das análises estatísticas para cada variável de interesse no trabalho.

	Medidas objetivas	Medidas subjetivas
NHST frequentista	Análise pareada: $x_{\text{EX}} - x_{\text{EXIM}}$ Teste t (ou de Wilcoxon) Hipótese nula: $\mu = 0$ Hipótese alternativa: $\mu > 0$	Análise pareada: $x_{\text{EXIM}} - x_{\text{EX}}$ Teste de Wilcoxon Hipótese nula: $\tilde{\mu} = 0$ Hipótese alternativa: $\tilde{\mu} > 0$
Estimação de parâmetros Bayesianas	Análise pareada: $x_{\text{EX}} - x_{\text{EXIM}}$ Modelo métrico com distribuição t Parâmetros: μ, σ e ν	Análise com dois grupos Modelo ordinal com distribuição t acumulada Parâmetros: $\mu_g, \sigma_g, \nu_g, \theta_1^{[i]}, \dots, \theta_{K-1}^{[i]}$

5

RESULTADOS E AVALIAÇÃO DOS
EXPERIMENTOS DE INTERAÇÃO

A realização do experimento de interação foi aprovada pelo Comitê de Ética em Pesquisa da UFMG e a pesquisa está identificada pelo número CAAE (Certificado de Apresentação para Apreciação Ética) 44110621.5.0000.5149¹. Os experimentos foram realizados entre julho e dezembro de 2021 e, por causa das restrições impostas pela pandemia de COVID-19, participantes foram pessoas da comunidade da UFMG, em especial aqueles que já estavam realizando atividades presenciais na Escola de Engenharia (onde a infraestrutura de comunicação foi montada), e pessoas próximas aos pesquisadores e alheias ao trabalho.

Ao todo, 30 pessoas participaram do experimento de interação. Quatro delas não são consideradas na comparação entre as configurações de comunicação por terem tido algum desvio avaliado como significativo em relação ao protocolo experimental: participante 4 interagiu apenas com uma agente virtual, porque o experimento foi interrompido após um problema com o sistema de rastreamento da cadeia cinemática do humano; participante 9 preencheu apenas uma parte do questionário de avaliação após a interação com a primeira agente virtual, finalizando-o apenas depois de interagir com a segunda agente virtual; participante 10 teve dificuldade de entender as instruções para a fase 1 e pediu ajuda para a experimentadora durante a interação, além de um comando externo ter sido enviado em momento inadequado; participante 22 não entendeu as instruções da fase 1 e executou a

¹A aprovação pode ser confirmada acessando-se <https://plataformabrasil.saude.gov.br/> e informando o número CAAE na opção *Confirmar Aprovação pelo CAAE ou Parecer*.

tarefa de modo atípico em relação ao esperado e a outros participantes, impossibilitando a contabilização de erros. Com isso, o tamanho amostral final é de 26 participantes e a Figura 5.1 resume o perfil do conjunto.

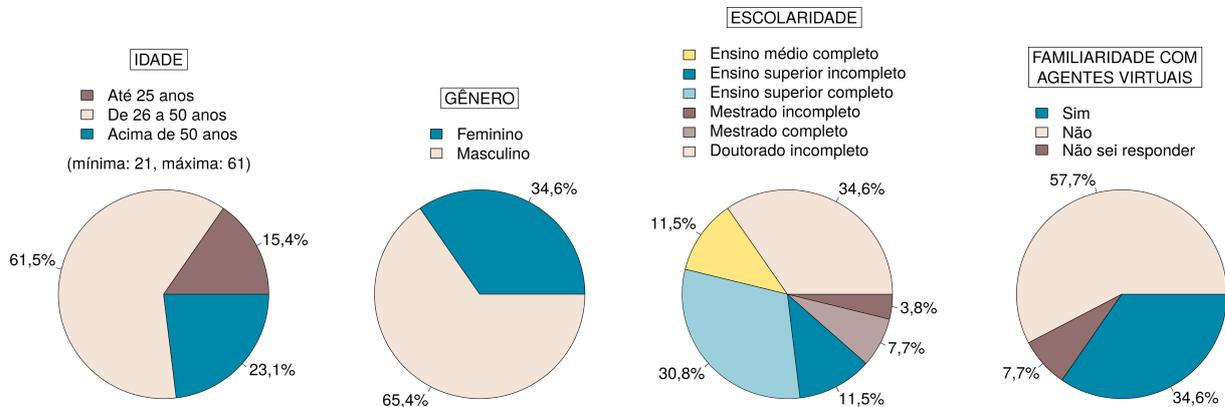


Figure 5.1: Resumo do perfil de participantes do experimento de interação.

Outros participantes também passaram por dificuldades com o reconhecimento da cadeia cinemática e, algumas vezes, foi necessário que a experimentadora intervisse, dando alguma dica ou reiniciando o sistema. Quando essa intervenção era mínima (sistema reiniciado remotamente ou com a experimentadora dando uma instrução rápida), o experimento continuava normalmente; se uma intervenção maior fosse necessária (reiniciar os sistemas localmente) e o problema tivesse ocorrido na primeira configuração (ou seja, a primeira agente virtual tinha apenas se apresentado e nenhuma instrução sobre as tarefas tinha sido dada ainda), o experimento era reiniciado em seguida ou em outro dia. Além disso, algumas vezes participantes não entenderam que a tarefa seria repetida e chamaram a experimentadora após a finalização com a primeira agente virtual. Nesses casos, o experimento continuava normalmente após breve instrução. A Tabela 5.1 lista ocorrências durante o experimento e suas respectivas resoluções, destacando os casos de exclusão.

Após a finalização de todos os experimentos, gravações das participações foram analisadas e os dados para análise foram ajustados. Para a fase 1, foram desconsiderados erros de gestos detectados quando a pessoa não estava apontando e erros contabilizados por atraso na sinalização de erros e acertos durante a interação (em alguns momentos, o sinal demorava e a pessoa mantinha o gesto de apontar, em espera, fazendo com que dois gestos fossem registrados ao invés de um, sendo pelo menos um deles sempre contabilizado como erro). Erros que poderiam ser interpretados como consequências indiretas de limitações e erros dos sistemas ou por falta de familiaridade de participantes com os sinais e os tempos da interação² foram mantidos, porque tais interpretações foram consideradas subjetivas e preferiu-se uma abordagem mais conservadora. Do total de erros em cada fase considerados

²Por exemplo, se a pessoa muda a indicação do objeto depois de tentar apontar para a cor correta e o sistema não reconhecer o gesto; se a pessoa continua apontando para um mesmo objeto após o sinal de erro ou de acerto; ou se indica as cores rapidamente em sequência, sem esperar os sinais e sem entender os tempos do sistema.

Table 5.1: Descrição de desvios do protocolo experimental e suas respectivas resoluções. Linhas destacadas indicam casos de desvios considerados significativos, que resultaram na exclusão de participante para comparação entre as configurações de comunicação.

Participante	Ocorrências	Resoluções
2	1. Dificuldade na detecção da cadeia cinemática na primeira configuração.	1. Sistema reiniciado localmente enquanto participante esperava do lado de fora da sala. Experimento recomeçou em seguida.
4	1. Dificuldade na detecção da cadeia cinemática na primeira configuração. 2. Rastreamento foi perdido durante a interação e na segunda configuração, novamente houve dificuldade na detecção da cadeia cinemática.	1. Sistema reiniciado localmente enquanto participante esperava do lado de fora da sala. Experimento recomeçou em seguida. 2. Como uma interação completa já tinha sido realizada e a experimentadora precisaria reiniciar os sistemas localmente mais uma vez, o experimento foi interrompido.
8	1. Dificuldade na detecção da cadeia cinemática na primeira configuração.	1. Experimento foi interrompido e iniciado novamente em outro dia.
9	1. Após a interação com a primeira agente virtual, participante não terminou de preencher o questionário de avaliação.	1. Participante finalizou o primeiro questionário apenas após interação com a segunda agente virtual, o que pode ter enviesado as respostas.
10	1. Na fase 1 da primeira configuração, participante chamou a experimentadora na sala para tirar dúvidas. 2. Comando para início da segunda configuração foi enviado enquanto participante ainda respondia o questionário.	1. Experimentadora respondeu as perguntas e experimento continuou, entretanto as tarefas deveriam ser executadas sem o auxílio da experimentadora. 2. Sistemas reiniciados remotamente e experimento continuou.

(continua na próxima página)

(continuação da Tabela 5.1)

Participante	Ocorrências	Resoluções
11	1. Dificuldade na detecção da cadeia cinemática na primeira configuração.	1. Sistemas reiniciados remotamente e experimento continuou.
22	1. Participante não entendeu instruções da fase 1 e moveu os objetos, colocando-os na ordem correta na frente da agente virtual.	1. Na primeira configuração, tempo limite foi alcançado e agente virtual finalizou a senha. Na segunda configuração, enquanto participante manipulava os objetos, indicações corretas foram detectadas e a senha foi preenchida. Não é possível realizar a contagem de erros como definido no protocolo experimental.
23	1. Participante se dirigiu para fora da sala após a interação com a primeira agente virtual. 2. Dificuldade na detecção da cadeia cinemática na segunda configuração.	1. Experimentadora instruiu participante a voltar e terminar o experimento. 2. Sistemas reiniciados remotamente, experimentadora instruiu participante para que detecção fosse concluída e experimento continuou.
25	1. Dificuldade na detecção da cadeia cinemática na primeira configuração.	1. Sistemas reiniciados remotamente, experimentadora instruiu participante para que detecção fosse concluída e experimento continuou.
26	1. Dificuldade na detecção da cadeia cinemática na primeira configuração.	1. Experimento foi interrompido e iniciado novamente em outro dia.
27	1. Participante se dirigiu para fora da sala após a interação com a primeira agente virtual.	1. Experimentadora instruiu participante a voltar e terminar o experimento.

na análise, 7,7% dos erros na fase 1 e 16,1% dos erros na fase 2 foram de cores não inseridas e valores de contagem não preenchidos por causa do tempo limite (ver Seção 4.3).

As próximas seções apresentam os resultados da comparação entre as duas configurações de comunicação entre humano e agente virtual, EX e EXIM. São mostradas as análises com NHST frequentista e estimação de parâmetros Bayesiana para todas as variáveis de interesse: tempo e número de erros na execução da tarefa, aceitação, sociabilidade e transparência das agentes virtuais e percepção de eficiência das interações. Em seguida, os resultados e os experimentos são discutidos e a Seção 5.4 apresenta análises sobre a influência de alguns fatores, como o perfil de participante e a ordem das interações, nas diferenças entre as configurações de comunicação.

5.1 Resultados das análises frequentistas

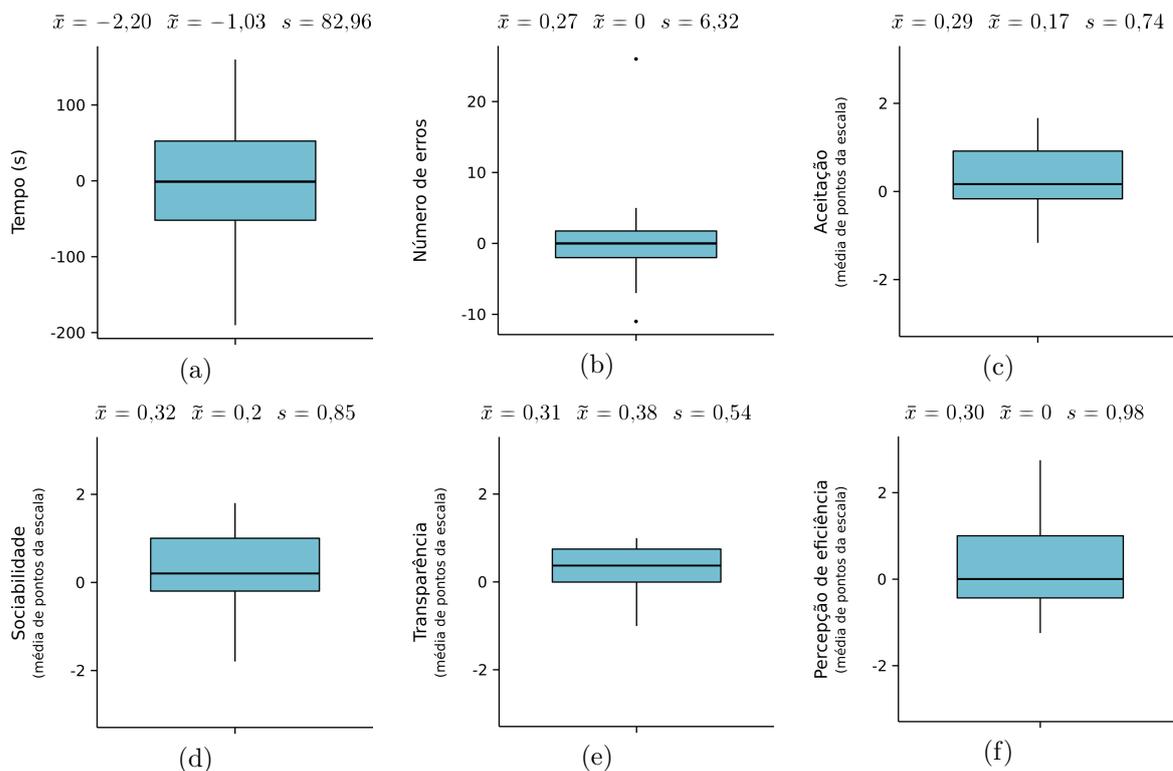


Figure 5.2: Diagramas de caixas do conjunto de dados de cada variável. Os dados são das diferenças $x_{EX} - x_{EXIM}$ para tempo e número de erros e $x_{EXIM} - x_{EX}$ para as variáveis de percepção, de modo que valores positivos são favoráveis às hipóteses deste trabalho. Média (\bar{x}), mediana (\tilde{x}) e desvio padrão (s) amostrais são indicados para cada conjunto.

A Figura 5.2 mostra diagramas de caixa resumindo o conjunto de dados de cada uma das variáveis e os valores amostrais de média (\bar{x}), mediana (\tilde{x}) e desvio padrão (s). Exceto para o conjunto de dados de tempo (em segundos), média e mediana para todas as outras variáveis foram sempre maiores ou iguais a zero. O espalhamento de dados é alto para a diferença de tempo e número de erros entre as configurações, sendo que para esta última

foram detectados dois valores atípicos (-11 e 26). Para as variáveis de percepção, que podiam assumir valores entre -4 e 4 (diferenças entre médias que podiam assumir valores de 1 a 5), as médias amostrais são parecidas e o espalhamento é maior para a medida de percepção de eficiência da interação e menor para a de transparência da agente virtual.

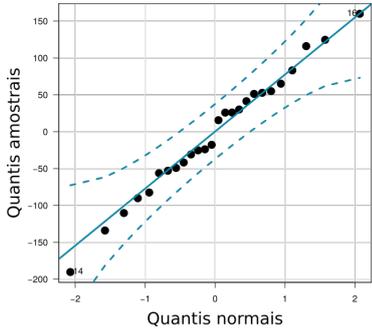
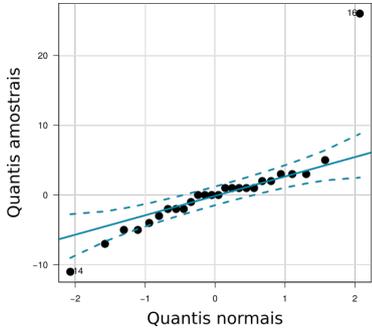
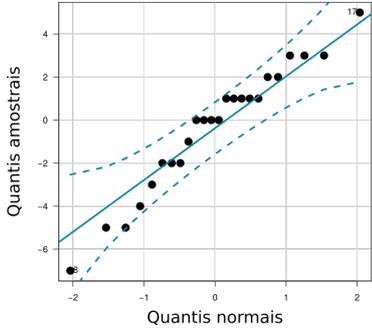
Considerando a maneira com que as diferenças foram tomadas para cada variável (ver Seção 4.4.1), diferenças positivas são observações favoráveis às hipóteses deste trabalho (ver Seção 4.1). A taxa aceitável de falsos positivos é de 5% , portanto, valores- p menores que $0,05$ indicam baixa probabilidade de se obter a estatística de teste sob a hipótese nula, que é então rejeitada.

A Tabela 5.2 mostra os resultados das análises com métodos NHST frequentistas para as diferenças de tempo e número de erros ($x_{EX} - x_{EXIM}$). A primeira coluna de cada variável mostra a verificação da premissa de que a distribuição de médias amostrais é normal a partir da análise da normalidade do conjunto de dados. Utilizando o teste de Shapiro-Wilk³ (a hipótese nula é de que a distribuição é normal, então se o valor- p do teste é maior que $0,05$, não se rejeita a hipótese nula) e a análise do gráfico quantil-quantil (se os dados forem normais, eles formam aproximadamente uma linha reta), decide-se qual teste será usado. O resultado do teste aplicado é mostrado na última coluna da tabela. Como foram identificados dois possíveis valores atípicos na amostra do número de erros, as análises foram feitas também para o conjunto de dados sem os valores atípicos e são mostradas na última linha da tabela. Em nenhum dos casos o valor- p foi menor que o limite de 5% , ou seja, não é possível rejeitar a hipótese nula de que não há diferença em tempo e número de erros entre as duas configurações de comunicação consideradas.

A Tabela 5.3 mostra os resultados dos testes com a diferença entre as médias de pontos nas escalas de aceitação, sociabilidade, transparência e percepção de eficiência das duas configurações de comunicação ($x_{EXIM} - x_{EX}$). Apesar de os diagramas de caixa sugerirem dados não simétricos para aceitação, sociabilidade e percepção de eficiência, o teste de Wilcoxon foi ainda assim aplicado, já que a distribuição das diferenças é considerada simétrica (ver Seção 4.4.1.1). Para a variável de aceitação, o valor- p obtido é bem próximo do limite de 5% , sinalizando algum indício de que a diferença entre as configurações pode não ser igual a zero. Para sociabilidade e transparência, o valor- p foi menor que o limite, indicando rejeição da hipótese nula em relação à hipótese alternativa de que a diferença é maior que zero. Entretanto, o conjunto de dados de transparência continha cinco observações nulas, *i.e.*, iguais ao valor sob a hipótese nula, descartadas para a análise com o teste de Wilcoxon, como descrito na Seção 4.4.1.1. Observações nulas são favoráveis à hipótese nula, então descartar essas cinco observações pode ter enviesado o teste contra ela e, por isso, esse resultado deve ser tratado com cautela. Por fim, o teste para a diferença da percepção de eficiência apresentou valor- p maior que o limite, sugerindo que a hipótese

³Função `shapiro.test` do pacote `stats` (versão 4.0.4), <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/shapiro.test.html>.

Table 5.2: Resultados dos testes de significância de hipótese nula para a diferença de tempo e número de erros entre as configurações de comunicação.

	Normalidade	Teste
Tempo	<p><i>Shapiro-Wilk:</i> valor-$p = 0,9926 > 0,05$</p> 	<p><i>Teste t</i></p> <p>$T = -0,13543$ (Equação (4.2))</p> <p>Valor-$p: 0,5533 > 0,05$</p>
	<p><i>Shapiro-Wilk:</i> valor-$p = 1,482 \times 10^{-5} < 0,05$ (desvio significativo de normalidade)</p> 	<p><i>Teste de Wilcoxon</i></p> <p>$T = 116$ (Equação (4.3))</p> <p>Valor-$p: 0,6348 > 0,05$</p> <p>Observações nulas: 4</p> <p>Grupos de empates: 3</p>
Erros sem valores atípicos	<p><i>Shapiro-Wilk:</i> valor-$p = 0,4394 > 0,05$</p> 	<p><i>Teste t</i></p> <p>$T = -0,5547$ (Equação (4.2))</p> <p>Valor-$p: 0,7078 > 0,05$</p>

nula não seja rejeitada. Nesse caso, a amostra continha seis observações nulas e, mesmo sem elas, a estatística de teste obtida continua sendo considerada provável sob a hipótese nula.

Table 5.3: Resultados dos testes de significância de hipótese nula para a diferença da média de pontos nas escalas de cada variável de percepção entre as configurações de comunicação.

Teste de Wilcoxon		
Aceitação	$T = 192$ (Equação (4.3)) Valor- p : $0,05121 > 0,05$	Observações nulas: 3 Grupos de empates: 0
Sociabilidade	$T = 201,5$ (Equação (4.3)) Valor- p : $0,02642 < 0,05$	Observações nulas: 3 Grupos de empates: 0
Transparência	$T = 188,5$ (Equação (4.3)) Valor- p : $0,004566 < 0,05$	Observações nulas: 5 Grupos de empates: 2
Percepção de eficiência	$T = 138,5$ (Equação (4.3)) Valor- p : $0,1091 > 0,05$	Observações nulas: 6 Grupos de empates: 2

Tamanho de efeito é uma medida do grau em que um efeito está presente e pode ser calculado como a diferença entre o parâmetro amostral e o valor nulo do parâmetro, relativa ao desvio padrão amostral, ou seja,

$$d = \frac{\bar{x} - \mu_0}{s}$$

para a média (Cohen, 1988). A partir do contexto de cada análise, define-se quais valores de tamanho de efeito são significativos. Para quando essa classificação não está bem definida, como é o caso deste trabalho, Cohen (1988) propõe uma convenção que define que tamanhos de efeito de 0,2 sejam considerados pequenos, de 0,5 sejam considerados médios e de 0,8 sejam considerados tamanhos de efeito grandes. Para as variáveis com indícios para rejeição da hipótese nula, incluindo aceitação, com o valor- p bem próximo do limite definido, os tamanhos de efeito, calculados com a mediana \tilde{x} , são: $d = 0,23$ para aceitação, $d = 0,23$ para sociabilidade e $d = 0,69$ para transparência. Para aceitação e sociabilidade, o tamanho de efeito calculado é considerado pequeno e para transparência está entre médio e grande.

A potência de um teste de hipóteses frequentista é dada pela probabilidade de se rejeitar a hipótese nula quando a hipótese alternativa é verdadeira, ou seja, é uma medida de sensibilidade do teste para detectar diferenças. O cálculo da potência depende da distribuição real dos dados, então para testes como o de Wilcoxon, que não assumem distribuições específicas, é ainda mais difícil se calcular e um valor exato muitas vezes não está disponível (Sprenst & Smeeton, 2001; Hollander et al., 2014). A potência do teste t de

uma amostra para detectar tamanhos de efeito de 0,2 com limite de falsos positivos de 5% e hipótese alternativa unidirecional (que define uma direção específica para o efeito, maior ou menor que o valor nulo) é de 25,7% para amostras com 26 observações. Essa potência é considerada baixa e, para uma potência de 80% na detecção de tamanhos de efeito de 0,2 no teste t com hipótese alternativa unidirecional, seria necessário um amostra com 310 observações⁴.

5.2 Resultados das análises Bayesianas

Os dois últimos passos da estimação de parâmetros Bayesianas (ver Seção 4.4.2) são a interpretação das distribuições *a posteriori* dos valores dos parâmetros de cada modelo e a verificação posterior se as estimativas condizem com os dados disponíveis. Para cada combinação de parâmetros obtida pelo processo MCMC, o tamanho de efeito é calculado, gerando uma amostra para a sua distribuição *a posteriori*: para as medidas objetivas, usando a diferença entre as configurações, o tamanho de efeito calculado é

$$d = \frac{\mu - \mu_0}{\sigma},$$

considerando o valor nulo $\mu_0 = 0$; para as medidas subjetivas, para as quais as análises são feitas com dois grupos, o tamanho de efeito calculado é

$$d = \frac{\mu_{\text{EXIM}} - \mu_{\text{EX}}}{\sqrt{\frac{\sigma_{\text{EX}}^2 + \sigma_{\text{EXIM}}^2}{2}}},$$

como sugerido por Kruschke (2013) and Kruschke (2015). Uma ROPE de $-0,1$ a $0,1$, abrangendo tamanhos de efeito de até metade do valor considerado pequeno pela convenção de Cohen (1988), é definida ao redor do valor nulo para o tamanho de efeito ($d = 0$), como ilustrado pela Figura 5.3, para enriquecer a interpretação da distribuição *a posteriori*.

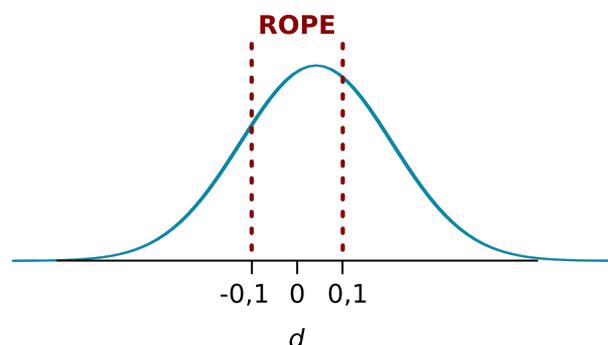


Figure 5.3: Exemplo de distribuição *a posteriori* para o tamanho de efeito d com ROPE de $-0,1$ a $0,1$, ao redor do valor nulo $d = 0$.

⁴Os cálculos relacionados à potência do teste foram feitos com a função `power.t.test` do pacote `stats` (versão 4.0.4), <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/power.t.test.html>.

5.2.1 Medidas objetivas

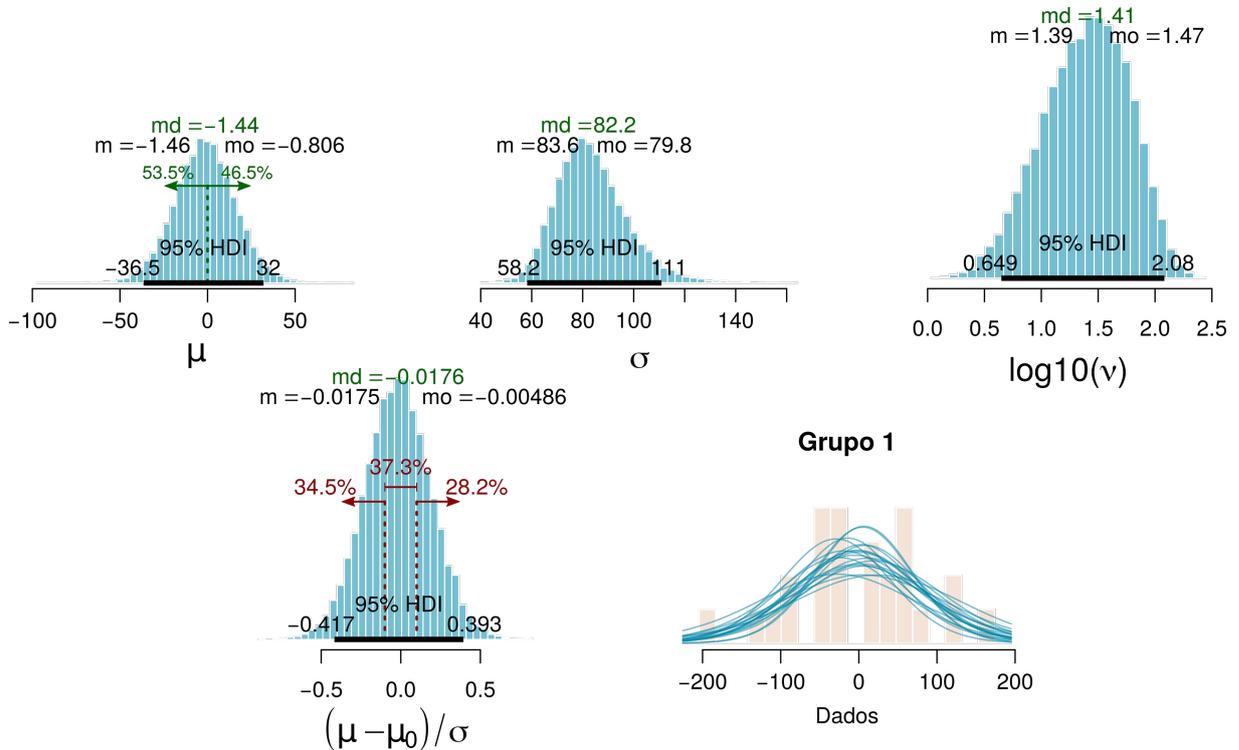


Figure 5.4: Resultados da inferência Bayesiana para a diferença de tempo $x_{EX} - x_{EXIM}$, em segundos. Na primeira linha, estão as distribuições *a posteriori* dos parâmetros μ , σ e ν (em escala logarítmica) da distribuição t e à esquerda na segunda linha, a distribuição *a posteriori* do tamanho de efeito considerando $\mu_0 = 0$. Para todas elas, estão anotadas média (m), mediana (md), moda (mo) e os limites do 95% HDI. Nas distribuições de μ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo μ_0 e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima de μ_0 e abaixo, dentro e acima da ROPE mostradas. No canto direito inferior, algumas distribuições t prováveis são sobrepostas ao histograma dos dados para checagem de adequação do modelo.

A Figura 5.4 mostra os resultados da inferência Bayesiana para a diferença de tempo (em segundos) entre as configurações de comunicação. A primeira linha da figura mostra as distribuições *a posteriori* de μ , σ e ν , sendo esta última mostrada em escala logarítmica para facilitar a visualização, já que a escala original é bastante assimétrica (a maior diferença nas caudas da distribuição t acontece em valores menores de ν). Em cada distribuição, estão anotadas a mediana (md), a média (m), a moda (mo) e os limites do 95% HDI. Para a distribuição de μ , o valor nulo é marcado com uma linha vertical pontilhada e a porcentagem da distribuição acima e abaixo dele são indicadas. O valor nulo $\mu_0 = 0$ para o tempo está entre os mais prováveis para o parâmetro μ e o 95% HDI abrange diferenças de $-36,5$ a 32 segundos. A distribuição *a posteriori* de σ dá credibilidade alta a valores entre $58,2$ e 111 segundos. O parâmetro ν na escala logarítmica pode assumir valores maiores ou iguais a zero, sendo que valores acima aproximadamente de $\log(\nu) = 1,47$ ($\nu = 30$ na

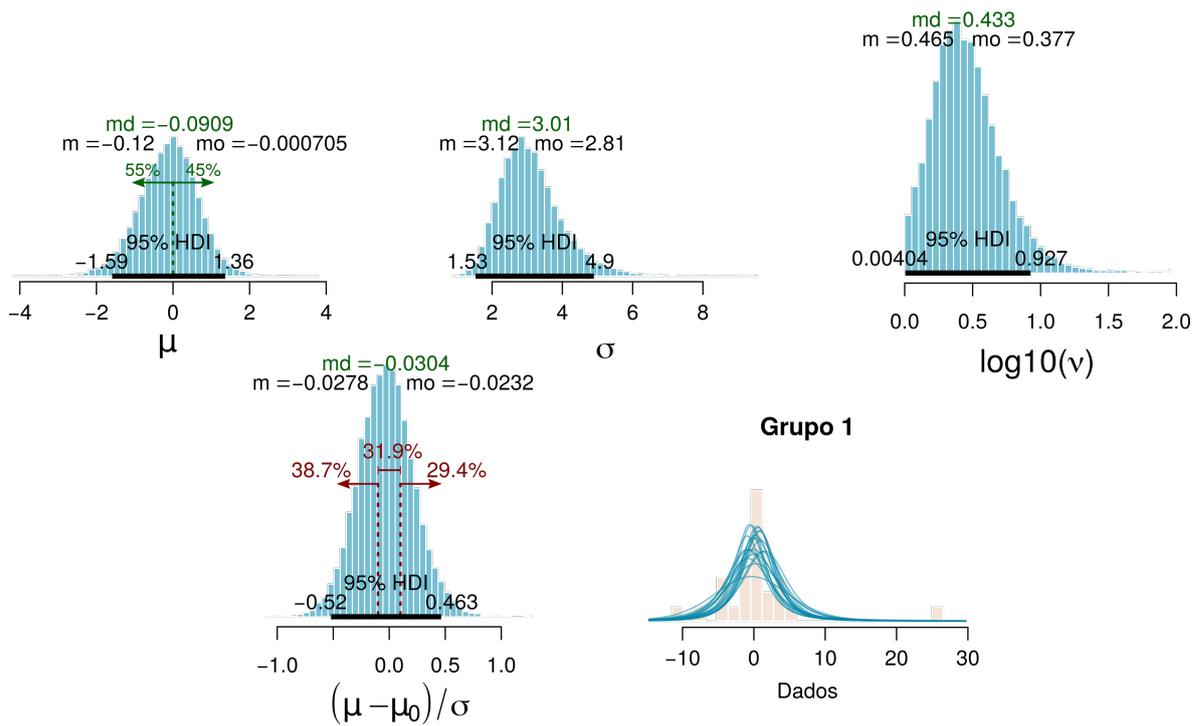
escala original) representam distribuições bem próximas da normal (Kruschke, 2015). Para a amostra de tempo, a distribuição *a posteriori* de ν tem 95% HDI largo, abrangendo valores que representam tanto distribuições com caudas mais pesadas quanto distribuições próximas da normal, ou seja, a distribuição não fornece estimativa precisa para ν , como já era esperado (ver Seção 4.4.2.1).

No canto esquerdo inferior da Figura 5.4, está a distribuição *a posteriori* do tamanho de efeito e as porcentagens da distribuição acima, dentro e abaixo da ROPE definida. Para o tamanho de efeito na diferença de tempo, os valores dentro da ROPE são bastantes prováveis, mas o 95% HDI é largo, indicando que a estimativa não é muito precisa, o que também já era esperado, considerando que a amostra é pequena. A distribuição *a posteriori* é centrada próxima ao zero, mas o 95% HDI contém valores até tamanhos de efeito quase médios nas duas direções, ou seja, não há precisão suficiente nas estimativas dos parâmetros para que se utilize a regra de decisão ilustrada na Figura 4.5 e se tire uma conclusão forte sobre a diferença no tempo de execução da tarefa entre as duas configurações de comunicação.

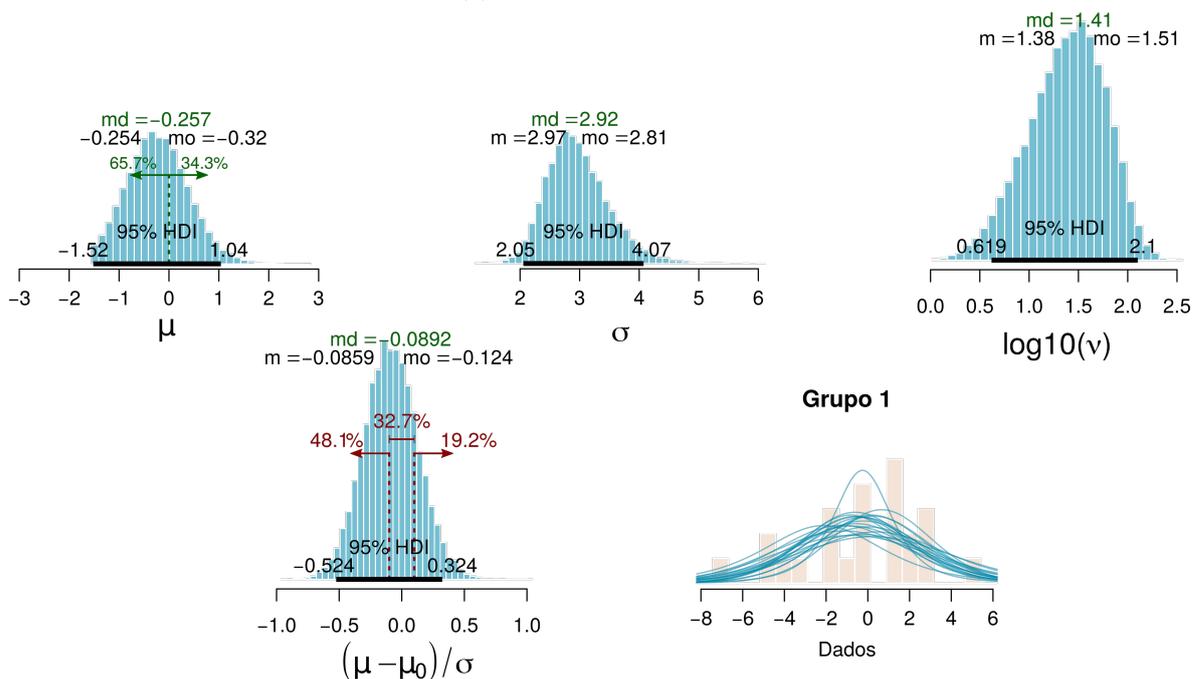
À direita, na segunda linha da Figura 5.4, é mostrada a checagem posterior. Na imagem, ao histograma dos dados sobrepõe-se algumas prováveis distribuições t , retiradas da amostra MCMC. As curvas estimadas parecem ter dificuldade em comportar os dados, porém, com um conjunto de dados tão pequeno, é difícil verificar visualmente a forma da distribuição, então, como não foram observados desvios muito críticos (fortes indícios de assimetria ou de distribuição multimodal, por exemplo), o modelo utilizado foi considerado suficientemente adequado.

A Figura 5.5a mostra os resultados da estimação de parâmetros para a diferença no número de erros entre as configurações de comunicação. As medidas de tendência central da distribuição de μ estão bem próximas de zero, com o 95% HDI abrangendo diferenças de menos de dois erros entre as condições, nas duas direções. A estimativa de σ dá credibilidade alta a valores de aproximadamente três erros, com limite superior do 95% HDI em quase cinco erros. Por causa dos valores atípicos no conjunto de dados, a distribuição *a posteriori* do parâmetro de normalidade atribui alta probabilidade a valores mais baixos de ν , ou seja, caudas mais pesadas, tentando ajustar a distribuição t para melhor acomodar os dados. Novamente, não se tem uma estimação precisa do tamanho de efeito, com a distribuição *a posteriori* muito levemente deslocada para a esquerda de zero, mas com 95% HDI contendo tamanhos de efeito médios nas duas direções. Com a checagem posterior, o modelo foi novamente considerado suficientemente adequado.

Assim como feito para o método NHST, a análise Bayesiana foi repetida para o conjunto de dados de erros sem os dois valores atípicos, em -11 e 26 . Os resultados são mostrados na Figura 5.5b. As estimativas de μ , σ e tamanho de efeito ficaram mais precisas (95% HDI mais estreitos) sem os valores atípicos, principalmente a do parâmetro de escala σ da distribuição t . A distribuição *a posteriori* do tamanho de efeito foi ligeiramente



(a) Com valores atípicos.



(b) Sem valores atípicos.

Figure 5.5: Resultados da inferência Bayesiana para a diferença de número de erros $x_{EX} - x_{EXIM}$, com e sem valores atípicos. Na primeira linha de cada figura, estão as distribuições *a posteriori* dos parâmetros μ , σ e ν (em escala logarítmica) da distribuição t e à esquerda na segunda linha, a distribuição do tamanho de efeito considerando $\mu_0 = 0$. Para todas elas, estão anotadas média (m), mediana (md), moda (mo) e os limites do 95% HDI. Nas distribuições de μ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo μ_0 e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima de μ_0 e abaixo, dentro e acima da ROPE mostradas. No canto direito inferior de cada figura, distribuições t prováveis são sobrepostas aos histogramas dos dados para checagem de adequação do modelo.

“comprimida” para a esquerda, reduzindo a porcentagem da distribuição acima do limite superior da ROPE e dando menos credibilidade a valores que são favoráveis às hipóteses deste trabalho. Entretanto, não há precisão suficiente nas estimativas para se tirar conclusões fortes sobre a existência ou não de diferença entre as duas configurações. Sem os valores atípicos, valores mais altos de ν foram considerados também prováveis, já que as caudas da distribuição t não precisam ser mais tão pesadas para acomodar os dados.

5.2.2 Medidas subjetivas

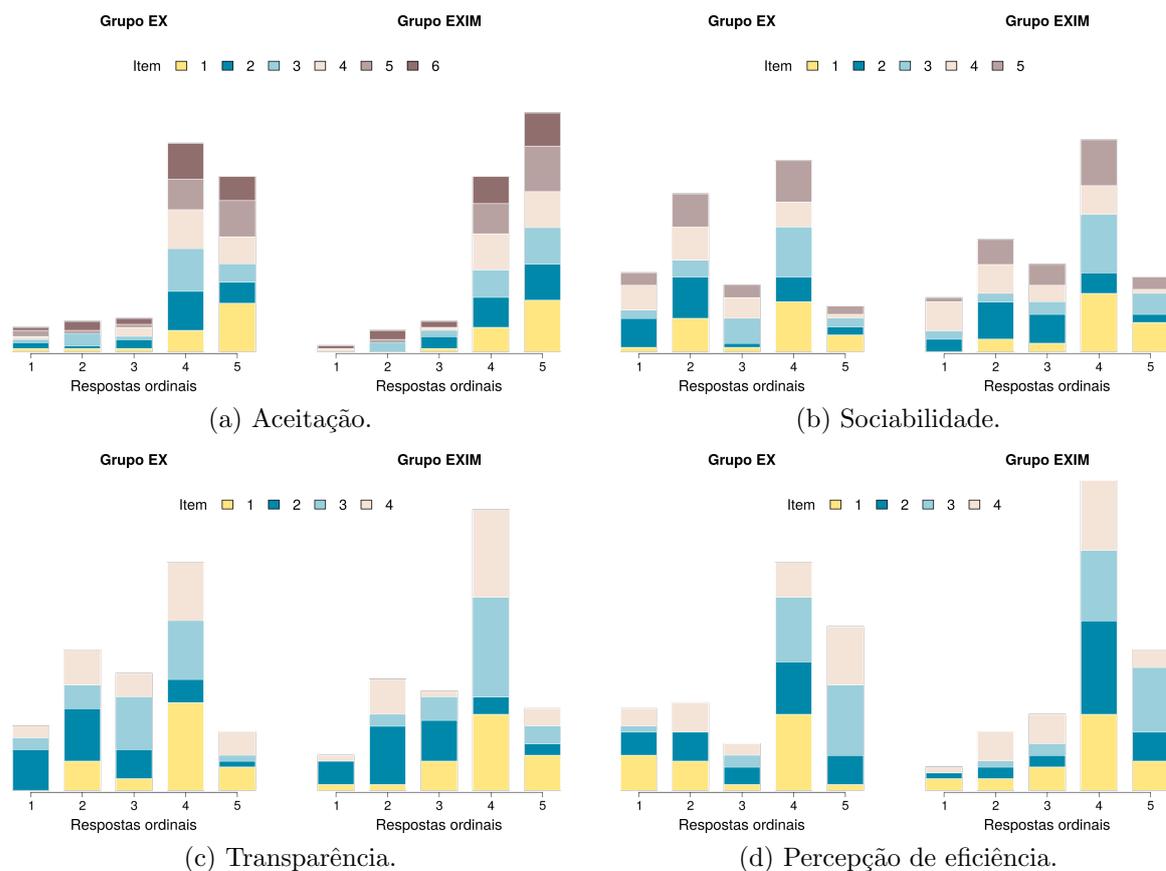


Figure 5.6: Histogramas das respostas nos itens das escalas Likert de aceitação, sociabilidade, transparência e percepção de eficiência para as configurações de comunicação EX e EXIM. As cores diferenciam os itens de cada escala.

Histogramas das respostas em cada um dos itens de aceitação, sociabilidade e transparência das agentes virtuais e a percepção de eficiência das interações para cada configuração de comunicação são mostrados na Figura 5.6. Para a escala de aceitação, as respostas se concentraram nos níveis mais altos para ambas as configurações, indicando que a agente virtual parece ter sido bem aceita independentemente do tipo de comunicação. Para as outras medidas subjetivas, os dados se espalharam mais pelos cinco níveis de resposta, mas os histogramas empilhados têm sempre o nível 4 mais alto, sendo o do grupo EXIM sempre mais alto que o do grupo EX. Em pelo menos metade dos itens nas escalas de

sociabilidade, transparência e percepção de eficiência, o número de respostas no nível 4 aumentou do grupo EX para o grupo EXIM. Já os níveis 1 e 2 são sempre mais altos no grupo EX do que no grupo EXIM para os histogramas empilhados dessas três medidas.

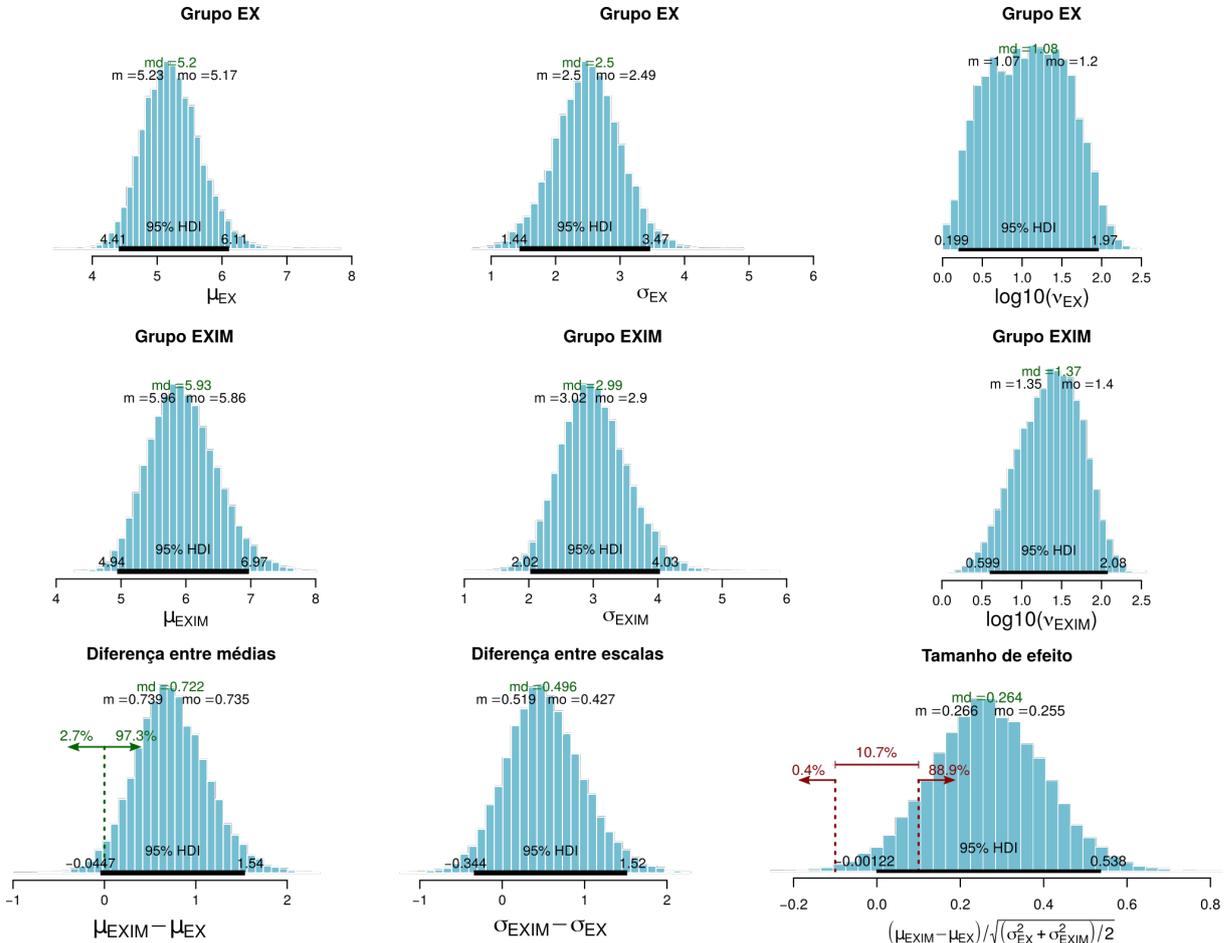


Figure 5.7: Distribuições *a posteriori* dos parâmetros μ , σ e ν (escala logarítmica) das distribuições *t* latentes de aceitação das agentes virtuais nas configurações de comunicação EX e EXIM. A última linha mostra as distribuições *a posteriori* das diferenças $\mu_{EXIM} - \mu_{EX}$ e $\sigma_{EXIM} - \sigma_{EX}$ e do tamanho de efeito. Para todas elas, estão anotadas média (*m*), mediana (*md*), moda (*mo*) e os limites do 95% HDI. Nas distribuições de $\mu_{EXIM} - \mu_{EX}$ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo ($\mu_{EXIM} - \mu_{EX} = 0$) e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima do valor nulo e abaixo, dentro e acima da ROPE mostradas.

A Figura 5.7 mostra os resultados da estimação de parâmetros para a variável de aceitação. As análises das medidas de percepção são feitas usando o modelo ordinal para dois grupos, por isso a figura mostra a estimação dos parâmetros μ , σ e ν para as duas configurações de comunicação separadamente. Para cada combinação de parâmetros na amostra MCMC, as diferenças $\mu_{EXIM} - \mu_{EX}$ e $\sigma_{EXIM} - \sigma_{EX}$ são obtidas e estas distribuições *a posteriori* também são mostradas da Figura 5.7. Na distribuição da diferença entre as médias, o valor nulo é indicado. As distribuições *a posteriori* de ν_{EX} e ν_{EXIM} da distribuição de aceitação de cada grupo têm 95% HDI abrangendo desde valores que

indicam caudas mais pesadas até aqueles que aproximam bastante a distribuição t de uma normal. Novamente, a distribuição *a posteriori* do tamanho de efeito é mostrada no canto inferior direito.

Os parâmetros μ , σ e ν são referentes às distribuições latentes, acessadas através dos itens da escala Likert, e a interpretação de μ é feita a partir dos cinco níveis ordinais de resposta (ver Seção 4.4.2.2, mais especificamente o exemplo ilustrado com a Figura 4.10). As distribuições *a posteriori* para os parâmetros μ das duas configurações dão alta credibilidade a valores mais próximos aos níveis altos da escala de respostas ordinal, ou seja, a estimativa sugere que a aceitação das agentes virtuais é alta nos dois grupos, confirmando o que se observou nos histogramas de dados (ver Figura 5.6a), sendo levemente maior no grupo EXIM. A distribuição *a posteriori* da diferença entre as médias tem 97,3% de seu volume acima do valor nulo, indicando probabilidade alta de que a aceitação seja realmente maior para a configuração EXIM. As distribuições dos parâmetros σ de cada grupo e da diferença entre eles indicam também valores maiores para a configuração de comunicações combinadas. Para o conjunto de dados de aceitação, foram adicionadas respostas extras em ambos os grupos, uma para EX (nível 2) e oito para EXIM (níveis 1, 2 e 3), portanto, se o comportamento observado nos testes com o modelo ajustado usando dados simulados, como discutido na Seção 4.4.2.2 (Tabela 4.3), se refletir nos dados reais, os valores considerados mais prováveis para σ_{EXIM} podem estar levemente maiores que o verdadeiro. A distribuição *a posteriori* do tamanho de efeito tem medidas de tendência central indicando efeito de pequeno a médio e, novamente, não se tem estimativas precisas o suficiente para uma decisão considerando o 95% HDI e a ROPE ao redor do valor nulo. Entretanto, 88,9% da distribuição está acima do limite superior da ROPE, dando alta credibilidade à existência de um efeito positivo, ou seja, de que a aceitação da agente virtual na configuração EXIM é maior.

A comparação entre a largura do 95% HDI do tamanho de efeito na Figura 5.7 e nas figuras dos resultados das variáveis objetivas (Figuras 5.4 a 5.5b) mostra que a estimativa para a variável de aceitação é mais precisa. Para tempo e erros, os parâmetros são estimados com um conjunto de dados de tamanho $N = 26$, uma observação por participante; já nas variáveis de percepção, o modelo considera que os parâmetros de μ , σ e ν do grupo são os mesmos para todos os itens, ou seja, para uma escala formada por Q itens, $26Q$ observações contribuem para a estimativa dos parâmetros comuns, tornando-a mais precisa.

Além dos parâmetros da distribuição latente, a inferência Bayesiana também estima os limites que a traduzem nas saídas ordinais. Os limites são fortemente correlacionados e as estimações são mostradas em conjunto, como sugerido por Kruschke (2015). Para ilustrar como os resultados são apresentados, a Figura 5.8 mostra as distribuições *a posteriori* dos limites do item 3 da escala de aceitação, representadas pelas nuvens de pontos azuis. O espalhamento das nuvens indica a dispersão das distribuições *a posteriori* e as linhas verticais tracejadas mostram as médias de cada limite. Para a aceitação, há

maior quantidade de dados nos níveis mais altos de respostas ordinais (ver Figura 5.6a) e, portanto, as estimativas dos limites mais altos são mais precisas que as dos limites mais baixos. As elipses na Figura 5.8 abrangem 95% das nuvens dos limites θ_1 e θ_4 e a elipse correspondente ao limite θ_1 (mais à esquerda) é mais larga que a do limite θ_4 , indicando que a distribuição *a posteriori* de θ_4 é mais compacta.

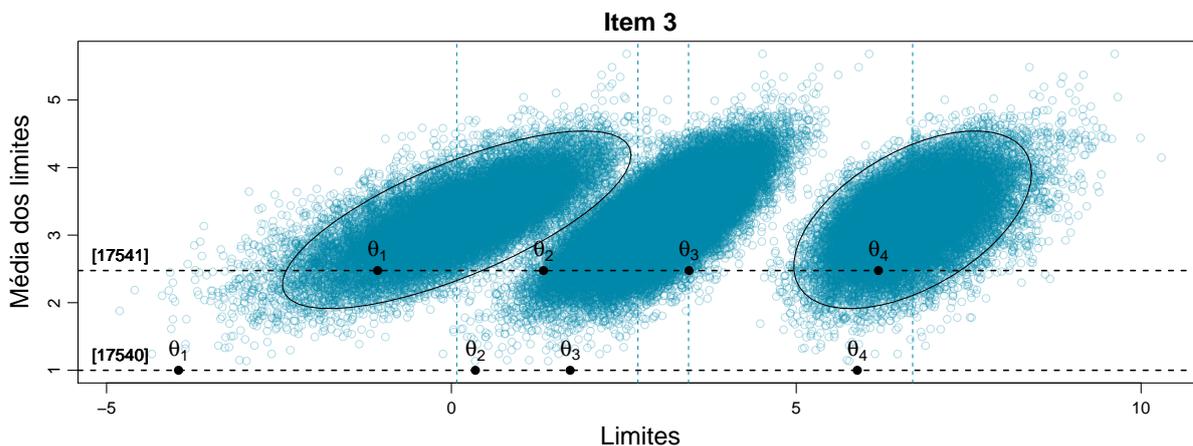


Figure 5.8: Distribuições *a posteriori* dos limites θ_1 , θ_2 , θ_3 e θ_4 do item 3 da escala de aceitação. Os pontos representam valores dos limites em cada combinação de parâmetros na amostra MCMC e a coordenada vertical é a média deles em cada combinação. Linhas verticais tracejadas mostram a média de cada distribuição e o espalhamento da nuvem de pontos indica sua dispersão. As elipses mostradas abrangem 95% dos pontos de θ_1 e θ_4 . Linhas horizontais tracejadas estão relacionadas com os passos 17540 e 17541 da construção da amostra MCMC e os limites em cada um deles são marcados com pontos pretos.

Os pontos na Figura 5.8 representam os valores dos limites em cada combinação de parâmetros na amostra MCMC e a coordenada vertical é a média de todos eles naquela combinação. Desse modo, para cada passo na construção da amostra MCMC, os pontos de todos os limites estão a uma mesma altura no gráfico. As linhas horizontais tracejadas na Figura 5.8 estão relacionadas com dois passos subsequentes na geração da amostra MCMC, de números 17540 e 17541 (a amostra MCMC gerada contém 20000 combinações de parâmetros), e a altura delas é a média dos limites (indicados pelos pontos pretos) naquele passo. Durante a construção da amostra MCMC, se, por exemplo, um valor mais alto é aleatoriamente escolhido para um limite, todos os outros limites daquele item precisam se ajustar e tendem a ser também mais altos para manter as probabilidades das saídas ordinais. Cada novo passo, portanto, acaba deslocando o conjunto para cima e para a direita ou para baixo e para a esquerda, como se observa com os passos subsequentes mostrados. Os resultados das estimativas dos limites de todos os itens da escala de aceitação são mostrados na Figura 5.9. Os limites das extremidades no item 1 estão sempre em 1,5 e 4,5, já que foram fixados nesses valores.

A checagem posterior da adequação do modelo aos dados de aceitação das agentes

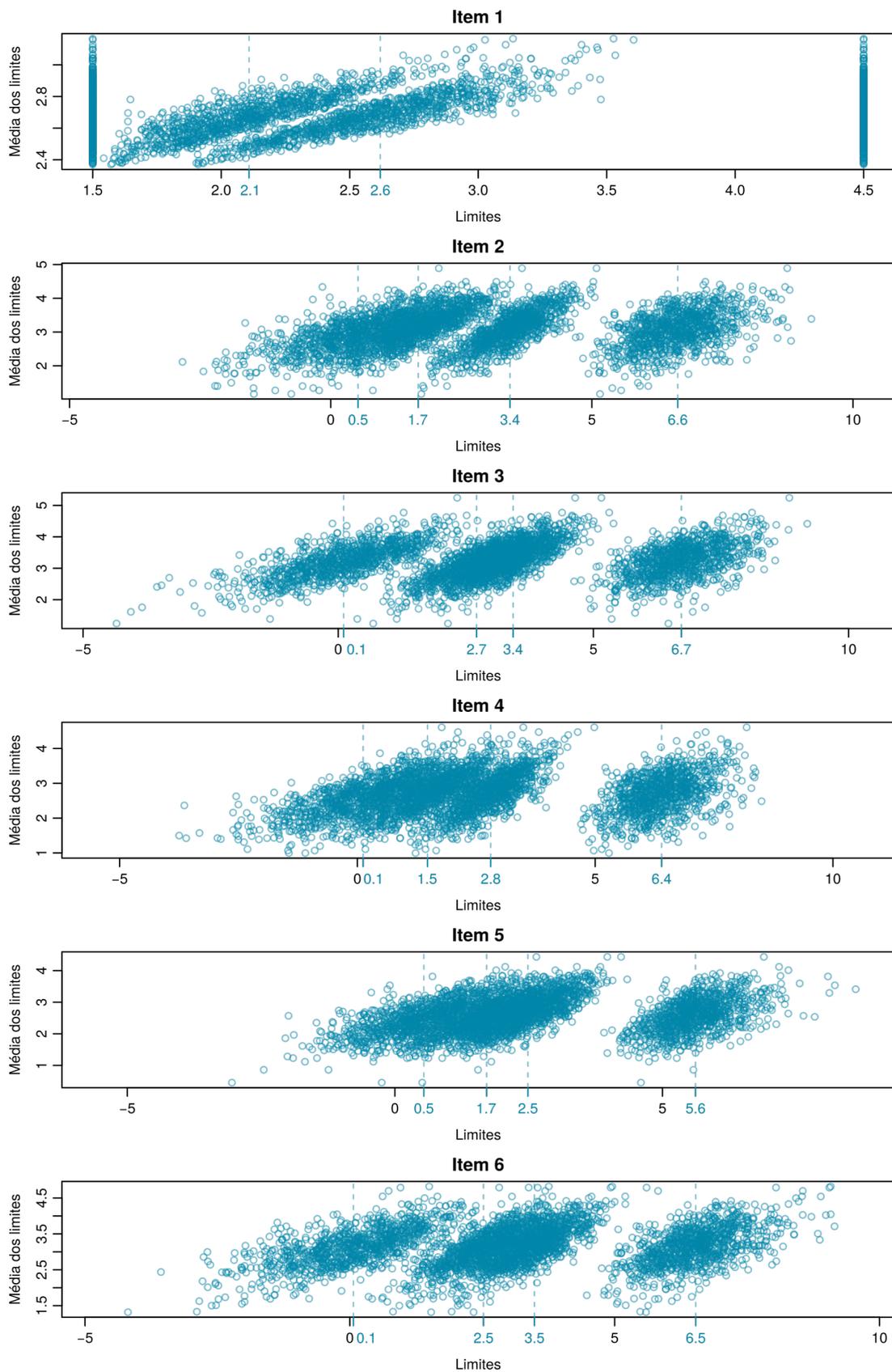


Figure 5.9: Distribuições *a posteriori* dos limites relacionados a cada item da escala Likert de aceitação das agentes virtuais. Linhas tracejadas indicam as médias das estimativas dos limites.

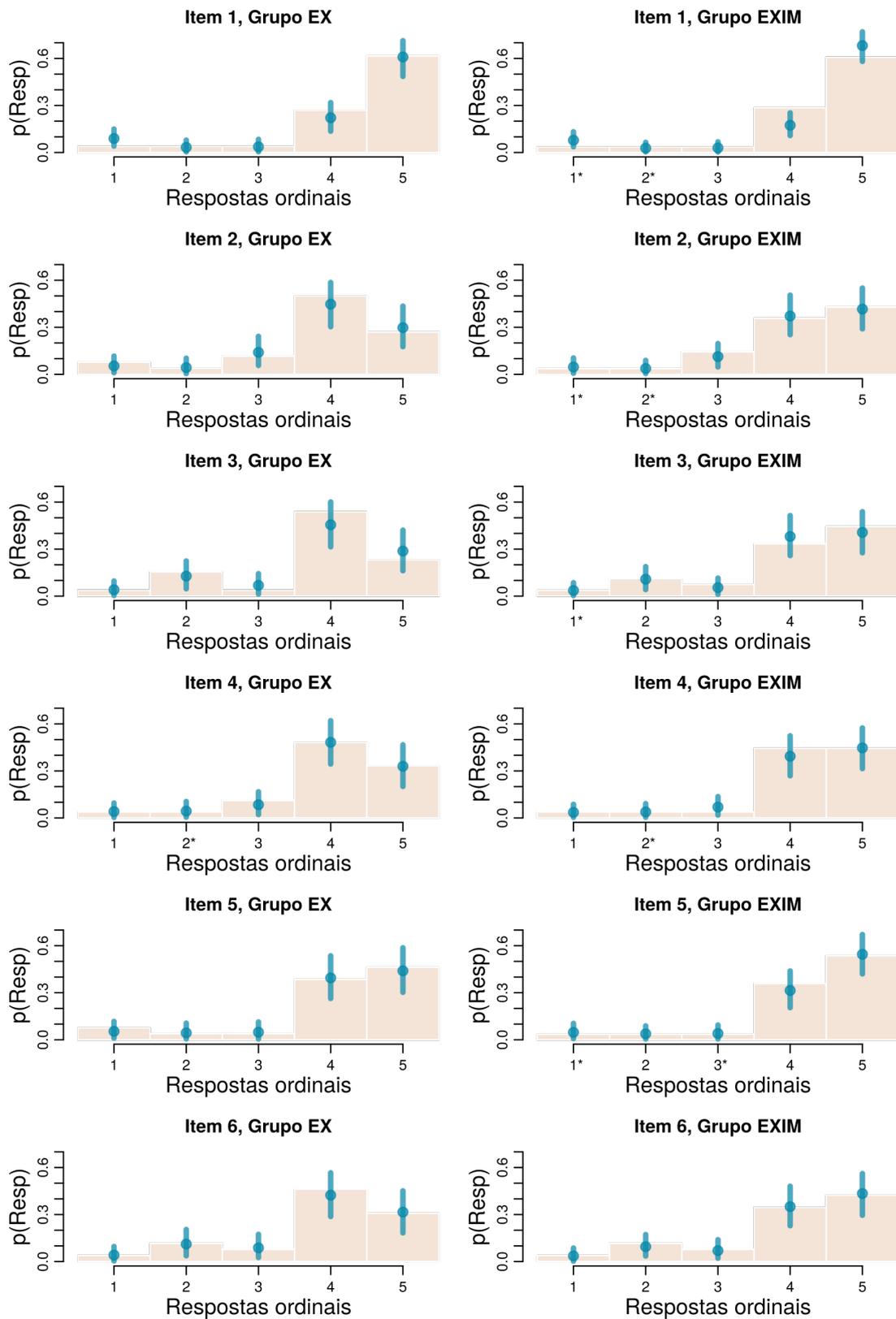


Figure 5.10: Histogramas dos dados de aceitação sobrepostos com probabilidades estimadas para checagem posterior de adequação do modelo. Níveis originalmente vazios, para os quais se adicionou uma resposta extra, são indicados por um asterisco (*).

virtuais nos dois grupos é mostrada na Figura 5.10. Nela, histogramas dos dados usados na análise (ou seja, com respostas adicionadas em níveis vazios) são sobrepostos com a probabilidade estimada de cada nível de resposta ordinal de todos os itens para cada grupo. A probabilidade é calculada para cada combinação de parâmetros na amostra MCMC e o círculo azul mostra a mediana da estimação e o segmento vertical mostra o intervalo 95% HDI. As probabilidades calculadas se aproximam dos dados e, portanto, o modelo é considerado adequado. Níveis originalmente vazios, onde foram adicionadas respostas extras (uma em cada nível), são indicados por um asterisco (*) junto aos seus rótulos no eixo de respostas ordinais dos gráficos na Figura 5.10.

As figuras com os resultados das inferências Bayesianas de sociabilidade e transparência das agentes virtuais e percepção de eficiência das interações são mostradas no Apêndice C e podem ser interpretadas seguindo o mesmo procedimento mostrado para as variáveis de tempo, erros e aceitação. A Tabela 5.4 mostra um resumo das informações consideradas mais relevantes para este trabalho sobre cada variável de percepção. A primeira linha mostra os resultados para a aceitação, retirados das distribuições mostradas na Figura 5.7, e as linhas restantes apresentam os resultados das demais medidas subjetivas, retirados das figuras no Apêndice C.

A média de sociabilidade para a agente virtual na configuração EX ficou próxima ao meio da escala ordinal, ou seja, em média, a sociabilidade não se destacou em nenhuma das duas direções. Já a sociabilidade da agente virtual na configuração EXIM foi percebida como um pouco mais alta do que na configuração EX, sendo que o 95% HDI da diferença entre as médias não inclui o valor nulo. Foi adicionada apenas uma resposta extra no grupo EXIM (nível 1) para o conjunto de dados de sociabilidade e os parâmetros σ das duas configurações parecem semelhantes. Os valores mais prováveis para o tamanho de efeito de sociabilidade entre os grupos foi ainda maior que para aceitação, com 92,1% da distribuição acima do limite superior da ROPE ao redor do zero, sugerindo que a agente virtual da configuração EXIM é percebida como mais sociável que a da configuração EX.

A terceira linha da Tabela 5.4 apresenta as informações da inferência sobre a transparência das agentes virtuais em cada configuração de comunicação. Assim como para sociabilidade, as médias de transparência estão próximas do meio da escala ordinal e o parâmetro μ estimado é levemente maior para a agente virtual EXIM, com o valor nulo fora do 95% HDI da diferença $\mu_{EXIM} - \mu_{EX}$. Os parâmetros de escala estimados são parecidos, sendo que apenas uma resposta extra foi adicionada em cada grupo (em ambos, no nível 1). A distribuição *a posteriori* do tamanho de efeito também não é precisa o suficiente para uma decisão com a ROPE, mas dá indicativos da presença de um efeito positivo, com a agente virtual da configuração EXIM sendo considerada mais transparente.

Já para a percepção de eficiência da interação, a diferença entre as médias de cada configuração, apesar de ainda positiva, foi menor, com o valor nulo mais perto do centro da distribuição de $\mu_{EXIM} - \mu_{EX}$ do que para as outras variáveis de percepção. Estimou-se

espalhamento maior para a distribuição latente do grupo EX do que para o grupo EXIM, sendo que uma resposta extra foi adicionada a cada conjunto de dados (nível 2 no grupo EX e nível 1 no grupo EXIM). O tamanho de efeito estimado é positivo, mas com medidas de tendência central indicando efeitos bem pequenos (menores que o tamanho de efeito pequeno de 0,2), com a ROPE ao redor do valor nulo abrangendo 30,9% da distribuição e com 62,9% dela acima de 0,1. A percepção de eficiência da interação, portanto, parece não ser tão influenciada pela configuração de comunicação.

5.3 Discussão dos experimentos

A Tabela 5.5 resume os resultados das análises feitas com cada variável de interesse. Para ambas as abordagens de análise de dados, as variáveis de aceitação, sociabilidade e transparência foram as que apresentaram resultados mais favoráveis às hipóteses deste trabalho, de que a combinação de comunicações explícitas e implícitas melhora a interação entre humanos e agentes virtuais.

Table 5.5: Resumo das análises para as variáveis de interesse do trabalho. Linhas destacadas indicam variáveis que apresentaram resultados favoráveis às hipóteses deste trabalho.

	NHST frequentista	Estimação de parâmetros Bayesiana
Tempo	Hipótese nula não rejeitada.	Sem indicativos de tamanho de efeito diferente de zero.
Erros	Hipótese nula não rejeitada.	Sem indicativos de tamanho de efeito diferente de zero.
Aceitação	Valor- p próximo ao limite que levaria à rejeição da hipótese nula.	Indicativos de tamanho de efeito positivo entre pequeno e médio.
Sociabilidade	Hipótese nula rejeitada.	Indicativos de tamanho de efeito positivo entre pequeno e médio.
Transparência	Hipótese nula rejeitada.	Indicativos de tamanhos de efeito positivo entre pequeno e médio.
Percepção de eficiência	Hipótese nula não rejeitada.	Indicativos de tamanhos de efeito positivo bem pequeno.

Das 29 pessoas que passaram pelas duas configurações de comunicação (participante 4 interagiu apenas na configuração EXIM), 21 responderam que preferiram interagir com a agente virtual EXIM. Comentários sobre as agentes virtuais no questionário final foram feitos por quatro participantes⁵:

Part. 14: *“Achei bastante interessante a interação e, principalmente, após interagir com a Luna, notei o quanto fez a diferença em como me sentia em relação a tarefa*

⁵Os comentários são transcritos aqui da maneira como foram apresentados no questionário final.

o simples fato do sistema 'olhar' na minha direção."

Part. 19: *"Achei tanto a Luna como a Sofia amigáveis e sua forma de interação bem interessante."*

Part. 22: *"Adorei interagir com as assistentes virtuais!!"*

Part. 27: *"a voz é irritante"*

Em outros comentários, feitos para a experimentadora após a finalização da participação ou durante a interação e gravados em vídeo, participantes disseram que gostaram de participar, que foi "legal" e que Luna (nesse caso, a agente virtual da configuração EXIM) era "muito fofinha", enquanto outra pessoa comentou com a experimentadora antes e depois da interação que não gostava "desses(as) agentes virtuais". Diversos participantes sorriram e falaram com Luna e Sofia, tanto respondendo ou agradecendo as agentes virtuais quanto fazendo perguntas quando não entendiam as instruções. No geral, a reação das pessoas às agentes virtuais pareceu ser positiva, o que é evidenciado especialmente pelas análises dos dados de aceitação.

Aceitação, sociabilidade e transparência são variáveis mais relacionadas com a agente virtual, enquanto tempo, número de erros e percepção de eficiência da interação dependem mais diretamente da tarefa a ser executada. Na maioria dos comentários feitos no questionário final, participantes relataram dificuldades para entender os objetivos da tarefa, indicando que não consideraram as instruções claras o suficiente. Considerando as 29 participações completas, com duas interações por pessoa, o tempo limite de fase foi atingido cinco vezes na fase 1 (três delas usadas nas análises) e oito vezes na fase 2 (cinco delas usadas nas análises), com as tarefas sendo finalizadas pela agente virtual. As gravações dos experimentos mostram que as pessoas tiveram mais dificuldade principalmente com a fase 2, demorando a encontrar ou muitas vezes não vendo as imagens para contagem fixadas no ambiente. Mais de uma pessoa mencionou que o laboratório utilizado para os experimentos era muito carregado visualmente e algumas vezes foram adicionados apenas valores genéricos (como "1" para todos os objetos) nos campos de preenchimento da fase 2, indicando que a pessoa não entendeu as instruções ou não viu as imagens fixadas à frente. As medidas de objetivas podem, portanto, ter sido influenciadas por esses fatores.

Contudo, na configuração EXIM, as comunicações implícitas não eram usadas apenas para tornar Luna e Sofia mais agradáveis, sociáveis e transparentes, mas também com a intenção de facilitar a execução da tarefa, com a agente virtual olhando para os objetos que acreditava que as pessoas iriam apontar na fase 1 e usando a direção do olhar do humano para estimar seu foco de atenção e dar dicas na fase 2. Esperava-se, portanto, que a inclusão dessas comunicações implícitas amenizasse as dificuldades que as pessoas pudessem ter em relação à tarefa. De fato, algumas pessoas que parecem não ter visto as imagens de contagem na fase 2, adicionaram valores corretos aparentemente apenas

confiando nas dicas da agente virtual. O sistema de detecção do olhar do humano não é um sistema de alto desempenho, mas, na tarefa, a sua utilização era sempre complementada pelo olhar da própria agente virtual, ou seja, mesmo que a pessoa não estivesse olhando para o ponto que o sistema detectou, a agente virtual dava a dica enquanto olhava para a imagem para a qual acreditava que a pessoa estivesse olhando, então era possível inferir sobre qual objeto a dica se referia e acrescentar o valor no campo correto. Outras pessoas, porém, ou não entenderam ou não levaram em consideração as dicas da agente virtual e inseriram valores incorretos mesmo assim. Quanto ao olhar da agente virtual durante a fase 1, as gravações sugerem que talvez muitas pessoas não tenham notado essa comunicação implícita. Algumas delas perceberam que a agente virtual as seguia (participante 14 comentou no questionário final e outras pessoas pareceram “brincar” com isso durante a interação e ficaram se movimentando para ver a agente virtual as seguindo), mas outras não pareciam nem olhar para a tela da agente virtual durante a execução dos gestos de apontar, já que a senha e a evolução da tarefa eram mostradas em uma tela diferente. Além disso, as pessoas podem não ter atribuído intenção ao olhar da agente virtual. No conjunto de dados de erros, um dos valores atípicos resultou de uma situação, na configuração EXIM, em que a pessoa parece não ter entendido que a primeira cor da senha era o preto e insistia em apontar para a segunda cor da sequência (amarelo). Após a finalização da participação, essa pessoa relatou para a experimentadora que a agente virtual não reconhecia para onde ela estava apontando e que parecia que ela ficava olhando para o outro lado (no caso, para o objeto preto), ou seja, o olhar da agente virtual foi percebido mas parece não ter sido interpretado. Apesar de tudo, a tarefa não era colaborativa, isto é, não precisava necessariamente ser feita em conjunto, mesmo que a agente virtual pudesse ajudar. Acredita-se que isso possa ter influenciado na percepção de eficiência da interação, com as pessoas talvez atribuindo muito pouco ou nenhum crédito à agente virtual pela execução da tarefa.

5.4 Influência de fatores na diferença entre as configurações de comunicação

O questionário preenchido ao final do experimento coletava idade, gênero e escolaridade de participantes. Essas e outras características podem influenciar a interação entre humanos e robôs ou agentes virtuais e são necessários estudos planejados especialmente com o objetivo de medir a influência desses fatores. Neste trabalho, em que o experimento não foi projetado para essa investigação, as informações de participantes são usadas para se fazer uma análise geral sobre essas influências e tentar levantar hipóteses para estudos futuros.

A idade foi dividida em três categorias: até 25 anos, de 26 a 50 anos e acima de 50 anos. Para gênero, são consideradas as duas únicas opções usadas por participantes no

questionário, Feminino e Masculino. Os grupos de escolaridade são: de ensino fundamental incompleto a ensino médio completo (“Até ens. médio”), ensino superior incompleto e completo (“Ens. superior”) e de mestrado incompleto a doutorado completo (“Pós-graduação”). Esta análise inclui todas as 29 participações completas, ou seja, com interação em ambas as configurações de comunicação.

A Figura 5.11 mostra os conjuntos de dados de cada variável analisada no trabalho, caracterizando idade, gênero e escolaridade de participantes. A coordenada horizontal dos pontos, que têm representação definida pelo perfil de participante, é a medida feita na configuração EX e a coordenada vertical, na configuração EXIM. Por exemplo, o gráfico no canto esquerdo superior mostra que um participante do gênero masculino, com idade de 26 a 50 anos e escolaridade de mestrado incompleto a doutorado completo (indicado por um triângulo rosa preenchido e rotulado com o número 1) demorou pouco menos de dois minutos e meio na configuração EX e mais de três minutos na configuração EXIM. Para as variáveis de percepção, foi usada a média dos pontos na escala Likert de cada uma. Pontos acima da linha diagonal indicam medidas mais altas na configuração EXIM e abaixo da linha diagonal, mais altas na configuração EX.

A maioria de participantes com até 25 anos, teve desempenho pior para a configuração EX, tanto em tempo quanto em número de erros. O gráfico com as medidas de transparência mostra que a maioria das pessoas do gênero feminino acharam a agente virtual da configuração EXIM mais transparente, assim como a maioria de participantes com ensino superior incompleto ou completo. A maioria das pessoas com pelo menos mestrado incompleto avaliaram melhor a agente virtual da configuração EXIM em todas as medidas subjetivas e o contrário acontece com a maioria das pessoas com até ensino médio completo. A escolaridade, portanto, pode ser um fator importante na percepção das pessoas sobre interações com agentes virtuais.

Além de idade, gênero e escolaridade, o questionário final também perguntava a familiaridade de participantes com agentes virtuais como Luna e Sofia e a preferência entre elas (e, conseqüentemente, entre as duas configurações de comunicação). A ordem das configurações no experimento foi definida aleatoriamente para cada participante e as instruções eram dadas na primeira delas, sendo que apenas as instruções da segunda fase poderiam ser rapidamente repetidas na segunda configuração. Com isso, é natural que o desempenho na segunda repetição da tarefa seja melhor, considerando ainda que participantes relataram dificuldades em entender o que deveriam fazer. A Figura 5.12 mostra novamente as medidas em ambas as configurações, mas com os pontos caracterizados pela configuração de preferência, pela primeira configuração de interação e a familiaridade de participantes com agentes virtuais.

Como esperado, na maioria das vezes, o desempenho foi pior na primeira configuração, sendo ela EX ou EXIM. Relacionando as Figuras 5.11 e 5.12, tem-se uma possível explicação para o pior desempenho na configuração EX para a maioria de participantes de até 25 anos

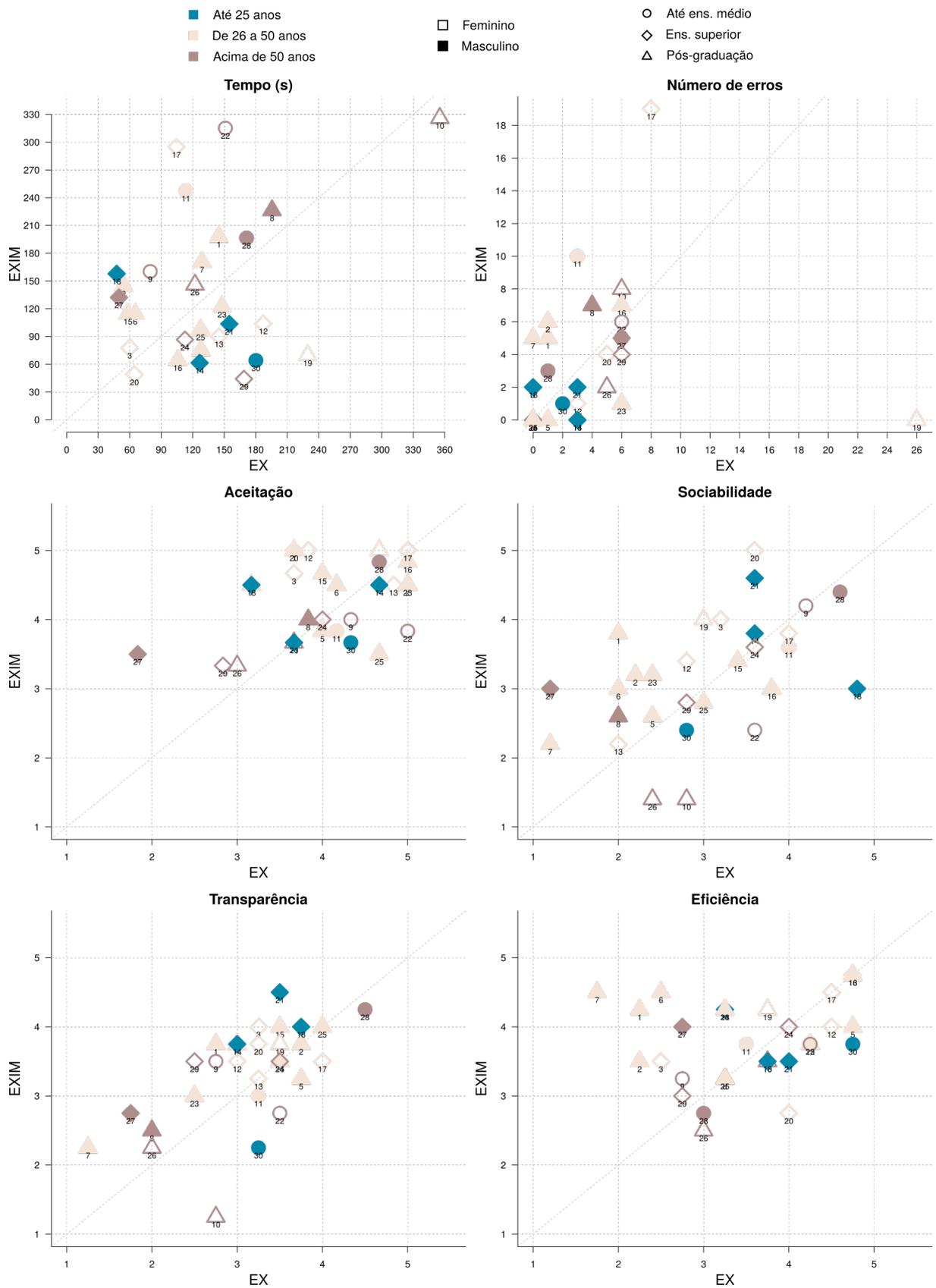


Figure 5.11: Conjuntos de dados para cada variável caracterizando perfis de participantes por idade, gênero e escolaridade. Para as variáveis de percepção, são mostradas as médias de pontos nos itens de cada escala.

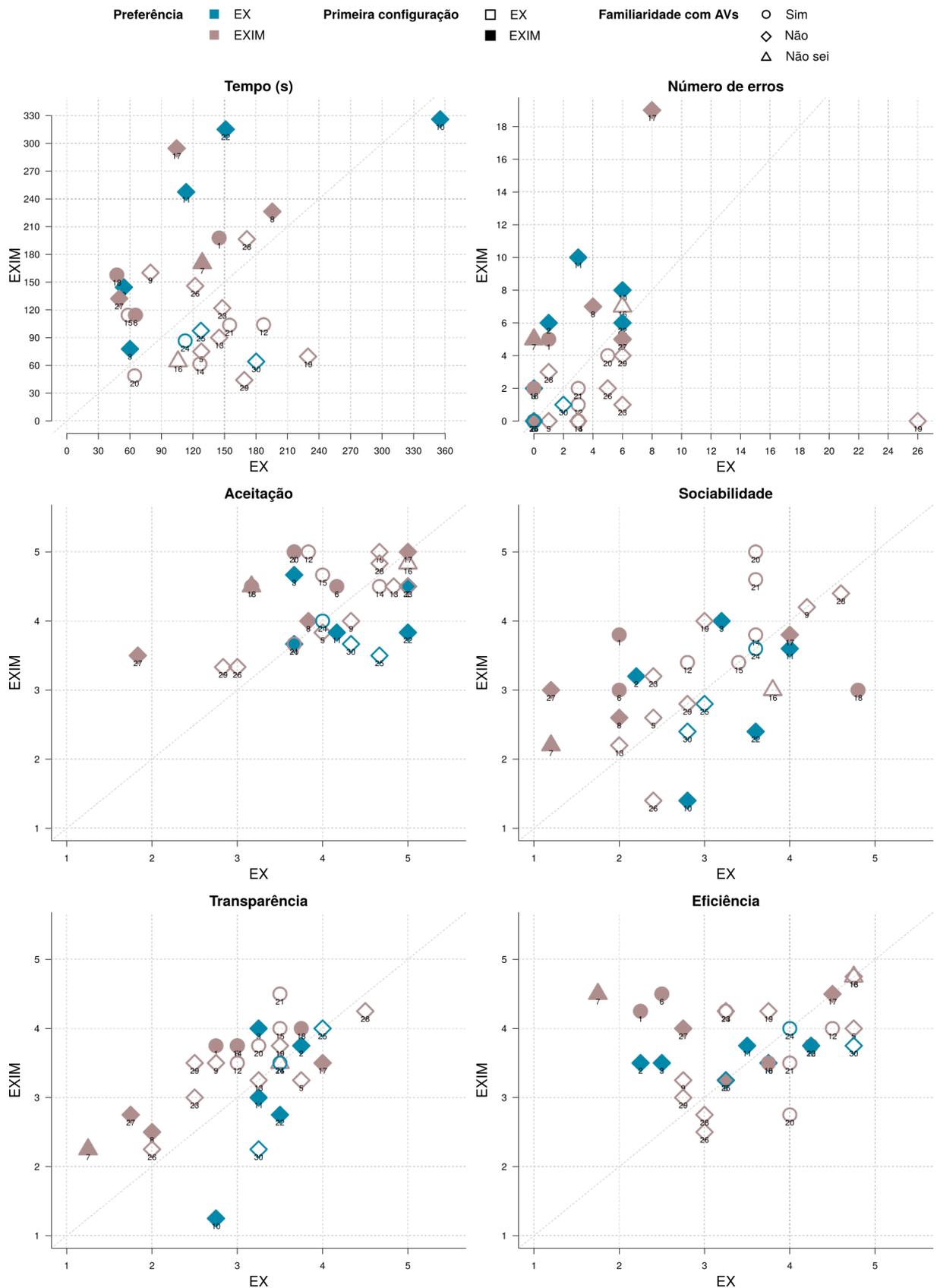


Figure 5.12: Conjuntos de dados para cada variável caracterizando perfis de participantes por configuração de preferência, primeira configuração de interação e familiaridade com agentes virtuais. Para as variáveis de percepção, são mostradas as médias de pontos nos itens de cada escala.

(participantes 14, 21 e 30): a primeira configuração nesses casos foi justamente a EX. A ordem das configurações não parece ter influenciado muito as variáveis subjetivas, já que, por exemplo, a maioria das pessoas que interagiu primeiro na configuração EXIM ainda avaliou melhor a agente virtual relacionada a ela. A preferência indicada por participantes não parece ter tido relação com o tempo de interação, entretanto, talvez tenha sido afetada pelo número de erros, visto que a maioria de participantes que preferiram a agente virtual da configuração EX erraram mais na configuração EXIM e a maioria que preferiu a configuração EXIM, errou mais na configuração EX. No geral, a preferência se relaciona com avaliações melhores sobre aceitação, sociabilidade e transparência da agente virtual, mas não tanto com a percepção de eficiência da interação, com a maioria das pessoas que preferiram a configuração EX tendo avaliado a interação com a agente virtual EXIM como mais eficiente. Por fim, a familiaridade com agentes virtuais como Luna e Sofia não parece ter tido efeito no desempenho da tarefa nem na percepção de eficiência, mas a maioria das pessoas que declararam ter alguma familiaridade, avaliaram melhor a agente virtual da configuração EXIM em aceitação, sociabilidade e transparência.

5.5 Conclusão

As análises dos resultados dos experimentos de interação indicaram que, apesar de variáveis mais relacionadas à tarefa, ou seja, tempo, número de erros e percepção de eficiência das interações, não terem sido influenciadas pela configuração de comunicação, a percepção das agentes virtuais em medidas de aceitação, sociabilidade e transparência é melhor na configuração EXIM, em que humano e agente virtual combinam comunicações explícitas e implícitas. Os tamanhos de efeito estimados com a análise Bayesiana para aceitação, sociabilidade e transparência foram entre pequenos e médios em favor da configuração EXIM, dando indícios favoráveis às hipóteses de trabalho relacionadas a essas variáveis. Acredita-se que características específicas da tarefa proposta podem ter sido mais relevantes para o desempenho e a percepção de eficiência do que o tipo de comunicação dos agentes.

Finalmente, os dados de todas as participações completas foram usados para análises gerais sobre a influência do perfil de participante (idade, gênero, escolaridade e familiaridade com agentes virtuais), da ordem das configurações e da agente virtual preferida nas diferenças entre as configurações de comunicação. As observações feitas podem se tornar hipóteses de trabalho de estudos para a investigação da influência desses fatores nas interações entre humanos e agentes virtuais.

6

CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho investigou os efeitos da combinação de comunicações explícitas e implícitas em medidas de desempenho e na percepção das pessoas em uma interação com agentes virtuais. Para isso, uma infraestrutura de comunicação humano-robô com modalidades de comunicação explícita e implícita foi implementada, com soluções simples e de baixo custo. Entre as tecnologias de comunicação, estão sistemas de reconhecimento e interpretação de gestos de apontar e direção do olhar do humano. Para esses sistemas, foram propostas análises geométricas, usando pontos da cadeia cinemática do humano, atributos faciais na imagem do rosto e primitivas geométricas. Além disso, foi desenvolvida uma agente virtual para desempenhar o papel de robô nas interações. Imagens com diferentes expressões faciais foram criadas e a direção do olhar da agente virtual é simulada a partir de geometria projetiva.

Para testar hipóteses sugeridas pela literatura, de que a combinação de comunicações explícitas e implícitas melhoram medidas objetivas e subjetivas da interação, um experimento de interação entre humano e agentes virtuais foi proposto. O protocolo experimental foi definido para que duas configurações de comunicação fossem comparadas: uma apenas com comunicações explícitas de ambos os agentes e outra com comunicações explícitas e implícitas. O conjunto de dados de 26 participações no experimento foram analisados usando testes de significância de hipótese nula e estimação de parâmetros Bayesiana. Os resultados sugerem que combinar comunicações explícitas e implícitas melhora a aceitação, a sociabilidade e a transparência de agentes virtuais.

As medidas mais relacionadas à tarefa (tempo, número de erros e percepção de

eficiência da interação) não parecem ter sido influenciadas pelo tipo de comunicação, o que possivelmente pode ser atribuído a características específicas da interação proposta. No trabalho de Breazeal et al. (2005), a combinação de comunicações explícitas e implícitas do robô melhorou medidas objetivas da interação, em um cenário em que participantes deveriam guiar o robô na execução de uma tarefa simples de pressionar botões. Já na interação proposta neste trabalho, a agente virtual dava instruções durante a interação e podia auxiliar na realização das tarefas, especialmente na configuração de comunicações combinadas, mas a colaboração entre participante e agente virtual não era necessária. Além disso, de acordo com relatos de participantes, as instruções das tarefas podem não ter sido claras o suficiente e a análise dos vídeos sugere que as comunicações implícitas da agente virtual na fase 1 da interação podem não ter sido percebidas por algumas pessoas. Esses fatores podem ter prejudicado as medidas objetivas das diferenças entre as configurações de comunicação e a medida de percepção de eficiência da interação, já que é possível que participantes não tenham atribuído nenhum crédito à agente virtual pelo desempenho nas tarefas.

Restrições impostas pela pandemia de COVID-19 limitaram o número e a diversidade de participantes, por isso o experimento realizado é considerado um piloto. Apesar disso, foram observados indícios a favor de três das seis hipóteses de trabalho e a análise dos experimentos indica direções para possíveis melhorias no protocolo experimental. Uma interação mais colaborativa pode permitir que se investigue melhor os efeitos do tipo de comunicação em variáveis mais relacionadas à tarefa e a inserção de uma etapa de treino pode ajudar a reduzir efeitos do período de aprendizagem da tarefa.

Para as medidas subjetivas, foram usados questionários com escalas Likert para cada uma das variáveis. Trabalhos como os de Venkatesh et al. (2003), Bartneck et al. (2009) e Heerink et al. (2010) buscam propor modelos e instrumentos de medição de percepções das pessoas sobre tecnologias e robôs, porém ainda não há uma unificação nos trabalhos de IHR quanto a isso. Quando questionários com escalas Likert são usados, é comum que seja reportada uma medida chamada de *alfa de Cronbach* (Cronbach, 1951, apud Dunn et al., 2014), como uma avaliação da confiabilidade e consistência interna da escala de medição (alguns exemplos são Hinds et al. (2004), Mutlu et al. (2009), Crossman et al. (2018) e Tatsukawa et al. (2019)), entretanto há discussões sobre a adequação dessa medida (Dunn et al., 2014; Peters, 2014). Além de orientações para a definição de itens para as escalas, Likert (1932) também discute em seu trabalho checagens objetivas sobre a consistência interna e adequação dos itens. Esse tipo de análise e discussão não foi feita neste trabalho e a adequação das escalas usadas para medir as variáveis subjetivas precisa ser investigada em trabalhos futuros.

Apesar das limitações do método NHST frequentista estarem sendo cada vez mais discutidas, ele segue sendo amplamente aplicado em diferentes áreas, incluindo a de interação humano-robô (Baxter et al., 2016). O trabalho de Schrum et al. (2020), por

exemplo, faz uma revisão das práticas estatísticas de análises de dados de escalas Likert em estudos de IHR e faz recomendações sobre o assunto apenas dentro da perspectiva de testes de hipóteses frequentistas. Belpaeme (2020) recomenda que mais detalhes sobre os experimentos e os conjuntos de dados do estudo sejam apresentados nos trabalhos de IHR, incluindo estatísticas descritivas, como tamanho, média e desvio padrão da amostra, e perfis de participantes, e que tamanhos de efeito sejam sempre reportados junto com valores- p , para que a relevância do efeito possa ser discutida. Baxter et al. (2016) sugerem uma mudança de perspectiva, com foco em metodologias experimentais que possam aumentar incrementalmente o conhecimento sobre fenômenos de interesse, ou seja, uma perspectiva Bayesiana. No contexto de interação humano-computador, Kaptein & Robertson (2012) apresentam o teste de hipóteses com o fator de Bayes como alternativa aos métodos NHST com valores- p .

A utilização da abordagem Bayesiana para análise de dados, em especial a estimação de parâmetros com incerteza, tem sido discutida em outras áreas, como estatística e psicologia, a partir de trabalhos como os de Kruschke (2013), Kruschke & Liddell (2018), Wagenmakers et al. (2018), Kelter (2020) and Kelter (2021), mas não tanto nos trabalhos de interação humano-robô. A multidisciplinaridade da área de IHR conecta e aproxima diversas áreas de pesquisa, mas também cria a necessidade de se observar aspectos relacionados a todas elas, nesse caso, mais especificamente, sobre análises estatísticas. Livros como o de Kruschke (2015), que apresenta exemplos detalhados e disponibiliza códigos implementados e uma variedade de funções em R, e programas como JASP¹, de código livre e com interface gráfica para a realização de análises estatísticas frequentistas e Bayesianas (Wagenmakers et al., 2018), facilitam a transição dos métodos NHST comumente utilizados para métodos de análise Bayesiana. A estimação de parâmetros Bayesiana apresenta vantagens como a possibilidade de discussões mais completas e não apenas rejeição ou não de hipóteses, a obtenção de informação de probabilidade sobre cada valor de parâmetro e, além disso, ainda possibilita a incorporação de conhecimentos prévios em novos estudos. Os resultados obtidos com a inferência Bayesiana neste trabalho podem ser usados na definição de distribuições *a priori* para experimentos futuros, desde que as incertezas, especificidades e limitações do trabalho sejam levadas em consideração.

Finalmente, as análises das influências de fatores, como idade, gênero, escolaridade e familiaridade com agentes virtuais, nas medidas objetivas e subjetivas das diferentes configurações de comunicação foi feita apenas com a observação dos dados. Uma vantagem da abordagem Bayesiana é a possibilidade de se fazer diferentes análises com um mesmo conjunto de dados, sem que seja necessário ajustar critérios de decisão para controlar a taxa de falsos positivos, como é feito no método NHST, já que métodos Bayesianos não dependem de intenções de testes (ver Seção 4.4). Os dados dos experimentos podem ser divididos de acordo com as categorias de cada fator analisado e a estimação Bayesiana

¹<https://jasp-stats.org/>

dos parâmetros aplicada ao conjunto de cada grupo. As análises podem ser feitas em pares, comparando, por exemplo, as duas categorias de gênero considerando cada uma das configurações de comunicação separadamente ou então as diferenças entre as configurações de comunicação para as participantes do gênero feminino e depois para os do gênero masculino. Essas e outras comparações podem ser feitas em trabalhos futuros a partir do mesmo conjunto de dados. Conclusões fortes ainda não poderiam ser feitas, já que os tamanhos amostrais seriam ainda menores para cada grupo, mas as distribuições *a posteriori* talvez sugiram outros padrões e deem origem a novas hipóteses de trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abdi, J.; Al-Hindawi, A.; Ng, T.; & Vizcaychipi, M. P. (2018). “Scoping review on the use of socially assistive robot technology in elderly care”. In: *BMJ Open* 8.2, e018815. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2017-018815. URL: <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-018815>.
- Adorno, B. V. (2017). “Robot Kinematic Modeling and Control Based on Dual Quaternion Algebra - Part I: Fundamentals”. In: URL: <https://hal.archives-ouvertes.fr/hal-01478225>.
- Adorno, B. V.; & Marinho, M. M. (2021). “DQ Robotics: a Library for Robot Modeling and Control”. In: *IEEE Robotics & Automation Magazine* 28.3, pp. 102–116. ISSN: 1070-9932. DOI: 10.1109/MRA.2020.2997920. arXiv: 1910.11612. URL: <https://ieeexplore.ieee.org/document/9136790> / <http://arxiv.org/abs/1910.11612> <http://dx.doi.org/10.1109/MRA.2020.2997920>.
- Agrigoroaie, R.; Ciocirlan, S.-D.; & Tapus, A. (2020). “In the Wild HRI Scenario: Influence of Regulatory Focus Theory”. In: *Frontiers in Robotics and AI* 7.April, pp. 1–11. ISSN: 2296-9144. DOI: 10.3389/frobt.2020.00058. URL: <https://www.frontiersin.org/article/10.3389/frobt.2020.00058/full>.
- Ajoudani, A.; Zanchettin, A. M.; Ivaldi, S.; Albu-Schäffer, A.; Kosuge, K.; & Khatib, O. (2018). “Progress and prospects of the human-robot collaboration”. In: *Autonomous Robots* 42.5, pp. 957–975. ISSN: 0929-5593. DOI: 10.1007/s10514-017-9677-2. URL: <http://link.springer.com/10.1007/s10514-017-9677-2>.
- Asfour, T.; Regenstein, K.; Azad, P.; Schroder, J.; Bierbaum, A.; Vahrenkamp, N.; & Dillmann, R. (2006). “ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control”. In: *2006 6th IEEE-RAS International Conference on Humanoid Robots*. IEEE, pp. 169–175. ISBN: 1-4244-0199-2. DOI: 10.1109/ICHR.2006.321380. URL: <http://ieeexplore.ieee.org/document/4115597/>.

- Azari, B.; Lim, A.; & Vaughan, R. (2019). “Commodifying Pointing in HRI: Simple and Fast Pointing Gesture Detection from RGB-D Images”. In: *2019 16th Conference on Computer and Robot Vision (CRV)*. IEEE, pp. 174–180. ISBN: 978-1-7281-1838-3. DOI: 10.1109/CRV.2019.00031. arXiv: 1902.02636. URL: <https://ieeexplore.ieee.org/document/8781628/>.
- Baker, M. (2016). “Statisticians issue warning on P values”. In: *Nature* 351, p. 151.
- Bartneck, C.; Kulić, D.; Croft, E.; & Zoghbi, S. (2009). “Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots”. In: *International Journal of Social Robotics* 1.1, pp. 71–81. ISSN: 1875-4791. DOI: 10.1007/s12369-008-0001-3. URL: <http://link.springer.com/10.1007/s12369-008-0001-3>.
- Bauer, A.; Wollherr, D.; & Buss, M. (2008). “Human-Robot Collaboration: a Survey”. In: *International Journal of Humanoid Robotics* 05.01, pp. 47–66. ISSN: 0219-8436. DOI: 10.1142/S0219843608001303. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0219843608001303>.
- Baxter, P.; Kennedy, J.; Senft, E.; Lemaignan, S.; & Belpaeme, T. (2016). “From characterising three years of HRI to methodology and reporting recommendations”. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Vol. 2016-April. IEEE, pp. 391–398. ISBN: 978-1-4673-8370-7. DOI: 10.1109/HRI.2016.7451777. URL: <http://ieeexplore.ieee.org/document/7451777/>.
- Beckerle, P.; Salvietti, G.; Unal, R.; Prattichizzo, D.; Rossi, S.; Castellini, C.; Hirche, S.; Endo, S.; Amor, H. B.; Ciocarlie, M.; Mastrogiovanni, F.; Argall, B. D.; & Bianchi, M. (2017). “A Human-Robot Interaction Perspective on Assistive and Rehabilitation Robotics”. In: *Frontiers in Neurobotics* 11.MAY. ISSN: 1662-5218. DOI: 10.3389/fnbot.2017.00024. URL: <http://journal.frontiersin.org/article/10.3389/fnbot.2017.00024/full>.
- Belkaid, M.; Kompatsiari, K.; De Tommaso, D.; Zablith, I.; & Wykowska, A. (2021). “Mutual gaze with a robot affects human neural activity and delays decision-making processes”. In: *Science Robotics* 6.58. ISSN: 2470-9476. DOI: 10.1126/scirobotics.abc5044. URL: <https://www.science.org/doi/10.1126/scirobotics.abc5044>.
- Belpaeme, T. (2020). “Advice to New Human-Robot Interaction Researchers”. In: *Human-Robot Interaction. Springer Series on Bio- and Neurosystems*, pp. 355–369. DOI: 10.

- 1007/978-3-030-42307-0_14. URL: http://link.springer.com/10.1007/978-3-030-42307-0_14.
- Biswas, K. K.; & Basu, S. K. (2011). “Gesture recognition using Microsoft Kinect”. In: *The 5th International Conference on Automation, Robotics and Applications*. May. IEEE, pp. 100–103. ISBN: 978-1-4577-0330-0. DOI: 10.1109/ICARA.2011.6144864. URL: <http://ieeexplore.ieee.org/document/6144864/>.
- Breazeal, C. (2003). “Toward sociable robots”. In: *Robotics and Autonomous Systems* 42.3-4, pp. 167–175. ISSN: 09218890. DOI: 10.1016/S0921-8890(02)00373-1. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0921889002003731>.
- Breazeal, C.; Kidd, C.; Thomaz, A.; Hoffman, G.; & Berlin, M. (2005). “Effects of nonverbal communication on efficiency and robustness in human-robot teamwork”. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. ISBN: 0-7803-8912-3. DOI: 10.1109/IRROS.2005.1545011. URL: <http://ieeexplore.ieee.org/document/1545011/>.
- Bruce, A.; Nourbakhsh, I.; & Simmons, R. (2002). “The role of expressiveness and attention in human-robot interaction”. In: *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*. Vol. 4. IEEE, pp. 4138–4142. ISBN: 0-7803-7272-7. DOI: 10.1109/ROBOT.2002.1014396. URL: <http://ieeexplore.ieee.org/document/1014396/>.
- Campos, A. C. A.; & Adorno, B. V. (2019). “Human Gesture Recognition using Dual Quaternion Algebra”. In: *Workshop on Applications of Dual Quaternion Algebra to Robotics (ICAR-2019)*, pp. 1–2. DOI: 10.5281/zenodo.3566656. URL: <https://zenodo.org/record/3566656>.
- (2020). “Development of Human-Robot Communication Technologies for Future Interaction Experiments”. In: *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*. IEEE, pp. 1–6. ISBN: 978-0-7381-1153-7. DOI: 10.1109/LARS/SBR/WRE51543.2020.9306965. URL: <https://ieeexplore.ieee.org/document/9306965/>.
- Carifio, J.; & Perla, R. (2008). “Resolving the 50-year debate around using and misusing Likert scales”. In: *Medical Education* 42.12, pp. 1150–1152. ISSN: 03080110. DOI: 10.1111/j.1365-2923.2008.03172.x. URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.2008.03172.x>.

- Cazzato, D.; Leo, M.; Distante, C.; & Voos, H. (2020). “When I Look into Your Eyes: A Survey on Computer Vision Contributions for Human Gaze Estimation and Tracking”. In: *Sensors* 20.13, p. 3739. ISSN: 1424-8220. DOI: 10.3390/s20133739. URL: <https://www.mdpi.com/1424-8220/20/13/3739>.
- Che, Y.; Okamura, A. M.; & Sadigh, D. (2020). “Efficient and Trustworthy Social Navigation via Explicit and Implicit Robot-Human Communication”. In: *IEEE Transactions on Robotics*, pp. 1–16. ISSN: 1552-3098. DOI: 10.1109/TR0.2020.2964824. arXiv: 1810.11556. URL: <https://ieeexplore.ieee.org/document/8967120/>.
- Clifford (1871). “Preliminary Sketch of Biquaternions”. In: *Proceedings of the London Mathematical Society* s1-4.1, pp. 381–395. ISSN: 00246115. DOI: 10.1112/plms/s1-4.1.381. URL: <http://doi.wiley.com/10.1112/plms/s1-4.1.381>.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second. Lawrence Erlbaum Associates. ISBN: 9781134742707.
- Cronbach, L. J. (1951). “Coefficient alpha and the internal structure of tests”. In: *Psychometrika* 16.3, pp. 297–334. ISSN: 0033-3123. DOI: 10.1007/BF02310555. URL: <http://link.springer.com/10.1007/BF02310555>.
- Crossman, M. K.; Kazdin, A. E.; & Kitt, E. R. (2018). “The influence of a socially assistive robot on mood, anxiety, and arousal in children”. In: *Professional Psychology: Research and Practice* 49.1, pp. 48–56. ISSN: 1939-1323. DOI: 10.1037/pro0000177. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/pro0000177>.
- Crowe, E.; & Higgins, E. (1997). “Regulatory Focus and Strategic Inclinations: Promotion and Prevention in Decision-Making”. In: *Organizational Behavior and Human Decision Processes* 69.2, pp. 117–132. ISSN: 07495978. DOI: 10.1006/obhd.1996.2675. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0749597896926758>.
- Daniilidis, K. (1999). “Hand-Eye Calibration Using Dual Quaternions”. In: *The International Journal of Robotics Research* 18.3, pp. 286–298. ISSN: 0278-3649. DOI: 10.1177/02783649922066213. URL: <http://journals.sagepub.com/doi/10.1177/02783649922066213>.
- Dunn, T. J.; Baguley, T.; & Brunsdon, V. (2014). “From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation”. In: *British Journal*

- of Psychology* 105.3, pp. 399–412. ISSN: 00071269. DOI: 10.1111/bjop.12046. URL: <http://doi.wiley.com/10.1111/bjop.12046>.
- Feil-Seifer, D.; & Mataric, M. J. (2005). “Defining Socially Assistive Robotics”. In: *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005*. IEEE, pp. 465–468. ISBN: 0-7803-9003-2. DOI: 10.1109/ICORR.2005.1501143. URL: <http://ieeexplore.ieee.org/document/1501143/>.
- Fiore, S. M.; Wiltshire, T. J.; Lobato, E. J. C.; Jentsch, F. G.; Huang, W. H.; & Axelrod, B. (2013). “Toward understanding social cues and signals in human-robot interaction: effects of robot gaze and proxemic behavior”. In: *Frontiers in Psychology* 4.NOV, pp. 1–15. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00859. URL: <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00859/abstract>.
- Fong, T.; Nourbakhsh, I.; & Dautenhahn, K. (2003). “A survey of socially interactive robots”. In: *Robotics and Autonomous Systems* 42, pp. 143–166. ISSN: 09218890. DOI: 10.1016/S0921-8890(02)00372-X. URL: <https://linkinghub.elsevier.com/retrieve/pii/S092188900200372X>.
- Frith, C. (2009). “Role of facial expressions in social interactions”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pp. 3453–3458. ISSN: 0962-8436. DOI: 10.1098/rstb.2009.0142. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0142>.
- Glover, S. (2018). “Likelihood Ratios: A Tutorial on Applications to Research in Psychology”. In: *metaRxiv*. DOI: 10.31222/osf.io/g3j2k. URL: <https://osf.io/preprints/bitss/g3j2k/>.
- Gockley, R.; Bruce, A.; Forlizzi, J.; Michalowski, M.; Mundell, A.; Rosenthal, S.; Sellner, B.; Simmons, R.; Snipes, K.; Schultz, A.; & Jue Wang (2005). “Designing robots for long-term social interaction”. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 1338–1343. ISBN: 0-7803-8912-3. DOI: 10.1109/IROS.2005.1545303. URL: <http://ieeexplore.ieee.org/document/1545303/>.
- Gockley, R.; Forlizzi, J.; & Simmons, R. (2006). “Interactions with a moody robot”. In: *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*. New York, New York, USA: ACM Press, pp. 186–193. ISBN: 1595932941. DOI: 10.1145/1121241.1121274. URL: <http://portal.acm.org/citation.cfm?doid=1121241.1121274>.

- Gombolay, M.; Bair, A.; Huang, C.; & Shah, J. (2017). “Computational design of mixed-initiative human - robot teaming that considers human factors: situational awareness, workload, and workflow preferences”. In: *The International Journal of Robotics Research* 36.5-7, pp. 597–617. ISSN: 0278-3649. DOI: 10.1177/0278364916688255. URL: <http://journals.sagepub.com/doi/10.1177/0278364916688255>.
- Gombolay, M. C.; Gutierrez, R. A.; Clarke, S. G.; Sturla, G. F.; & Shah, J. A. (2015). “Decision-making authority, team efficiency and human worker satisfaction in mixed human - robot teams”. In: *Autonomous Robots* 39.3, pp. 293–312. ISSN: 0929-5593. DOI: 10.1007/s10514-015-9457-9. URL: <http://link.springer.com/10.1007/s10514-015-9457-9>.
- Goodrich, M. A.; & Schultz, A. C. (2007). “Human-Robot Interaction: A Survey”. In: *Foundations and Trends® in Human-Computer Interaction* 1.3, pp. 203–275. ISSN: 1551-3955. DOI: 10.1561/1100000005. URL: <http://www.nowpublishers.com/article/Details/HCI-005>.
- Gu, Y.; Do, H.; Ou, Y.; & Sheng, W. (2012). “Human gesture recognition through a Kinect sensor”. In: *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, pp. 1379–1384. ISBN: 978-1-4673-2127-3. DOI: 10.1109/ROBIO.2012.6491161. URL: <http://ieeexplore.ieee.org/document/6491161/>.
- Guznov, S.; Lyons, J.; Pfahler, M.; Heironimus, A.; Woolley, M.; Friedman, J.; & Neimeier, A. (2019). “Robot Transparency and Team Orientation Effects on Human-Robot Teaming”. In: *International Journal of Human-Computer Interaction*, pp. 650–660. ISSN: 1044-7318. DOI: 10.1080/10447318.2019.1676519. URL: <https://doi.org/10.1080/10447318.2019.1676519><https://www.tandfonline.com/doi/full/10.1080/10447318.2019.1676519>.
- Hamilton, W. R. (1844). “On Quaternions, Or On a New System Of Imaginaries In Algebra. Edited by David R. Wilkins. 18 vols”. In: *Philosophical Magazine. 18 vols. Ed. by David R. Wilkins*.
- Harpe, S. E. (2015). “How to analyze Likert and other rating scale data”. In: *Currents in Pharmacy Teaching and Learning* 7.6, pp. 836–850. ISSN: 18771297. DOI: 10.1016/j.cptl.2015.08.001. URL: <http://dx.doi.org/10.1016/j.cptl.2015.08.001><https://linkinghub.elsevier.com/retrieve/pii/S1877129715200196>.

- Heerink, M.; Kröse, B.; Evers, V.; & Wielinga, B. (2010). “Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model”. In: *International Journal of Social Robotics* 2.4, pp. 361–375. ISSN: 1875-4791. DOI: 10.1007/s12369-010-0068-5. URL: <http://link.springer.com/10.1007/s12369-010-0068-5>.
- Higgins, E. T. (1997). “Beyond pleasure and pain”. In: *American Psychologist* 52.12, pp. 1280–1300. ISSN: 1935-990X. DOI: 10.1037/0003-066X.52.12.1280. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.52.12.1280>.
- Hinds, P. J.; Roberts, T. L.; & Jones, H. (2004). “Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task”. In: *Human-Computer Interaction* 19.1-2, pp. 151–181. DOI: 10.1080/07370024.2004.9667343. URL: <https://www.tandfonline.com/doi/abs/10.1080/07370024.2004.9667343>.
- Hirano, T.; Shiomi, M.; Iio, T.; Kimoto, M.; Tanev, I.; Shimohara, K.; & Hagita, N. (2018). “How Do Communication Cues Change Impressions of Human-Robot Touch Interaction?”. In: *International Journal of Social Robotics* 10, pp. 21–31. ISSN: 1875-4791. DOI: 10.1007/s12369-017-0425-8. URL: <http://link.springer.com/10.1007/s12369-017-0425-8>.
- Hollander, M.; Wolfe, D. A.; & Chicken, E. (2014). *Nonparametric Statistical Methods*. Third. John Wiley & Sons, Inc. ISBN: 978-0-470-38737-5.
- Huang, C. M.; & Mutlu, B. (2016). “Anticipatory robot control for efficient human-robot collaboration”. In: *ACM/IEEE International Conference on Human-Robot Interaction*. Vol. 2016-April. Section V. IEEE, pp. 83–90. ISBN: 9781467383707. DOI: 10.1109/HRI.2016.7451737. URL: <http://ieeexplore.ieee.org/document/7451737/>.
- Huang, Y.; Chen, F.; Lv, S.; & Wang, X. (2019). “Facial Expression Recognition: A Survey”. In: *Symmetry* 11.10. ISSN: 2073-8994. DOI: 10.3390/sym11101189. URL: <https://www.mdpi.com/2073-8994/11/10/1189>.
- Iwasaki, M.; Zhou, J.; Ikeda, M.; Koike, Y.; Onishi, Y.; Kawamura, T.; & Nakanishi, H. (2019). ““That Robot Stared Back at Me!”: Demonstrating Perceptual Ability Is Key to Successful Human - Robot Interactions”. In: *Frontiers in Robotics and AI* 6.September, pp. 1–12. ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00085. URL: <https://www.frontiersin.org/article/10.3389/frobt.2019.00085/full>.

- Kalegina, A.; Schroeder, G.; Allchin, A.; Berlin, K.; & Cakmak, M. (2018). “Characterizing the Design Space of Rendered Robot Faces”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*. February. New York, New York, USA: ACM Press, pp. 96–104. ISBN: 9781450349536. DOI: 10.1145/3171221.3171286. URL: <http://dl.acm.org/citation.cfm?doid=3171221.3171286>.
- Kaptein, M.; & Robertson, J. (2012). “Rethinking statistical analysis methods for CHI”. In: *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1105–1113. DOI: 10.1145/2207676.2208557.
- Kelter, R. (2020). “Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP”. In: *BMC Medical Research Methodology* 20.1. ISSN: 1471-2288. DOI: 10.1186/s12874-020-00980-6. URL: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-00980-6>.
- (2021). “Bayesian and frequentist testing for differences between two groups with parametric and nonparametric two-sample tests”. In: *WIREs Computational Statistics* 13.6, pp. 1–29. ISSN: 1939-5108. DOI: 10.1002/wics.1523. URL: <https://onlinelibrary.wiley.com/doi/10.1002/wics.1523>.
- Kim, H.; Lee, S.; Lee, D.; Choi, S.; Ju, J.; & Myung, H. (2015). “Real-Time Human Pose Estimation and Gesture Recognition from Depth Images Using Superpixels and SVM Classifier”. In: *Sensors* 15.6, pp. 12410–12427. ISSN: 1424-8220. DOI: 10.3390/s150612410. URL: <http://www.mdpi.com/1424-8220/15/6/12410>.
- Kloke, J.; & Mckean, J. W. (2015). *The R Series Statistics Nonparametric Statistical Methods Using R*. CRC Press - Taylor & Francis Group.
- Knepper, R. A.; Mavrogiannis, C. I.; Proft, J.; & Liang, C. (2017). “Implicit Communication in a Joint Action”. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, pp. 283–292. ISBN: 9781450343367. DOI: 10.1145/2909824.3020226. URL: <https://dl.acm.org/doi/10.1145/2909824.3020226>.
- Kruschke, J. K. (2013). “Bayesian estimation supersedes the t test.” In: *Journal of Experimental Psychology: General* 142.2, pp. 573–603. ISSN: 1939-2222. DOI: 10.1037/a0029146. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0029146>.

- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. 2nd. Burlington, MA: Academic Press / Elsevier. ISBN: 9780124058880.
- Kruschke, J. K.; & Liddell, T. M. (2018). “The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective”. In: *Psychonomic Bulletin & Review* 25.1, pp. 178–206. ISSN: 1069-9384. DOI: 10.3758/s13423-016-1221-4. URL: <http://link.springer.com/10.3758/s13423-016-1221-4>.
- Lana, E. P.; Adorno, B. V.; & Tierra-Criollo, C. J. (2013). “Assistance Task Using a Manipulator Robot and User Kinematics Feedback”. In: *Simpósio Brasileiro de Automação Inteligente (SBAI)* 52.1, pp. 1–6.
- Lazzeri, N.; Mazzei, D.; & De Rossi, D. (2014). “Development and Testing of a Multimodal Acquisition Platform for Human-Robot Interaction Affective Studies”. In: *Journal of Human-Robot Interaction* 3.2. ISSN: 2163-0364. DOI: 10.5898/JHRI.3.2.Lazzeri. URL: <http://dl.acm.org/citation.cfm?id=3109830>.
- Lenz, A.; Skachek, S.; Hamann, K.; Steinwender, J.; Pipe, A. G.; & Melhuish, C. (2010). “The BERT2 infrastructure: An integrated system for the study of human-robot interaction”. In: *2010 10th IEEE-RAS International Conference on Humanoid Robots*. IEEE, pp. 346–351. ISBN: 978-1-4244-8688-5. DOI: 10.1109/ICHR.2010.5686319. URL: <http://ieeexplore.ieee.org/document/5686319/>.
- Liddell, T. M.; & Kruschke, J. K. (2018). “Analyzing ordinal data with metric models: What could possibly go wrong?” In: *Journal of Experimental Social Psychology* 79. August, pp. 328–348. ISSN: 00221031. DOI: 10.1016/j.jesp.2018.08.009. URL: <https://doi.org/10.1016/j.jesp.2018.08.009><https://linkinghub.elsevier.com/retrieve/pii/S0022103117307746>.
- Likert, R. (1932). “A Technique for the Measurement of Attitudes”. In: *Archives of Psychology*.
- Liu, H.; & Wang, L. (2018). “Gesture recognition for human-robot collaboration: A review”. In: *International Journal of Industrial Ergonomics* 68, pp. 355–367. ISSN: 18728219. DOI: 10.1016/j.ergon.2017.02.004. URL: <https://doi.org/10.1016/j.ergon.2017.02.004>.

- Ljungblad, S.; Kotrbova, J.; Jacobsson, M.; Cramer, H.; & Niechwiadowicz, K. (2012). “Hospital Robot at Work: Something Alien or an Intelligent Colleague?” In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*. New York, New York, USA: ACM Press, pp. 177–186. ISBN: 9781450310864. DOI: 10.1145/2145204.2145233. URL: <http://dl.acm.org/citation.cfm?doid=2145204.2145233>.
- Lun, R.; & Zhao, W. (2015). “A Survey of Applications and Human Motion Recognition with Microsoft Kinect”. In: *International Journal of Pattern Recognition and Artificial Intelligence* 29.05, p. 1555008. ISSN: 0218-0014. DOI: 10.1142/S0218001415550083. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0218001415550083>.
- Marinho, M. M.; Adorno, B. V.; Harada, K.; & Mitsuishi, M. (2018). “Active Constraints Using Vector Field Inequalities for Surgical Robots”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5364–5371. ISBN: 978-1-5386-3081-5. DOI: 10.1109/ICRA.2018.8461105. arXiv: 1804.11270. URL: <http://arxiv.org/abs/1804.11270><https://ieeexplore.ieee.org/document/8461105/>.
- Maris, A. van; Zook, N.; Caleb-Solly, P.; Studley, M.; Winfield, A.; & Dogramadzi, S. (2020). “Designing Ethical Social Robots - A Longitudinal Field Study With Older Adults”. In: *Frontiers in Robotics and AI* 7.January. ISSN: 2296-9144. DOI: 10.3389/frobt.2020.00001. URL: <https://www.frontiersin.org/article/10.3389/frobt.2020.00001/full>.
- Matthews, R.; Wasserstein, R.; & Spiegelhalter, D. (2017). “The ASA’s p -value statement, one year on”. In: *Significance* 14.2, pp. 38–41. ISSN: 17409705. DOI: 10.1111/j.1740-9713.2017.01021.x. URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01021.x>.
- Mavridis, N. (2015). “A review of verbal and non-verbal human-robot interactive communication”. In: *Robotics and Autonomous Systems* 63, pp. 22–35. ISSN: 09218890. DOI: 10.1016/j.robot.2014.09.031. URL: <http://dx.doi.org/10.1016/j.robot.2014.09.031><https://linkinghub.elsevier.com/retrieve/pii/S0921889014002164>.
- Meek, G. E.; Ozgur, C.; & Dunning, K. (2007). “Comparison of the t vs. Wilcoxon Signed-Rank Test for Likert Scale Data and Small Samples”. In: *Journal of Modern Applied Statistical Methods* 6.1, pp. 91–106. ISSN: 1538-9472. DOI: 10.22237/jmasm/1177992540. URL: <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/10>.

- Meghdari, A.; Shariati, A.; Alemi, M.; Nobaveh, A. A.; Khamooshi, M.; & Mozafari, B. (2018). "Design Performance Characteristics of a Social Robot Companion "Arash" for Pediatric Hospitals". In: *International Journal of Humanoid Robotics* 15.05, p. 1850019. ISSN: 0219-8436. DOI: 10.1142/S0219843618500196. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0219843618500196>.
- Montgomery, D. C.; & Runger, G. C. (2011). *Applied Statistics and Probability for Engineers, Fifth Edition*. 5th. John Wiley & Sons, Inc. ISBN: 9780470053041. URL: www.wiley.com/college/montgomery.
- Mower, E.; Feil-Seifer, D. J.; Mataric, M. J.; & Narayanan, S. (2007). "Investigating Implicit Cues for User State Estimation in Human-Robot Interaction Using Physiological Measurements". In: *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 1125–1130. ISBN: 978-1-4244-1634-9. DOI: 10.1109/ROMAN.2007.4415249. URL: <http://ieeexplore.ieee.org/document/4415249/>.
- Mutlu, B.; Yamaoka, F.; Kanda, T.; Ishiguro, H.; & Hagita, N. (2009). "Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior". In: *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction - HRI '09*. New York, New York, USA: ACM Press, pp. 69–76. ISBN: 9781605584041. DOI: 10.1145/1514095.1514110. URL: <http://portal.acm.org/citation.cfm?doid=1514095.1514110>.
- Nanna, M. J.; & Sawilowsky, S. S. (1998). "Analysis of Likert scale data in disability and medical rehabilitation research". In: *Psychological Methods* 3.1, pp. 55–67. ISSN: 1082-989X. DOI: 10.1037//1082-989x.3.1.55.
- Narayanan, V.; Manoghar, B. M.; Dorbala, V. S.; Manocha, D.; & Bera, A. (2020). "ProxEmo: Gait-based emotion learning and multi-view proxemic fusion for socially-aware robot navigation". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8200–8207. ISBN: 9781728162119. arXiv: 2003.01062.
- Parkinson, B. (2005). "Do Facial Movements Express Emotions or Communicate Motives?" In: *Personality and Social Psychology Review* 9.4, pp. 278–311. ISSN: 1088-8683. DOI: 10.1207/s15327957pspr0904_1. URL: http://journals.sagepub.com/doi/10.1207/s15327957pspr0904_1.

- Pasquali, D.; Aroyo, A. M.; Gonzalez-Billandon, J.; Rea, F.; Sandini, G.; & Sciutti, A. (2020). "Your Eyes Never Lie: A Robot Magician Can Tell if You Are Lying". In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, pp. 392–394. ISBN: 9781450370578. DOI: 10.1145/3371382.3378253. URL: <https://dl.acm.org/doi/10.1145/3371382.3378253>.
- Perneger, T. V. (2021). "How to use likelihood ratios to interpret evidence from randomized trials". In: *Journal of Clinical Epidemiology* 136, pp. 235–242. ISSN: 08954356. DOI: 10.1016/j.jclinepi.2021.04.010. URL: <https://doi.org/10.1016/j.jclinepi.2021.04.010><https://linkinghub.elsevier.com/retrieve/pii/S0895435621001323>.
- Peters, G.-J. Y. (2014). "The alpha and the omega of scale reliability and validity". In: *The European Health Psychologist* 16.2, pp. 56–69. ISSN: 22256962.
- Plummer, M. (2003). "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling". In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Vienna, Austria.
- Rau, P. P.; Li, Y.; & Li, D. (2009). "Effects of communication style and culture on ability to accept recommendations from robots". In: *Computers in Human Behavior* 25.2, pp. 587–595. ISSN: 07475632. DOI: 10.1016/j.chb.2008.12.025. URL: <http://dx.doi.org/10.1016/j.chb.2008.12.025><https://linkinghub.elsevier.com/retrieve/pii/S0747563208002367>.
- Robins, B.; Dautenhahn, K.; & Nadel, J. (2018). "Kaspar, the social robot and ways it may help children with autism - an overview". In: *Enfance* 1.1, p. 91. ISSN: 0013-7545. DOI: 10.3917/enf2.181.0091. URL: <http://www.cairn.info/revue-enfance-2018-1-page-91.htm?ref=doi>.
- Roesler, E.; Manzey, D.; & Onnasch, L. (2021). "A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction". In: *Science Robotics* 6.58, pp. 1–11. ISSN: 2470-9476. DOI: 10.1126/scirobotics.abj5425. URL: <https://www.science.org/doi/10.1126/scirobotics.abj5425>.
- Schrum, M. L.; Johnson, M.; Ghuy, M.; & Gombolay, M. C. (2020). "Four Years in Review: Statistical Practices of Likert Scales in Human-Robot Interaction Studies". In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, pp. 43–52. ISBN: 9781450370578. DOI: 10.

- 1145/3371382.3380739. arXiv: 2001.03231. URL: <https://dl.acm.org/doi/10.1145/3371382.3380739>.
- Scoglio, A. A.; Reilly, E. D.; Gorman, J. A.; & Drebing, C. E. (2019). “Use of Social Robots in Mental Health and Well-Being Research: Systematic Review”. In: *Journal of Medical Internet Research* 21.7, e13322. ISSN: 1438-8871. DOI: 10.2196/13322. URL: <http://www.jmir.org/2019/7/e13322/>.
- Sebanz, N.; Bekkering, H.; & Knoblich, G. (2006). “Joint action: bodies and minds moving together”. In: *Trends in Cognitive Sciences* 10.2, pp. 70–76. ISSN: 13646613. DOI: 10.1016/j.tics.2005.12.009. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1364661305003566>.
- Seib, V.; Giesen, J.; Grüntjens, D.; & Paulus, D. (2013). “Enhancing human-robot interaction by a robot face with facial expressions and synchronized lip movements”. In: *21st International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, WSCG 2013 - Communication Papers Proceedings*, pp. 70–77.
- Shah, J.; Wiken, J.; Williams, B.; & Breazeal, C. (2011). “Improved human-robot team performance using chaski, a human-inspired plan execution system”. In: *Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11*. New York, New York, USA: ACM Press, pp. 29–36. ISBN: 9781450305617. DOI: 10.1145/1957656.1957668. URL: <http://portal.acm.org/citation.cfm?doid=1957656.1957668>.
- Sprenst, P.; & Smeeton, N. C. (2001). *Applied nonparametric statistical methods*. Chapman & Hall/CRC. ISBN: 1-58488-145-3.
- Takayama, L.; & Pantofaru, C. (2009). “Influences on proxemic behaviors in human-robot interaction”. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 5495–5502. ISBN: 978-1-4244-3803-7. DOI: 10.1109/IRoS.2009.5354145. URL: <http://ieeexplore.ieee.org/document/5354145/>.
- Tanaka, M.; Ishii, A.; Yamano, E.; Ogikubo, H.; Okazaki, M.; Kamimura, K.; Konishi, Y.; Emoto, S.; & Watanabe, Y. (2012). “Effect of a human-type communication robot on cognitive function in elderly women living alone”. In: *Medical Science Monitor* 18.9, CR550–CR557. ISSN: 1234-1010. DOI: 10.12659/MSM.883350. URL: <http://www.medscimonit.com/abstract/index/idArt/883350>.

- Tatsukawa, K.; Takahashi, H.; Yoshikawa, Y.; & Ishiguro, H. (2019). “Android Pretending to Have Similar Traits of Imagination as Humans Evokes Stronger Perceived Capacity to Feel”. In: *Frontiers in Robotics and AI* 6.September, pp. 1–9. ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00088. URL: <https://www.frontiersin.org/article/10.3389/frobt.2019.00088/full>.
- Timm, F.; & Barth, E. (2011). “Accurate Eye Centre Localisation by Means of Gradients”. In: *Proceedings of the International Conference on Computer Vision Theory and Applications*. SciTePress - Science, and Technology Publications, pp. 125–130. ISBN: 978-989-8425-47-8. DOI: 10.5220/0003326101250130.
- Toh, L. P. E.; Causo, A.; Tzuo, P. W.; Chen, I. M.; & Yeo, S. H. (2016). “A review on the use of robots in education and young children”. In: *Educational Technology and Society* 19.2, pp. 148–163. ISSN: 14364522.
- Tölgyessy, M.; Dekan, M.; & Hubinský, P. (2018). “Human-Robot Interaction Using Pointing Gestures”. In: *Proceedings of the 2nd International Symposium on Computer Science and Intelligent Control*. New York, NY, USA: ACM, pp. 1–5. ISBN: 9781450366281. DOI: 10.1145/3284557.3284718. URL: <https://dl.acm.org/doi/10.1145/3284557.3284718>.
- Unhelkar, V. V.; Dorr, S.; Bubeck, A.; Lasota, P. A.; Perez, J.; Siu, H. C.; Boerkoel, J. C.; Tyroller, Q.; Bix, J.; Bartscher, S.; & Shah, J. A. (2018). “Mobile Robots for Moving-Floor Assembly Lines: Design, Evaluation, and Deployment”. In: *IEEE Robotics & Automation Magazine* 25.2, pp. 72–81. ISSN: 1070-9932. DOI: 10.1109/MRA.2018.2815639. URL: <https://ieeexplore.ieee.org/document/8360770/>.
- Unhelkar, V. V.; Yang, X. J.; & Shah, J. A. (2017). “Challenges for Communication Decision-Making in Sequential Human-Robot Collaborative Tasks”. In: *Workshop on Mathematical Models, Algorithms, and Human-Robot Interaction at Robotics: Science and Systems*.
- Venkatesh, V.; Morris, M. G.; Davis, G. B.; & Davis, F. D. (2003). “User Acceptance of Information Technology: Toward a Unified View”. In: *MIS Quarterly* 27.3, pp. 425–478. ISSN: 02767783. DOI: 10.2307/30036540. URL: <http://www.crcnetbase.com/doi/10.1201/9780849375477.ch230https://www.jstor.org/stable/10.2307/30036540>.
- Venture, G.; & Kulić, D. (2019). “Robot Expressive Motions: A Survey of Generation and Evaluation Methods”. In: *ACM Transactions on Human-Robot Interaction* 8.4. ISSN:

25739522. DOI: 10.1145/3344286. URL: <http://dl.acm.org/citation.cfm?doid=3372354.3344286>.
- Viola, P.; & Jones, M. (2001). "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*. IEEE Comput. Soc. ISBN: 0-7695-1272-0. DOI: 10.1109/CVPR.2001.990517. URL: <http://ieeexplore.ieee.org/document/990517/>.
- Wagenmakers, E.-J.; Marsman, M.; Jamil, T.; Ly, A.; Verhagen, J.; Love, J.; Selker, R.; Gronau, Q. F.; Šmíra, M.; Epskamp, S.; Matzke, D.; Rouder, J. N.; & Morey, R. D. (2018). "Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications". In: *Psychonomic Bulletin & Review* 25.1, pp. 35–57. ISSN: 1069-9384. DOI: 10.3758/s13423-017-1343-3. URL: <http://link.springer.com/10.3758/s13423-017-1343-3>.
- Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6, pp. 80–83. ISSN: 00994987. DOI: 10.2307/3001968. URL: <http://academic.oup.com/jee/article/39/2/269/2203647/Individual-Comparisons-of-Grouped-Data-by-Ranking><http://www.ncbi.nlm.nih.gov/pubmed/20983181><https://www.jstor.org/stable/10.2307/3001968?origin=crossref>.
- Zhang, H.; Fricker, D.; & Yu, C. (2010). "A Multimodal Real-Time Platform for Studying Human-Avatar Interactions". In: *International Conference on Intelligent Virtual Agents*, pp. 49–56. ISBN: 3642158919. DOI: 10.1007/978-3-642-15892-6_6. URL: http://link.springer.com/10.1007/978-3-642-15892-6_6.
- Zhang, Z. (2012). "Microsoft Kinect Sensor and Its Effect". In: *IEEE Multimedia* 19.2, pp. 4–10. ISSN: 1070-986X. DOI: 10.1109/MMUL.2012.24. URL: <http://ieeexplore.ieee.org/document/6190806/>.
- Zoghbi, S.; Parker, C. A. C.; Croft, E. A.; & Van der Loos, H. F. M. (2010). "Enhancing Collaborative Human-Robot Interaction Through Physiological-Signal Based Communication". In: *IEEE Conference on Robotics and Automation: Workshop on Multimodal Human-Robot Interfaces*.



AVALIAÇÕES DOS SISTEMAS DE COMUNICAÇÃO

A.1 Gestos de apontar

O sistema para reconhecimento e interpretação de gestos de apontar, descrito na Seção 3.2, foi avaliado em um experimento. Os resultados foram apresentados nos trabalhos Campos & Adorno (2019) e Campos & Adorno (2020). Participantes deveriam executar uma sequência de 10 gestos indicando um dos quatro objetos presentes no ambiente ou nenhum deles. A sequência executada era escolhida pela pessoa, que a informava à experimentadora ao final do experimento. Os objetos utilizados foram caixas de dimensões e localizações conhecidas e os planos criados para os modelos consideravam dimensões 20% maiores que as reais, para garantir que os objetos estivessem contidos em seus modelos. Sem poder ver a execução dos gestos, a experimentadora era responsável por documentar as respostas do sistema, que deveria informar se algum objeto estava sendo indicado e, se sim, qual deles. A Figura A.1 mostra o ambiente onde os experimentos foram realizados e uma ilustração da vista de topo, indicando os objetos, a câmera RGB-D, a experimentadora e uma aproximação da área em que a pessoa podia se movimentar (área pontilhada). O experimento foi filmado com o conhecimento e consentimento dos participantes.

Ao todo, 11 pessoas participaram do experimento, mas duas foram excluídas da análise por utilizarem gestos que não eram considerados gestos de apontar. Esses participantes apenas levantavam o braço e o posicionavam acima do objeto indicado, ao invés de

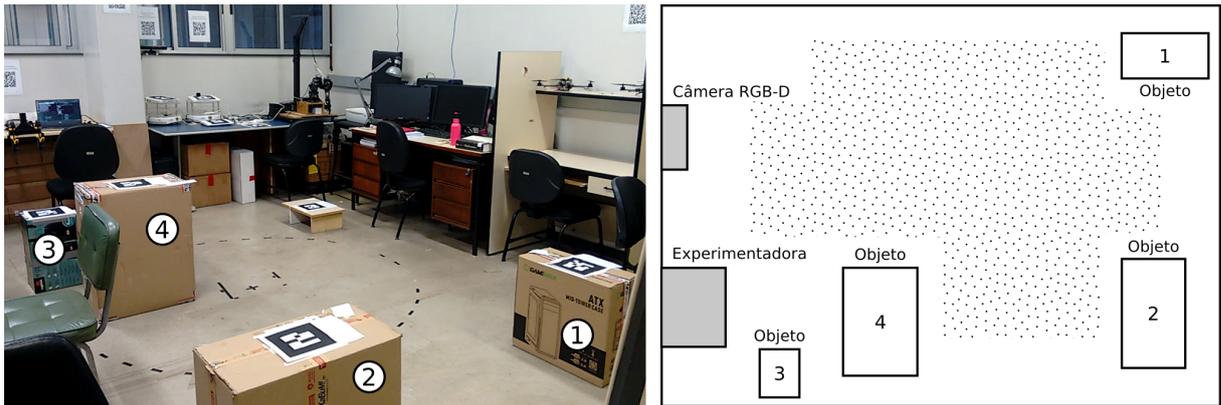


Figure A.1: Ambiente utilizado para os experimentos, com os objetos numerados. À direita, a representação da vista de topo indica a posição de cada um dos objetos, da câmera RGB-D e da experimentadora, além do espaço de movimentação dos participantes (área pontilhada).

apontarem diretamente para ele. Além disso, um terceiro participante informou uma sequência diferente da executada, e, portanto, a sequência de referência foi corrigida para a análise, de acordo com os vídeos do experimento. A Figura A.2 ilustra um gesto de apontar, mostrando os modelos dos objetos com dimensões maiores que as reais e a linha criada a partir dos pontos de cotovelo e mão.

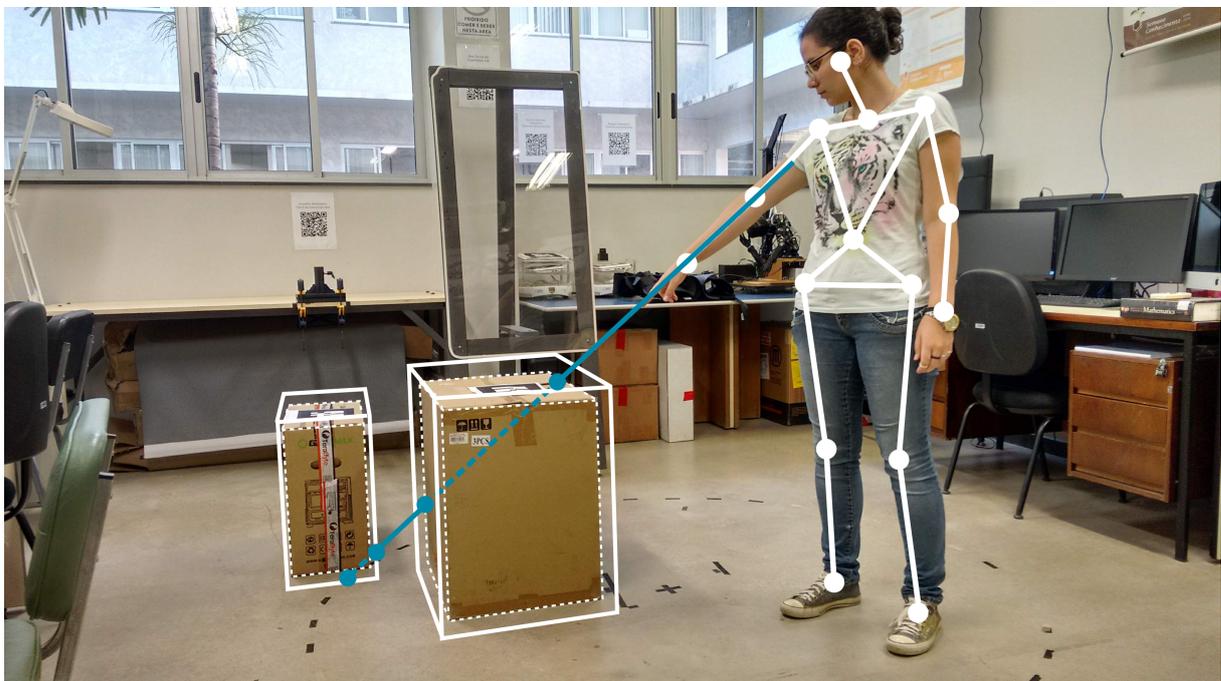


Figure A.2: Exemplo de um gesto de apontar. Objetos foram definidos com dimensões maiores que as reais. Na situação ilustrada, a linha do gesto (azul) cruza dois objetos ao mesmo tempo e o mais próximo do ponto no braço (caixa maior) é escolhido como indicação.

O número de acertos e erros do sistema para cada participante é mostrado na Figura A.3.

A taxa média de acertos foi de 77,8% com desvio padrão de 16,4%. Considerando todos os gestos executados por todos os participantes, a taxa de falsos negativos, ou seja, quando o sistema não detecta a indicação de um objeto enquanto o participante aponta para um deles, foi de 32,2%. Não ocorreram casos de falso positivo, isto é, quando o sistema detecta a indicação de um objeto enquanto o participante não aponta para nenhum deles. Apenas uma vez o sistema indicou um objeto enquanto a pessoa apontava para outro. Esse erro envolvia os objetos 3 e 4 (ver Figura A.1), que foram posicionados de modo a possibilitarem uma ambiguidade como a ilustrada na Figura A.2: dependendo do ponto de vista da pessoa, o objeto 3 estava oculto pelo objeto 4, que tinha dimensões maiores. Duas vezes o sistema oscilou entre indicar o objeto 3 ou o objeto 4 e, nesses casos, se o objeto apontado pelo participante era um deles, foi considerado um acerto.

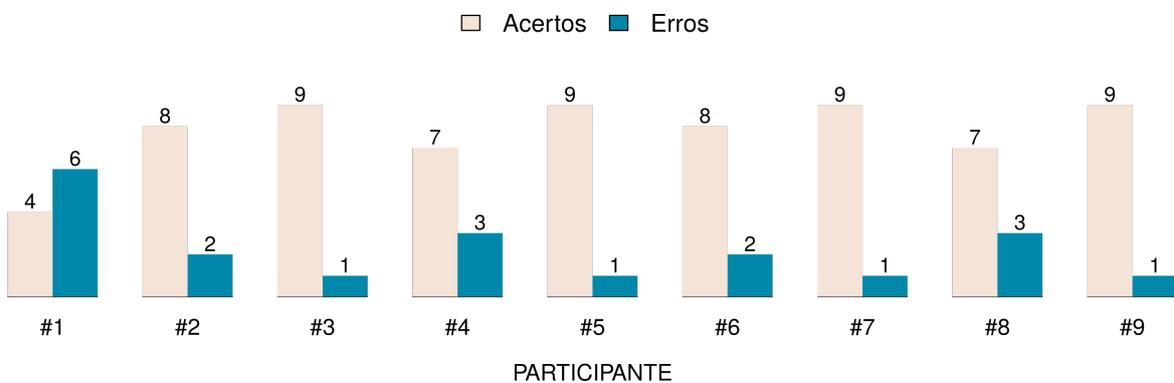


Figure A.3: Acertos e erros do sistema sobre os gestos de cada participante.

Ao todo foram registrados 20 erros. A maioria dos erros (oito) são de falsos negativos relacionados ao objeto 3. Já durante os testes para a realização do experimento, observou-se maior dificuldade no reconhecimento de gestos que apontavam para o objeto 3, mesmo quando realizados pela especialista do sistema. Além disso, foram observados erros nas medições das posições das juntas, que, para a realização dos experimentos, foram compensados com uma transformação obtida experimentalmente. Um objeto só é detectado pelo sistema se a linha do gesto cruza o interior do modelo criado com os planos, fazendo com que a detecção de objetos maiores seja mais fácil. O objeto 3 era o menor entre os utilizados nos experimentos, portanto, acredita-se que isso pode ter contribuído para o número maior de erros relacionados a ele, juntamente com as incertezas de medição da câmera RGB-D. Falsos negativos relacionados a outros objetos somaram quatro erros. A partir da análise dos vídeos do experimento, outros sete erros de falsos negativos foram atribuídos à oclusão do braço de apontar. Em alguns momentos, o participante, com o próprio corpo, ocultava da câmera o braço que usava para apontar, causando perda de informação e dificuldade no reconhecimento e interpretação do gesto.

Os resultados indicam que soluções simples, como a utilização de mais sensores espalhados pelo ambiente para rastrear a cadeia cinemática do humano, podem melhorar o desempenho do sistema, reduzindo os casos de oclusão do braço que aponta e os erros

causados por incertezas de medição. Objetos de dimensões grandes e bem espalhados pelo ambiente facilitam a detecção correta, por isso a inserção de mais sensores, tornando as medições mais precisas, pode ser bastante vantajoso em aplicações em que os objetos e regiões são menores e se encontram mais próximos uns dos outros. Além disso, na comunicação humana, a identificação de gestos de apontar não é tão rígida; duas pessoas interagindo conseguem interpretar os gestos uma da outra pelo contexto ou por eliminação, não sendo necessário que se aponte precisamente para o objeto de interesse. Dependendo do ambiente e do contexto da interação, as condições para detecção de objetos indicados podem ser relaxadas, para reduzir as taxas de falsos negativos, e outras informações podem ser consideradas para auxiliar na interpretação dos gestos de apontar. Por fim, em aplicações em que as pessoas interagem por um longo tempo ou repetidas vezes com o sistema e que permitem que se saiba quando o sistema detecta ou não o objeto indicado, o desempenho também pode ser melhor, já que o humano se familiariza e consegue se adaptar às habilidades do sistema, assim como faria ao interagir com outra pessoa.

A.2 Detecção de atributos faciais

O algoritmo final de detecção dos atributos faciais de interesse passou por testes preliminares e os resultados foram apresentados no trabalho Campos & Adorno (2020). Duas pessoas foram gravadas executando uma sequência de movimentos com a cabeça, gerando sete vídeos. Manualmente, pontos dos centros das íris e da boca em cada quadro de cada vídeo foram obtidos para serem usados como referência. Seja D_m a distância calculada com os pontos obtidos manualmente e D_d a distância calculada com os pontos detectados pelo algoritmo, sendo D_m e D_d as distâncias horizontais entre os olhos ou verticais entre olhos e boca. Os erros $E_{px} = |D_d - D_m|$, em pixels, para cada quadro foram calculados e convertidos em centímetros usando a razão entre a distância real entre os olhos $D_{o,real_{cm}}$ e a distância de referência $D_{o,ref_{px}}$ (quando a pessoa olha para frente) em pixels, obtendo-se $E_{cm} = RE_{px}$, onde $R = D_{o,real_{cm}}/D_{o,ref_{px}}$, tanto para a distância entre os centros das íris quanto para entre olhos e boca. A distância entre a pessoa e a câmera foi mantida fixa durante a execução dos movimentos.

A partir da média de erros de cada vídeo, as médias globais de erro (e desvios padrão) foram calculadas: 0,52 cm (desvio padrão de 0,29 cm) para a distância horizontal entre os centros das íris e 0,67 cm (desvio padrão de 0,31 cm) para a distância vertical entre olhos e boca. A Figura A.4 mostra exemplos de erros extraídos dos vídeos testados. A primeira imagem mostra falha na detecção de um olho e da boca (indicados pela ausência de um retângulo verde relacionado ao olho direito da pessoa e de um retângulo vermelho relacionado à boca). A segunda imagem mostra um momento quando, após algumas iterações sem detecção do olho direito, o olho esquerdo é detectado como se fosse o da direita, atualizando o filtro com um viés (indicado pelos retângulos duplicados sobre o

olho esquerdo). Na última imagem, o olho direito foi novamente detectado incorretamente (indicado pelo retângulo verde próximo à boca). Os filtros se mostraram suficientemente capazes de manter os erros baixos quando ocorreram falhas de detecção e, forçando a reinicialização após algumas iterações sem detecção ou quando dois olhos detectados eram provavelmente o mesmo, o algoritmo foi capaz de se recuperar de erros como os das duas últimas imagens da Figura A.4.

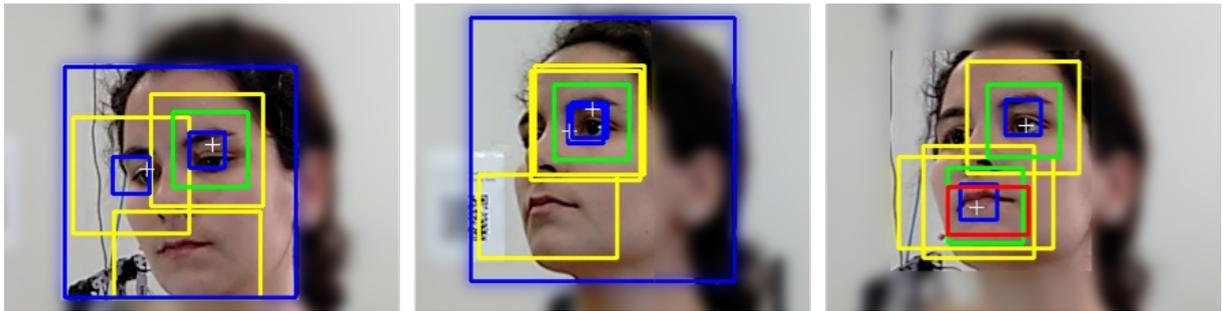


Figure A.4: Erros na detecção de atributos faciais. Da esquerda para a direita: olho direito e boca não detectados; olho esquerdo detectado como o direito; e olho direito detectado próximo à boca.

A.3 Expressões faciais da agente virtual

O reconhecimento das expressões faciais criadas para a agente virtual, descrita na Seção 3.4, foi avaliado utilizando dois grupos: participantes do experimento de interação descrito na Seção 4.2 e pessoas que não tiveram interação com as agentes virtuais. Um formulário apresentava cada uma das seis expressões criadas e pedia que as pessoas escolhessem uma dentre dez opções de resposta, sendo elas “neutra”, “feliz”, “confusa”, “triste”, “raiva”, “satisfeita”, “medo”, “surpresa”, “nojo” e “outra”, esta última acompanhada de um campo para preenchimento. As opções extras de resposta (“medo”, “surpresa” e “nojo”) foram escolhidas arbitrariamente apenas para que se tivesse mais opções de respostas do que as que correspondem às expressões mostradas. A ordem das expressões apresentadas e das opções de resposta eram definidas aleatoriamente para cada participante.

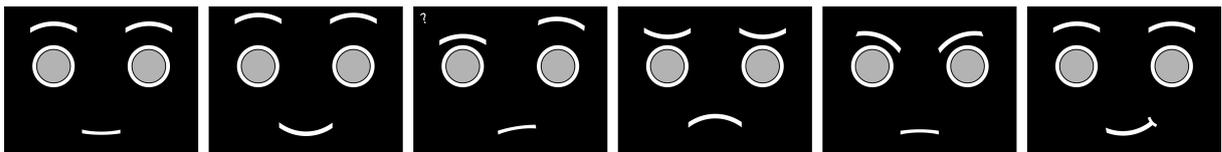


Figure A.5: Imagens usadas para avaliação do reconhecimento das expressões faciais.

Para participantes do experimento presencial, o questionário era respondido virtualmente ao final da interação com as duas agentes virtuais. Participantes sem interação foram pessoas próximas da equipe de pesquisa e responderam remotamente através de

formulário online. Para ambos os grupos, as imagens mostradas eram de agentes virtuais com olhos de cor cinza, como mostrado na Figura A.5, para diferenciar das agentes Luna e Sofia com quem participantes do experimento presencial interagiam (ver Figura 4.1).



*Outras expressões mencionadas por participantes (C) com e (S) sem interação:
 Neutra: (C) apática; Confusa: (C) dúvida, (S) questionamento; Triste: (C) decepcionada (2x); Satisfeita: (C) indecisa, (S) pensativa

Figure A.6: Resultados da avaliação do reconhecimento das expressões faciais para participantes com e sem interação com as agentes virtuais. Cada dupla de barras (com e sem interação) mostra o número de respostas em cada uma das dez opções dadas (indicadas na parte inferior da figura, eixo horizontal) para cada uma das seis expressões apresentadas (indicadas à esquerda na figura, eixo vertical).

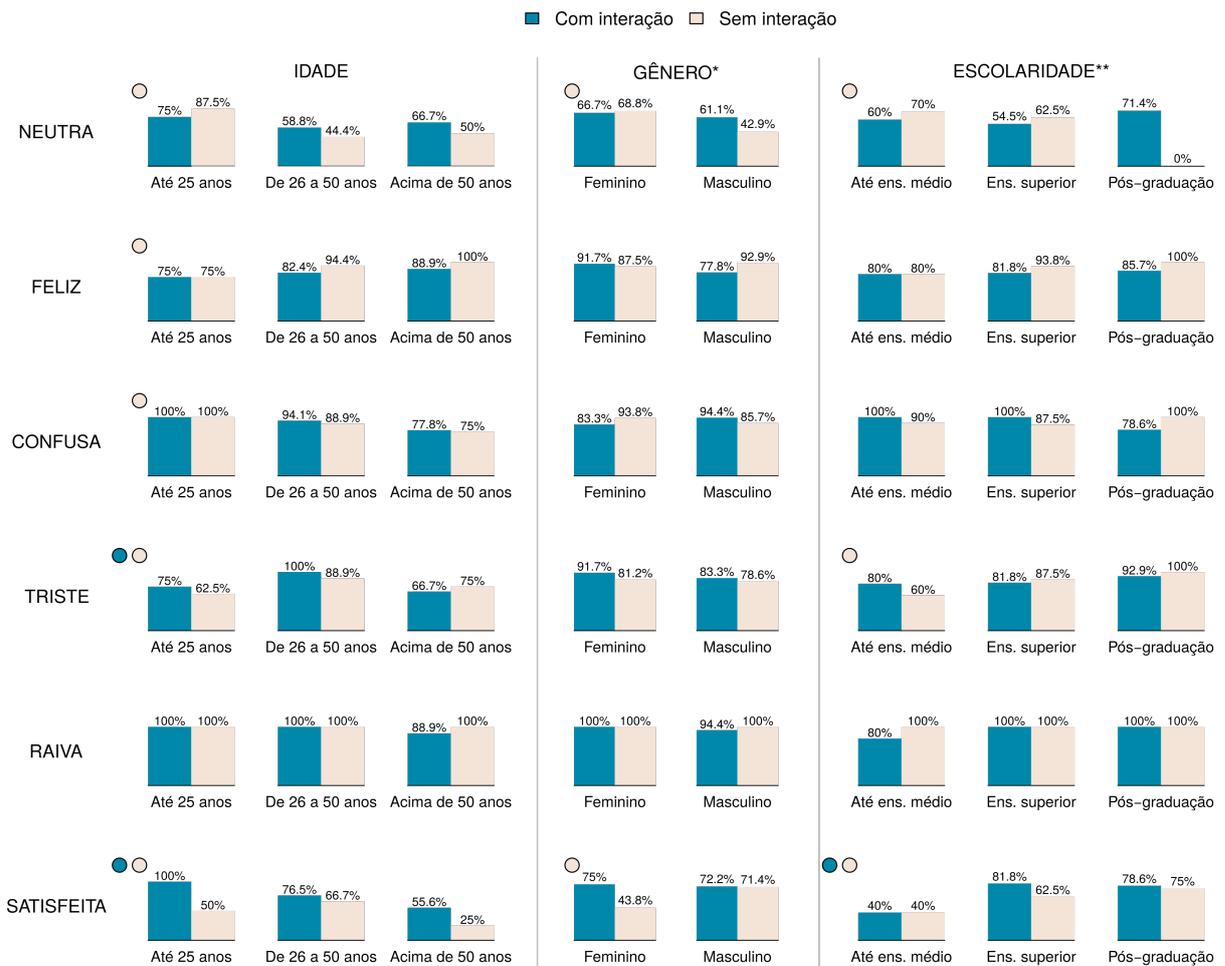
Ao todo, 60 pessoas responderam ao questionário, 30 de cada grupo. A Figura A.6 mostra os resultados da avaliação. As expressões foram corretamente reconhecidas pela maioria das pessoas com e sem interação com as agentes virtuais. A expressão neutra foi mais confundida com as expressões feliz, satisfeita e surpresa; a expressão de raiva foi a mais reconhecida, com apenas um erro de participante com interação; e a expressão satisfeita foi mais confundida com neutra, feliz e surpresa. Exceto pela expressão de raiva,

todas as outras foram criadas para serem usadas na interação proposta neste trabalho (ver Seção 4.2). Na interação, a expressão satisfeita é exibida quando a pessoa indica a cor correta e a sua intenção é demonstrar a satisfação ou aprovação da agente virtual pelo acerto. A expressão feliz poderia ter sido usada nesse contexto, mas desejava-se uma alternativa mais sutil e, por isso, já se esperava que a expressão satisfeita pudesse ser interpretada como feliz, o que não causava prejuízo para a interação proposta. Já a expressão neutra foi criada para ser a expressão padrão quando outra não estivesse sendo utilizada, especialmente quando a agente virtual não usava expressões faciais para se comunicar (nesse caso, apenas a neutra era exibida). As expressões neutra e satisfeita já eram, portanto, consideradas as mais sutis e passíveis de erros. Por outro lado, o fato de seis pessoas terem interpretado alguma expressão como uma expressão de medo não era esperado, já que nenhuma das imagens foi criada com a intenção de transmitir nada parecido com medo ou nojo e as opções de resposta extras foram escolhidas arbitrariamente.

Idade, gênero (opções *Feminino*, *Masculino*, *Prefiro não dizer* e *Outros*, sendo a última com campo para preenchimento) e escolaridade (opções de *Ensino fundamental incompleto* a *Doutorado completo*) de participantes também foram coletados com os questionários. No grupo com interação, a média de idade de participantes foi de 37,2 anos (mínima: 21, máxima: 61) e 60% se declararam do gênero masculino. Já no grupo sem interação, a média de idade foi de 33,8 anos (mínima: 18, máxima: 61) e 53,33% de participantes do gênero feminino. A escolaridade do grupo com interação abrangeu desde ensino médio completo até doutorado incompleto, com maioria de doutorado incompleto e ensino superior completo (33,33% e 26,67%, respectivamente). Participantes sem interação tinham desde ensino médio incompleto até doutorado completo, sendo ensinos superior e médio completos as categorias mais representadas (46,67% e 23,33%, respectivamente).

A Figura A.7 mostra a taxa de acertos para cada expressão considerando perfis de idade, gênero e escolaridade. A idade foi dividida em três categorias: até 25 anos; de 26 a 50 anos; e acima de 50 anos. Apenas as duas primeiras opções (Feminino e Masculino) para gênero foram utilizadas por participantes e por isso elas são as únicas categorias mostradas. Por fim, escolaridade foi dividida também em três categorias: de ensino fundamental incompleto até ensino médio completo; ensino superior incompleto e completo; e de mestrado incompleto a doutorado completo.

Círculos no canto esquerdo superior de cada bloco de histogramas na Figura A.7 indicam diferenças de pelo menos 25% entre categorias. Por exemplo, foi observada diferença de 25% ou mais na taxa de acertos entre as categorias de idade no reconhecimento da expressão neutra por pessoas que não interagiram com a agente virtual (círculo rosa no bloco *neutra-idade*); para a expressão satisfeita, houve diferença de pelo menos 25% entre categorias de escolaridade tanto para quem interagiu com as agentes virtuais quanto para quem não interagiu (círculo azul e círculo rosa no bloco *satisfeita-escolaridade*). Idade foi o fator que mais apresentou diferenças de 25% ou mais na taxa de acertos das expressões, especialmente



*As opções de resposta para gênero nos questionários eram Feminino, Masculino, Prefiro não dizer e Outros (com campo para preenchimento). Apenas as opções Feminino e Masculino foram selecionadas por participantes.
 **As opções de resposta para escolaridade nos questionários iam de Ensino fundamental incompleto até Doutorado completo. A categoria "Até ens. médio" inclui ensinos fundamental e médio incompletos e completos; "Ens. superior" corresponde a ensino superior incompleto e completo; e "Pós-graduação" engloba mestrado e doutorado incompletos e completos.

Figure A.7: Taxa de acertos no reconhecimento das expressões faciais considerando idade, gênero e escolaridade de participantes com e sem interação com as agentes virtuais.

para o grupo sem interação, para o qual apenas a expressão de raiva teve reconhecimento similar entre as três categorias. Para o grupo sem interação, foram observadas diferenças de pelo menos 25% na taxa de acertos das expressões neutra e satisfeita entre categorias de todos os três fatores e a expressão satisfeita foi a única do grupo com interação para a qual se observou diferença em mais de um deles (idade e escolaridade).

Para quem interagiu com as agentes virtuais, dois outros fatores foram analisados e os resultados são mostrados na Figura A.8. O formulário do grupo com interação questionava sobre a familiaridade com agentes virtuais como Luna e Sofia e com qual delas a pessoa preferiu interagir. As agentes virtuais diferenciavam duas configurações de comunicação, EX e EXIM (ver Seção 4.2), e a agente virtual de preferência indicava qual a configuração de comunicação preferida. Na configuração EX apenas a expressão neutra era mostrada; na configuração EXIM, as expressões neutra, feliz e satisfeita eram sempre exibidas no decorrer da interação, mas as expressões confusa e triste podiam ser usadas ou não, dependendo

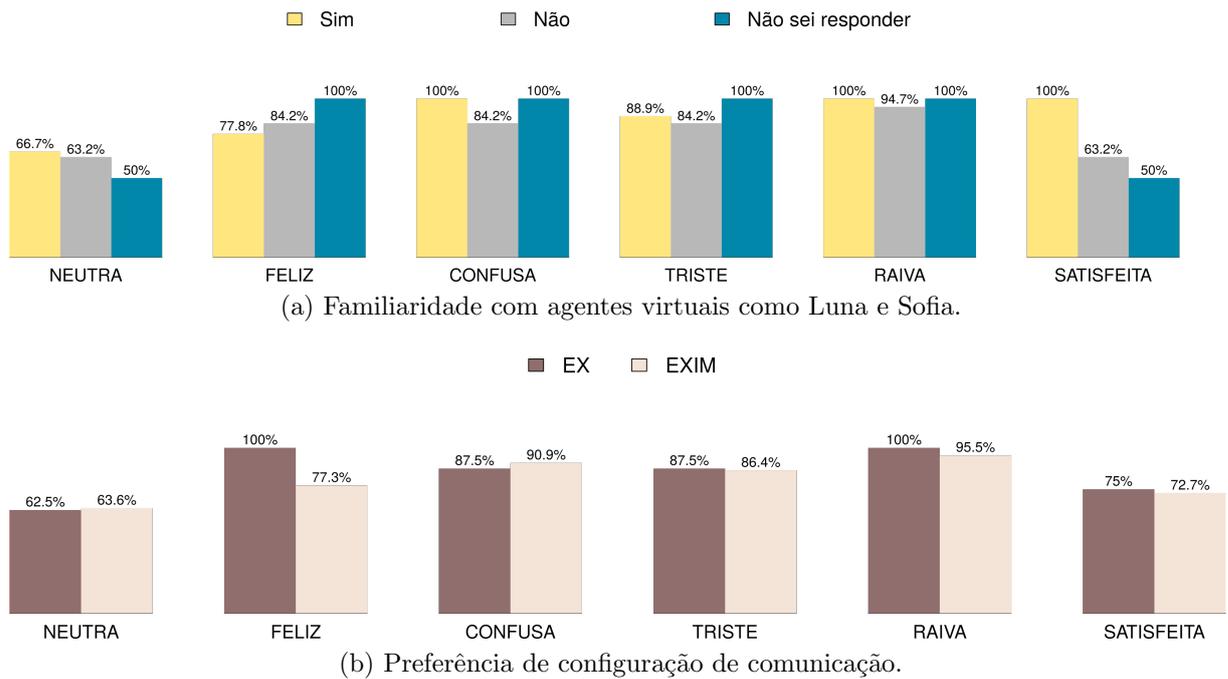


Figure A.8: Taxa de acertos no reconhecimento das expressões faciais para participantes com interação considerando familiaridade com agentes virtuais e preferência de configuração de comunicação.

da necessidade durante a realização da tarefa; a expressão de raiva nunca era exibida em nenhuma das duas configurações de comunicação. Diferenças de 25% ou mais na taxa de acertos foram observadas apenas entre categorias de familiaridade com agentes virtuais para a expressão satisfeita. A preferência de configuração de comunicação não parece ter influenciado no reconhecimento das expressões faciais.

As Figuras A.9 e A.10 mostram todas as respostas para ambos os grupos, com e sem interação. Os pontos são caracterizados pelo perfil de participante de cada resposta, considerando idade, gênero e escolaridade. Por exemplo, para o grupo com interação (Figura A.9), a expressão confusa foi confundida com a expressão de raiva por uma pessoa com idade acima de 50 anos, do gênero feminino e com mestrado ou doutorado incompleto ou completo (triângulo marrom não preenchido na área *confusa-raiva* da Figura A.9); já no grupo sem interação (Figura A.10), uma pessoa de até 25 anos, do gênero masculino e com escolaridade no máximo até ensino médio completo interpretou a expressão feliz como surpresa (círculo azul preenchido na área *feliz-surpresa* da Figura A.10).

Dos três erros no reconhecimento da expressão confusa por participantes com interação, mostrados na Figura A.9, a expressão foi exibida durante a interação apenas para a participante que a confundiu com raiva e a expressão mencionada pela participante que respondeu “Outra” foi *dúvida*. Já para os quatro erros na expressão triste (Figura A.9), apenas para o participante do gênero masculino com até 25 anos a expressão não foi exibida e as duas vezes que a opção “Outra” foi escolhida, a expressão *decepcionada* foi citada.

As Figuras A.6 a A.10 permitem apenas algumas análises gerais, já que os conjuntos

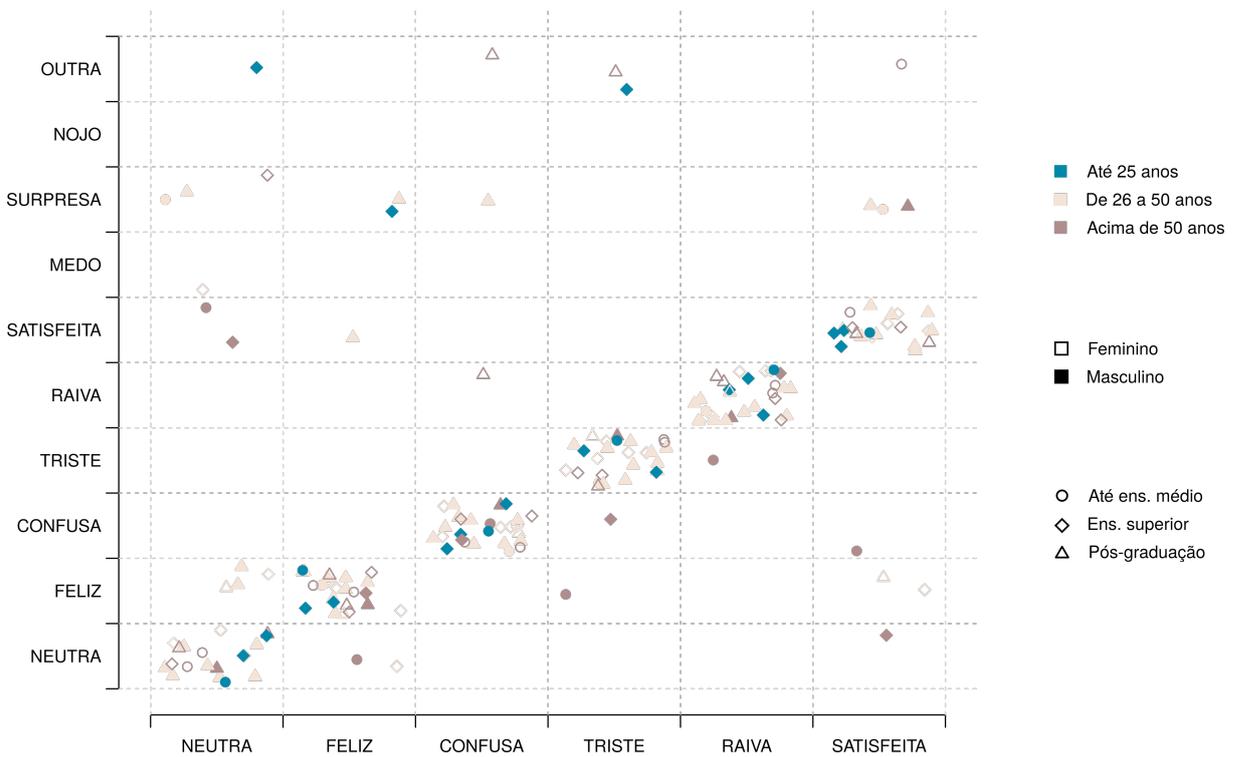


Figure A.9: Todas as respostas de participantes *com* interação com as agentes virtuais sobre as expressões faciais, caracterizando-se o perfil de participante de cada uma delas.

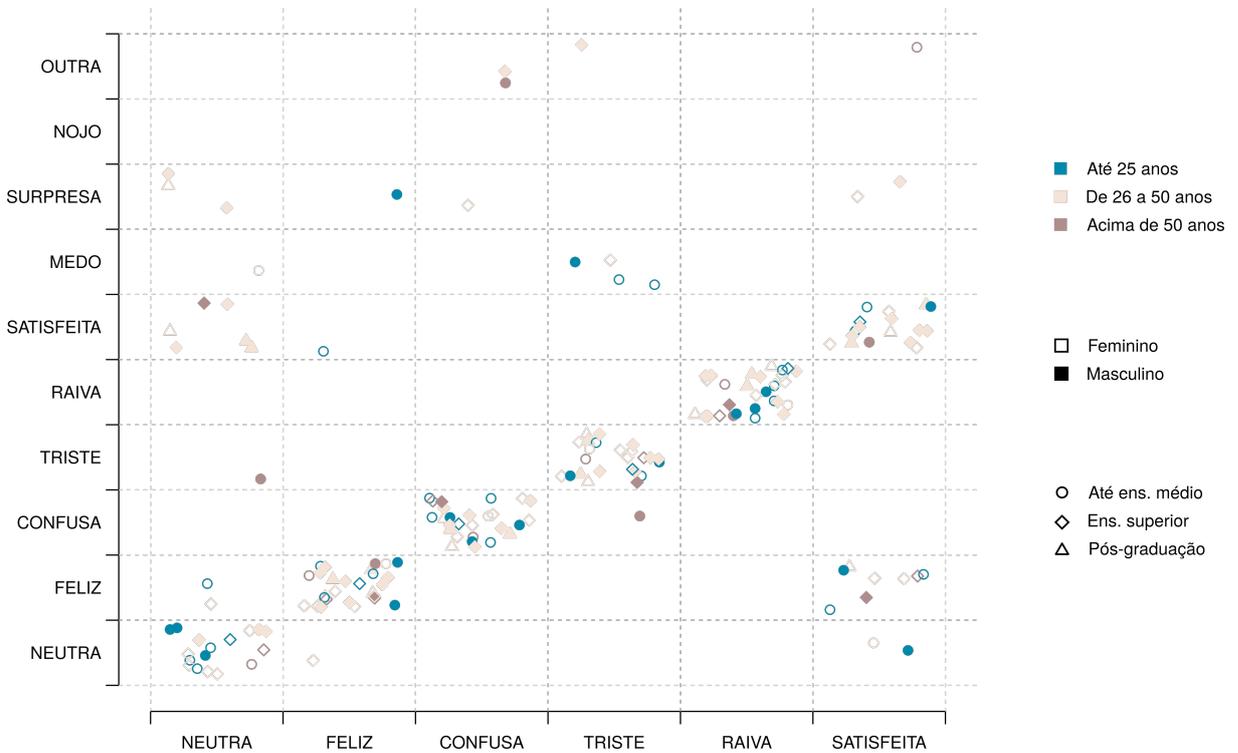


Figure A.10: Todas as respostas de participantes *sem* interação com as agentes virtuais sobre as expressões faciais, caracterizando-se o perfil de participante de cada uma delas.

de dados são pequenos e não muito diversos e o foco da pesquisa não está nas expressões faciais criadas para a agente virtual, sendo que um planejamento direcionado para isso seria necessário para se tirar conclusões fortes sobre o reconhecimento. Entretanto, os resultados podem ser usados como guia para possíveis melhorias e para o levantamento de hipóteses para pesquisas futuras. Os resultados gerais sobre o reconhecimento, mostrados na Figura A.6, sugerem que se a interação proposta entre humano e agente virtual necessitar do reconhecimento mais específico das expressões, algumas delas talvez precisem ser melhor caracterizadas, como as expressões neutra e satisfeita. Os resultados mostrados na Figura A.7 sugerem que a idade pode ser um fator importante no reconhecimento de expressões faciais e, caso o alvo de uma aplicação seja uma faixa etária específica, pode ser interessante se realizar estudos mais profundos sobre isso. Já as Figuras A.9 e A.10 podem ser usadas para se buscar possíveis padrões que possam ser melhor investigados em outras pesquisas. Por exemplo, a Figura A.10 mostra que a expressão neutra foi confundida majoritariamente por pessoas de 26 a 50 anos, especialmente do gênero masculino, e que a interpretação da expressão triste como uma expressão de medo foi feita, em sua maioria, por pessoas de até 25 anos com escolaridade até ensino médio completo.

B

DETALHES DA INTERAÇÃO PROPOSTA

A Figura B.1 mostra o fluxo do experimento de interação. A participação se inicia com uma interação com a experimentadora, que apresenta o termo de consentimento da pesquisa, informa as medidas de biossegurança tomadas em relação à pandemia de COVID-19¹ e dá instruções gerais sobre o experimento. Participantes então entram e se fecham na sala em que os sistemas foram montados enquanto a experimentadora permanece do lado de fora. O experimento começa pela leitura de um texto de contextualização da tarefa, contendo algumas instruções iniciais, incluindo como iniciar a interação (ver Figura B.2). O comando de início é enviado pela experimentadora, que assiste à interação sem que isso seja informado a quem participa.

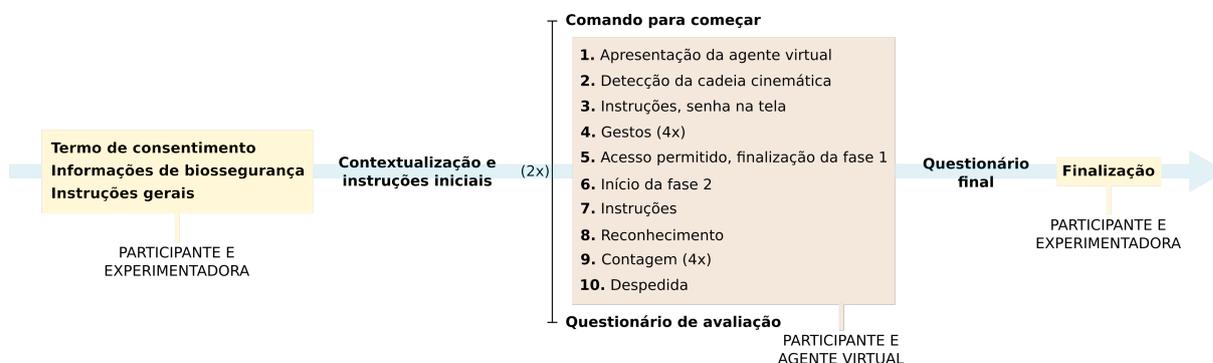


Figure B.1: Fluxo de participação no experimento de interação.

A interação começa com a primeira agente virtual sendo exibida em uma tela, se

¹Os experimentos foram realizados entre julho e dezembro de 2021.



Figure B.2: Contextualização da interação e instruções iniciais.

apresentando e pedindo que a pessoa faça a pose mostrada (*“Antes de tudo, eu preciso encontrar você. Por favor, se posicione acima do X no chão, olhe para a câmera 1 e faça a pose indicada.”*), necessária para o início do rastreamento de sua cadeia cinemática. Após a detecção, a tela ao lado da agente virtual passa a exibir a aplicação relacionada à fase 1 (senha de cores) e a agente virtual dá as instruções da fase. A agente virtual mostra a senha de cores na tela e o estado da tarefa é atualizado à medida que as cores são indicadas através dos gestos de apontar. Quando a senha de cores é preenhida, a agente virtual pede que a pessoa se direcione para uma posição indicada, próxima ao computador em que a fase 2 (contagem de imagens) é realizada.

Ao iniciar a fase 2, a agente virtual passa a ser exibida junto com a aplicação em tela da segunda fase e começa dando as instruções. Em seguida, a agente virtual pede que a pessoa olhe diretamente para a câmera indicada (*“Antes de começarmos, preciso checar sua identidade. Olhe diretamente para a câmera 2 até que eu consiga te reconhecer, por favor.”*), para que os pontos de referência na imagem do rosto para a comunicação implícita através da direção do olhar da pessoa sejam obtidos. Após essa etapa, os botões na tela são habilitados e a contagem dos objetos e o preenchimento dos valores podem começar. Após o término da fase 2, a agente virtual finaliza, agradece e para de ser exibida. Uma mensagem em tela dá instruções para o preenchimento do formulário de avaliação da agente virtual (ver Seção 4.3). A Figura B.3 ilustra o fluxo de telas durante a interação.

Esse processo é repetido duas vezes, com as duas agentes virtuais, cada uma com uma configuração de comunicação, EX e EXIM. Exceto por algumas falas que são específicas da ordem das configurações (instruções mais longas, por exemplo, são dadas apenas na primeira configuração) e pelas dicas dos valores na fase 2, as frases usadas pelas agentes virtuais são levemente diferentes, para ajudar na diferenciação entre elas. Por exemplo, Luna se apresenta dizendo *“Olá, meu nome é Luna.”*, enquanto Sofia diz *“Oi, eu sou a Sofia.”*. Após o preenchimento do questionário sobre a segunda agente virtual, a experimentadora é chamada de volta à sala e indica o questionário final a ser preenchido (ver Seção 4.3). Participantes respondem o questionário final enquanto a experimentadora espera do lado de fora da sala e, ao terminarem, a participação é finalizada.

A sequência de cores da fase 1 é definida aleatoriamente pelo sistema para cada repetição com cada participante. São quatro cores disponíveis (azul, amarelo, branco e preto) e a única restrição é que uma cor não seja usada duas vezes sendo uma seguida da outra. Caixas das mesmas quatro cores disponíveis são os objetos que devem ser apontados para o preenchimento da senha e estão dispostas no ambiente de interação, como mostrado na Figura B.4. Marcações no chão indicam posições para onde participantes devem se direcionar em determinados momentos da interação.

Para a fase 2, quatro imagens de objetos embaralhados são fixadas no ambiente. A Figura B.5, que ilustra a visão da pessoa durante a fase 2, mostra a tela em que a aplicação e a agente virtual são exibidas durante a fase, a câmera utilizada pelo sistema

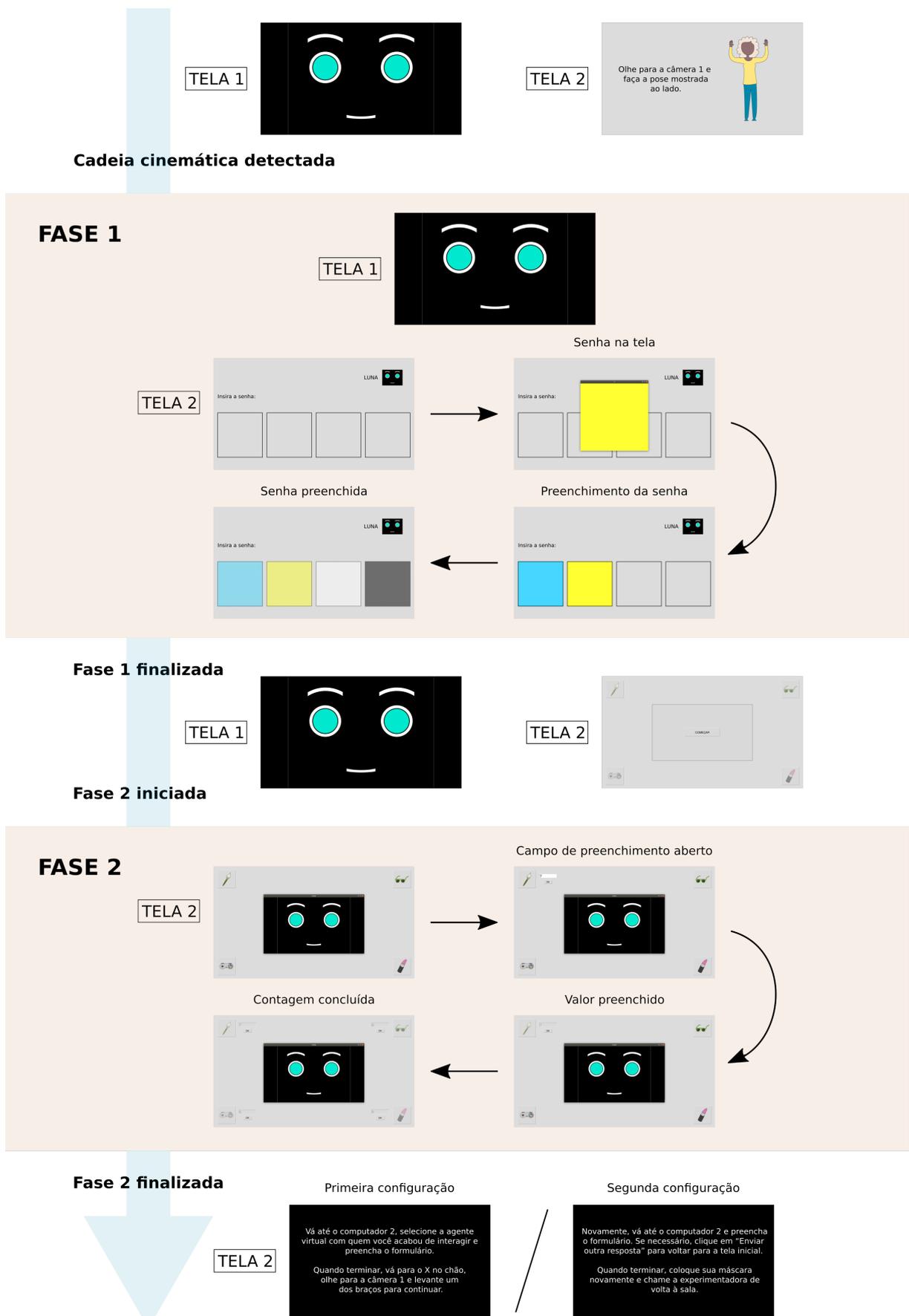


Figure B.3: Fluxo de telas durante a interação.



Figure B.4: Ambiente dos experimentos de interação. As quatro caixas coloridas são os objetos a serem indicados e marcações no chão indicam posições para onde participantes devem se direcionar.



Figure B.5: Visão de participantes durante a fase 2 e as quatro imagens para contagem fixadas no ambiente. As imagens são posicionadas de modo a se relacionarem com os botões mostrados em tela durante a interação.

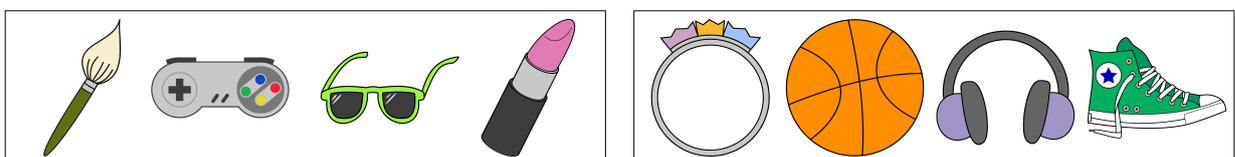


Figure B.6: Conjuntos de objetos para contagem.

de rastreamento da direção do olhar (câmera 2) e a câmera RGB-D que rastreia a cadeia cinemática do humano (câmera 1). As imagens para contagem são fixadas de modo a se relacionarem com botões mostrados na tela durante a interação, que indicam os objetos que devem ser contados. Por exemplo, o canto direito superior da tela mostra o objeto que deve ser contado na imagem posicionada mais acima e à direita da pessoa. As imagens para contagem também são mostradas na Figura B.5, de maneira similar a como são posicionadas no ambiente, e a Figura B.6 mostra os dois conjuntos de objetos que são exibidos nos botões em tela em cada repetição da fase. A ordem dos conjuntos de objetos entre as duas configurações de comunicação é definida aleatoriamente pelo sistema.

C

DISTRIBUIÇÕES DE PROBABILIDADE E HISTOGRAMAS DA ESTIMAÇÃO DE PARÂMETROS BAYESIANA

A seguir são mostradas figuras com os resultados da estimação de parâmetros Bayesianas para as medidas de sociabilidade e transparência das agentes virtuais e a percepção de eficiência das interações nas duas configurações de comunicação, EX e EXIM. O procedimento para interpretação das figuras é discutido na Seção 5.2. As Figuras C.1 a C.3 apresentam os resultados para a variável de sociabilidade, as Figuras C.4 a C.6 são sobre a transparência das agentes virtuais e as Figuras C.7 a C.9 mostram a estimação Bayesianas dos parâmetros de percepção de eficiência das interações.

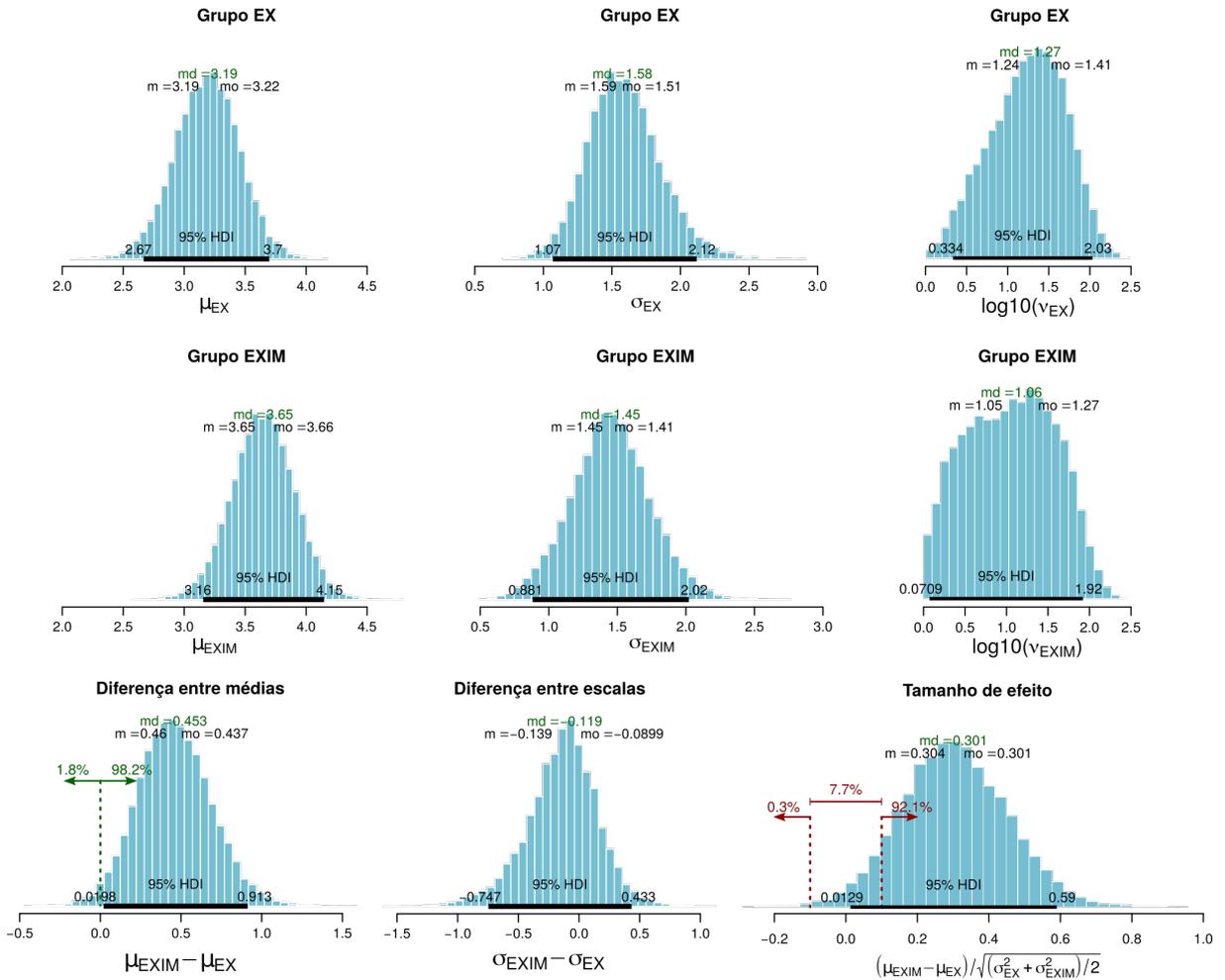


Figure C.1: Distribuições *a posteriori* dos parâmetros μ , σ e ν (escala logarítmica) das distribuições *t* latentes de sociabilidade das agentes virtuais nas configurações de comunicação EX e EXIM. A última linha mostra as distribuições *a posteriori* das diferenças $\mu_{EXIM} - \mu_{EX}$ e $\sigma_{EXIM} - \sigma_{EX}$ e do tamanho de efeito. Para todas elas, estão anotadas média (m), mediana (md), moda (mo) e os limites do 95% HDI. Nas distribuições de $\mu_{EXIM} - \mu_{EX}$ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo ($\mu_{EXIM} - \mu_{EX} = 0$) e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima do valor nulo e abaixo, dentro e acima da ROPE mostradas.

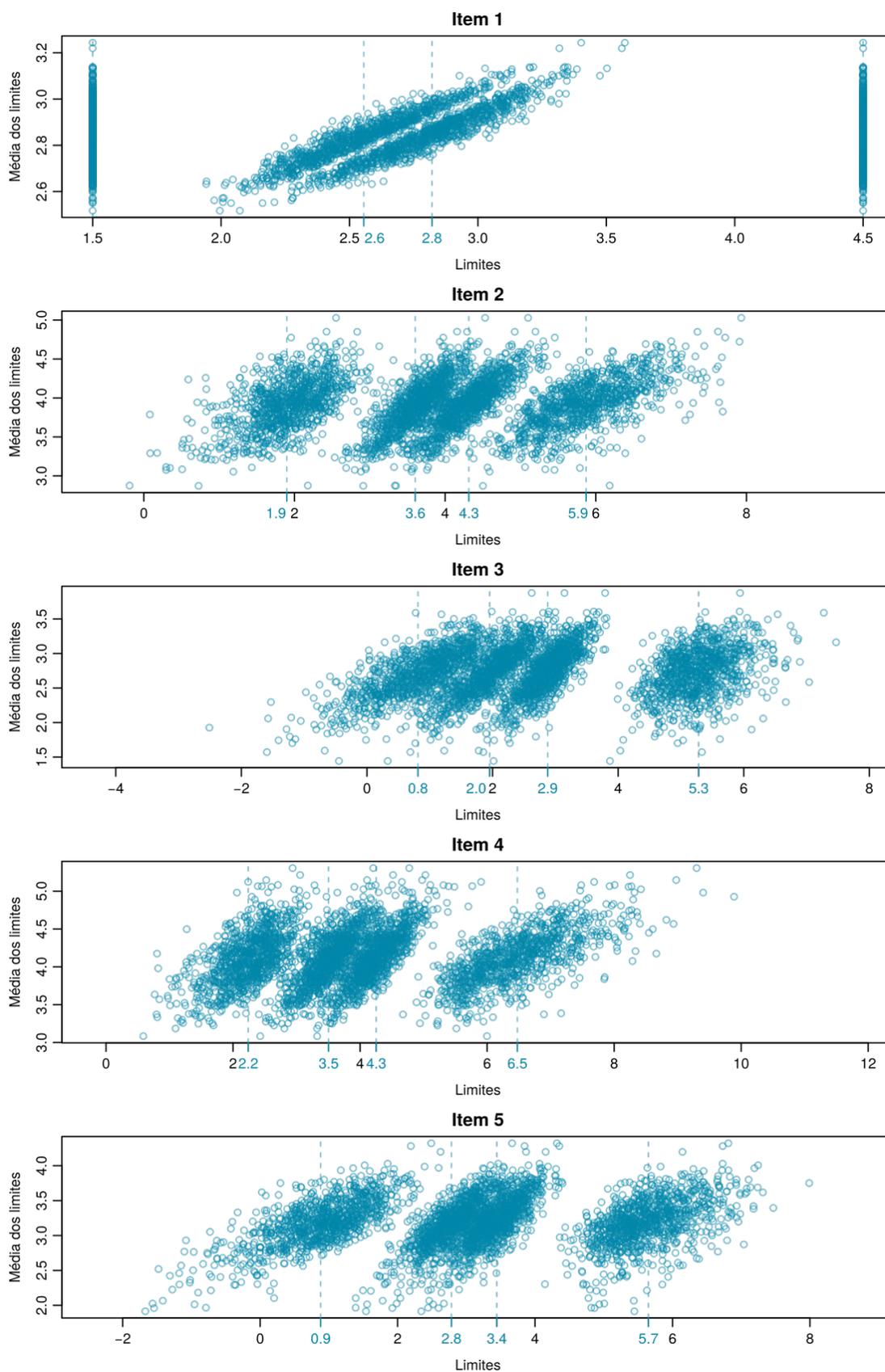


Figure C.2: Distribuições *a posteriori* dos limites relacionados a cada item da escala Likert de sociabilidade das agentes virtuais. Linhas tracejadas indicam as médias das estimativas dos limites.

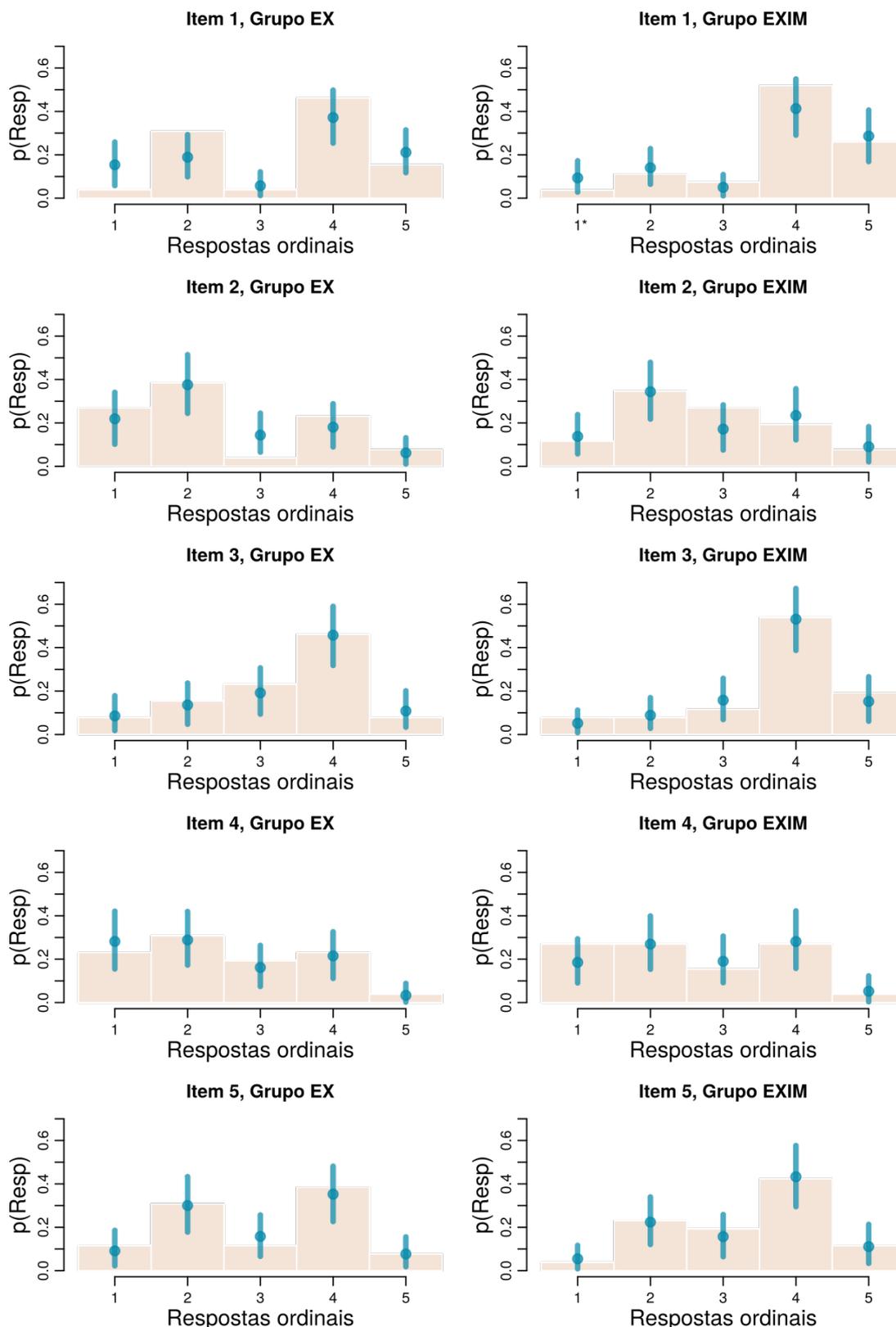


Figure C.3: Histogramas dos dados de sociabilidade sobrepostos com probabilidades estimadas para checagem posterior de adequação do modelo. Níveis originalmente vazios, para os quais se adicionou uma resposta extra, são indicados por um asterisco (*).

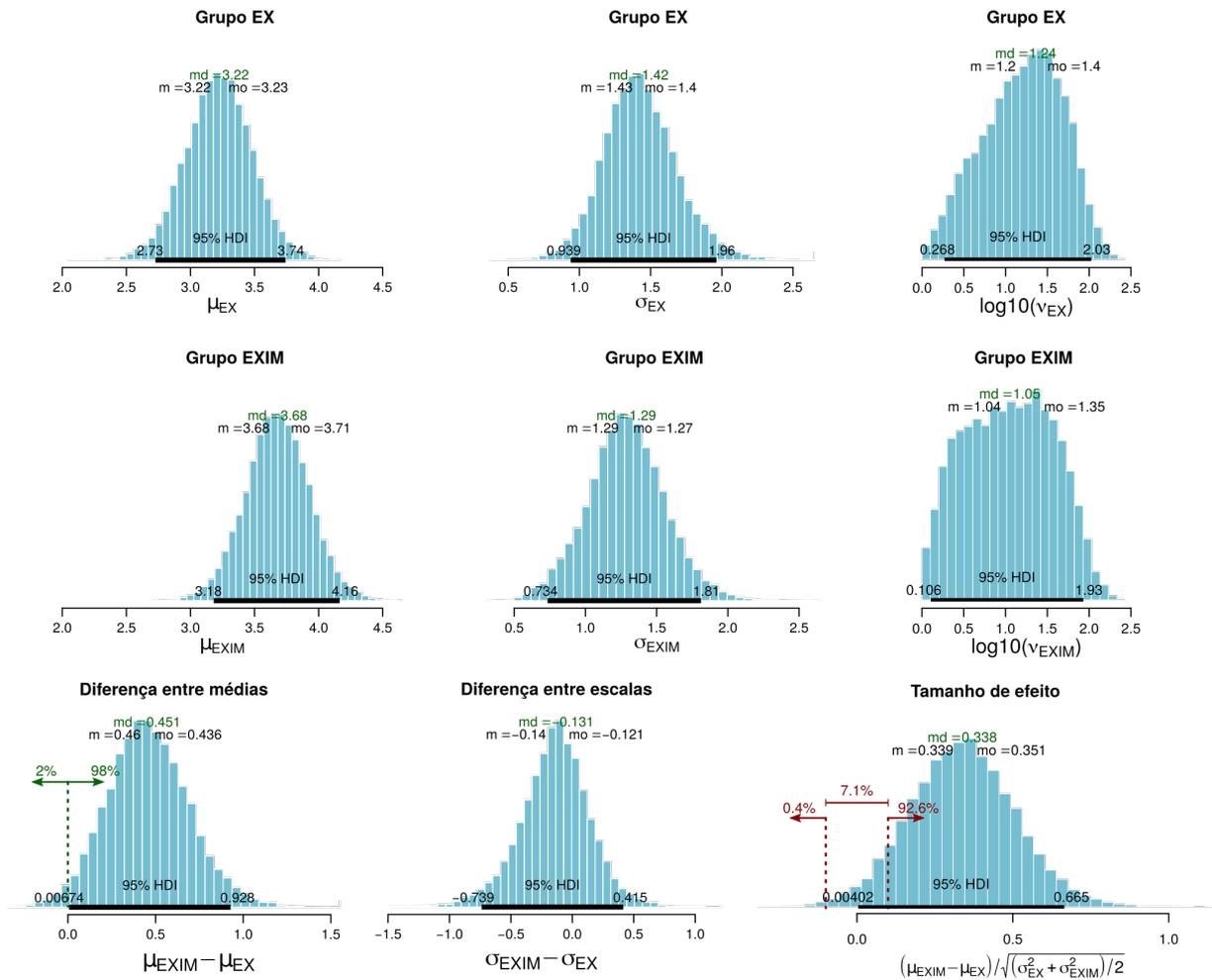


Figure C.4: Distribuições *a posteriori* dos parâmetros μ , σ e ν (escala logarítmica) das distribuições *t* latentes de transparência das agentes virtuais nas configurações de comunicação EX e EXIM. A última linha mostra as distribuições *a posteriori* das diferenças $\mu_{EXIM} - \mu_{EX}$ e $\sigma_{EXIM} - \sigma_{EX}$ e do tamanho de efeito. Para todas elas, estão anotadas média (m), mediana (md), moda (mo) e os limites do 95% HDI. Nas distribuições de $\mu_{EXIM} - \mu_{EX}$ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo ($\mu_{EXIM} - \mu_{EX} = 0$) e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima do valor nulo e abaixo, dentro e acima da ROPE mostradas.

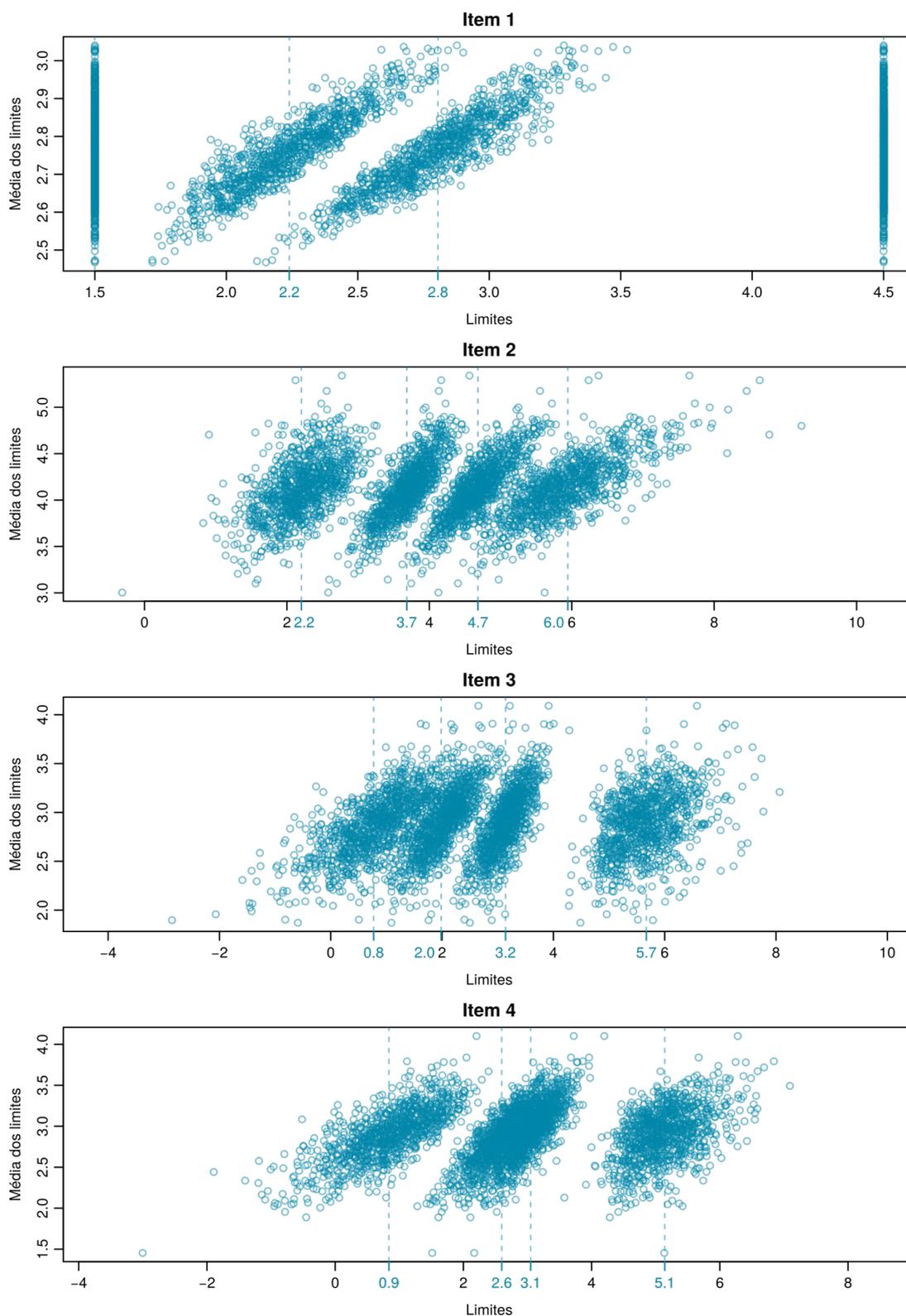


Figure C.5: Distribuições *a posteriori* dos limites relacionados a cada item da escala Likert de transparência das agentes virtuais. Linhas tracejadas indicam as médias das estimativas dos limites.

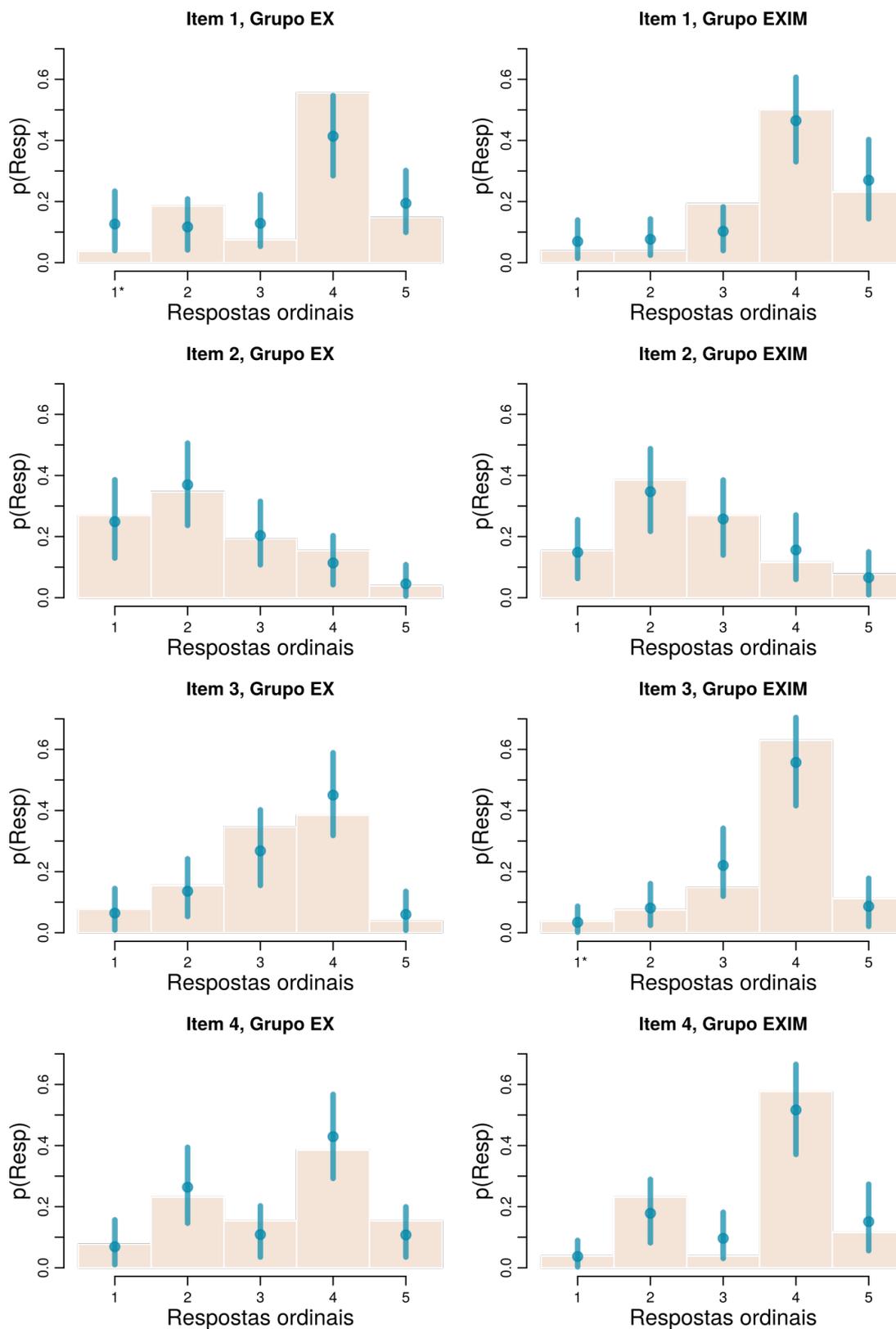


Figure C.6: Histogramas dos dados de transparência sobrepostos com probabilidades estimadas para checagem posterior de adequação do modelo. Níveis originalmente vazios, para os quais se adicionou uma resposta extra, são indicados por um asterisco (*).

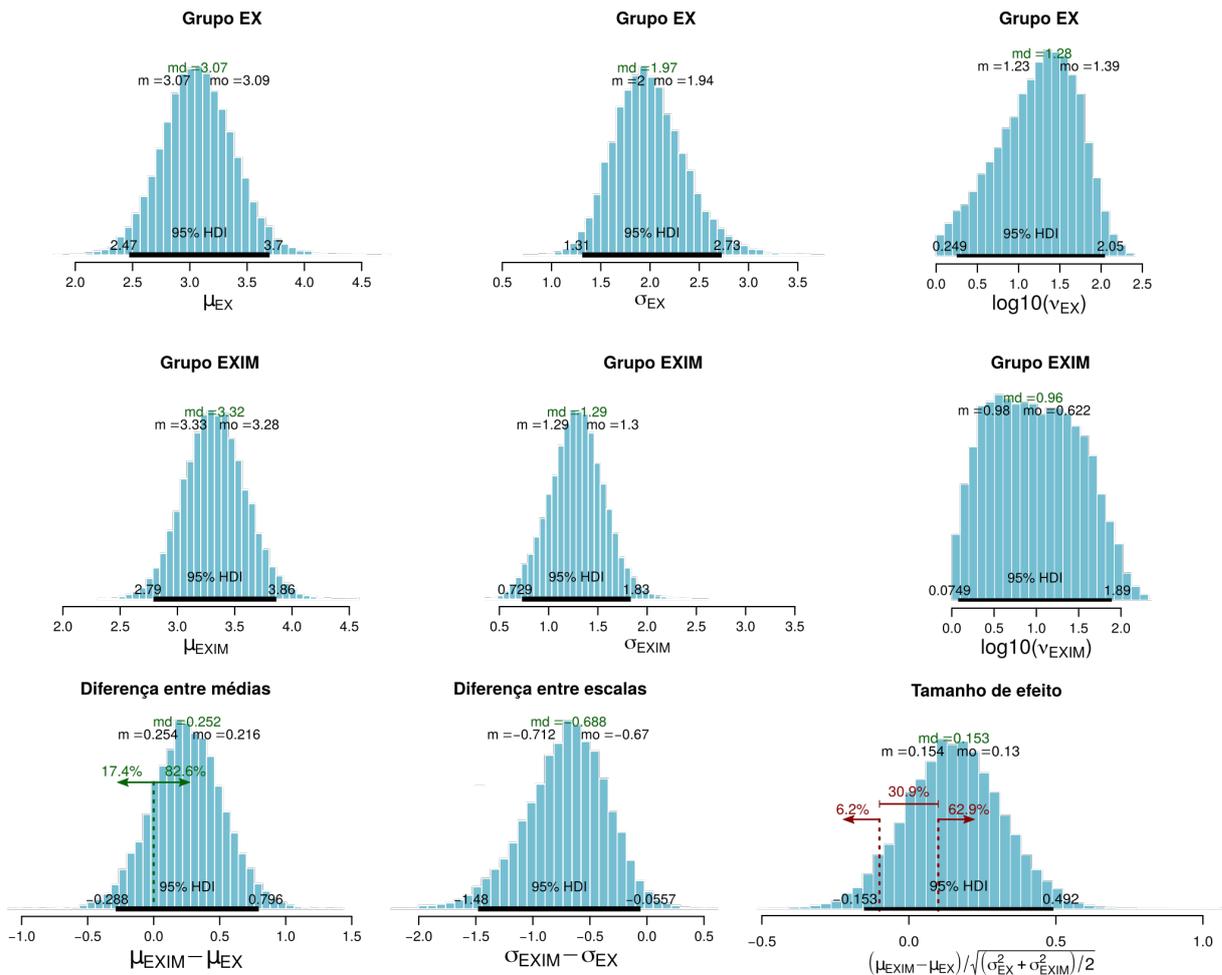


Figure C.7: Distribuições *a posteriori* dos parâmetros μ , σ e ν (escala logarítmica) das distribuições *t* latentes de percepção de eficiência das interações nas configurações de comunicação EX e EXIM. A última linha mostra as distribuições *a posteriori* das diferenças $\mu_{EXIM} - \mu_{EX}$ e $\sigma_{EXIM} - \sigma_{EX}$ e do tamanho de efeito. Para todas elas, estão anotadas média (m), mediana (md), moda (mo) e os limites do 95% HDI. Nas distribuições de $\mu_{EXIM} - \mu_{EX}$ e do tamanho de efeito, linhas verticais pontilhadas indicam, respectivamente, o valor nulo ($\mu_{EXIM} - \mu_{EX} = 0$) e a ROPE de $-0,1$ a $0,1$, com as porcentagens da distribuição abaixo e acima do valor nulo e abaixo, dentro e acima da ROPE mostradas.

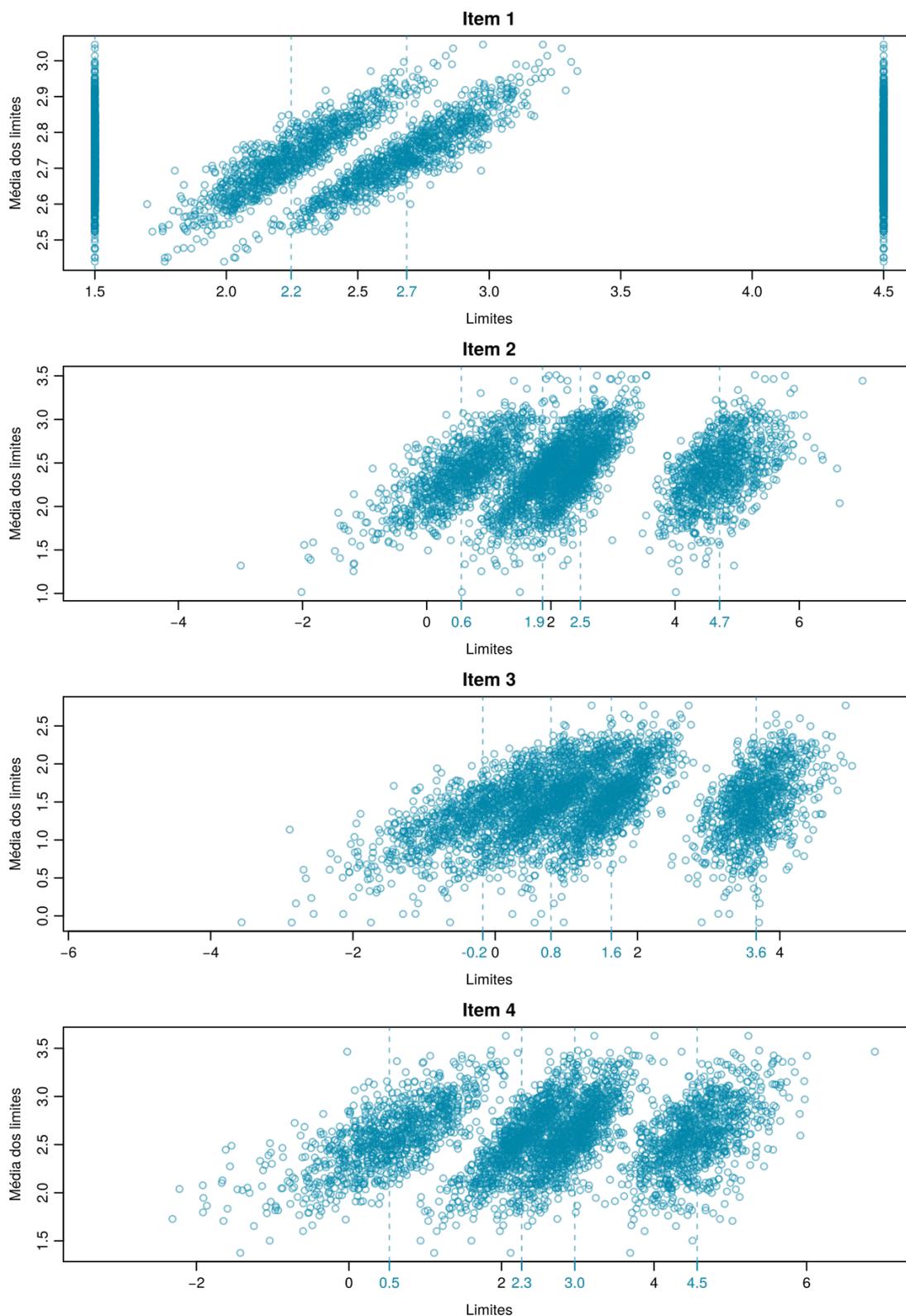


Figure C.8: Distribuições *a posteriori* dos limites relacionados a cada item da escala Likert de percepção de eficiência das interações. Linhas tracejadas indicam as médias das estimativas dos limites.

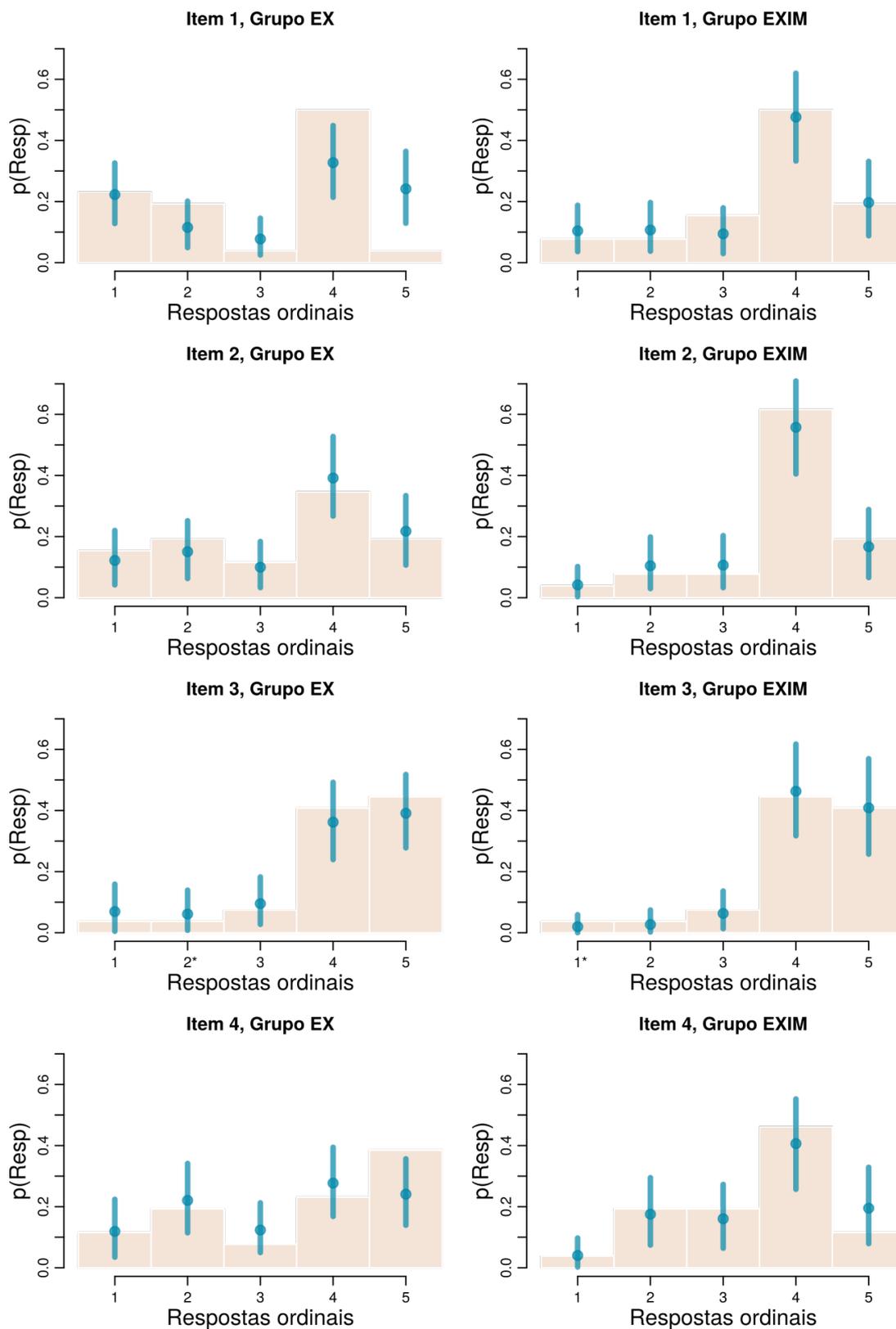


Figure C.9: Histogramas dos dados de percepção de eficiência sobrepostos com probabilidades estimadas para checagem posterior de adequação do modelo. Níveis originalmente vazios, para os quais se adicionou uma resposta extra, são indicados por um asterisco (*).