



Programa Interdisciplinar de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Edward Simon John

Interação Multimodal Adaptativa Embarcada em Robótica Assistiva
para Comunicação com Pessoas com Deficiência

São Leopoldo
2016

Edward Simon John

INTERAÇÃO MULTIMODAL ADAPTATIVA EMBARCADA EM ROBÓTICA
ASSISTIVA PARA COMUNICAÇÃO COM PESSOAS COM DEFICIÊNCIA

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa Interdisciplinar de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS

Orientador: Prof. Dr. Sandro José Rigo

Coorientador: Prof. Dr. Jorge Barbosa

São Leopoldo
2016

J65i

John, Edward Simon.

Interação multimodal adaptativa embarcada em robótica assistiva para comunicação com pessoas com deficiência / Edward Simon John. – 2016.

95 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2016.

“Orientador: Prof. Dr. Sandro José Rigo.”

“Coorientador: Prof. Dr. Jorge Barbosa.”

1. Robótica. 2. Interação homem-máquina. 3. Interfaces de usuários (sistemas de computação). 4. Pessoas com deficiência. I. Rigo, Sandro José. II. Barbosa, Jorge. III. Título.

CDU 004.896-056.26

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Raquel Herbcz França – CRB 10/1795)

Edward Simon John

Interação Multimodal Adaptativa Embarcada em Robótica Assistiva para
Comunicação com Pessoas com Deficiência

Dissertação apresentada à Universidade do Vale do
Rio dos Sinos – UNISINOS, como requisito parcial para
obtenção do título de Mestre em Computação Aplicada

Aprovado em 25/02/2016

BANCA EXAMINADORA

Prof. Dra. Liliana Maria Passerino – Universidade Federal do Rio Grande do Sul

Prof. Dr. João Carlos Gluz - Universidade do Vale do Rio dos Sinos

Prof. Dr. Sandro José Rigo (Orientador) - Universidade do Vale do Rio dos Sinos

Prof. Dr. Jorge Luis Victória Barbosa (Coorientador) - Universidade do Vale do Rio
dos Sinos

São Leopoldo,

Prof. Dr. Sandro José Rigo
Coordenador do PPG em Computação Aplicada

AGRADECIMENTOS

Inicialmente, meus agradecimentos são dirigidos à minha família, pelo suporte e compreensão. À minha namorada, pela sua contribuição na elaboração do protótipo e realização dos testes com usuários reais.

Agradeço aos professores do PIPCA, em especial ao meu orientador Prof. Dr. Sandro Rigo e ao coorientador Prof. Dr. Jorge Barbosa, pelos ensinamentos e apoio durante esses anos.

À professora Ariadne Leal e à diretora Thaís Monteiro, por terem oportunizado os testes com alunos da Escola de Surdos Bilíngue Salomão Watnik. À FENEIS, na pessoa de Alexandra Boeira, pelos contatos com as entidades representativas e de educação de pessoas com deficiência auditiva, e ao AFT Rafael Giger, pelos contatos com as associações de pessoas com deficiência visual.

Ao Adilso Corlassoli, presidente da União de Cegos do Rio Grande do Sul, por ter-nos oferecido o espaço para teste com o público com deficiência visual.

RESUMO

Pessoas que enfrentam dificuldades de comunicação em suas atividades diárias, por apresentarem necessidades especiais para ver ou ouvir, são uma significativa parcela da população brasileira. Técnicas de robótica assistiva vêm sendo desenvolvidas para apoiar este público em determinadas atividades, com bons resultados sendo observados. Entretanto os dispositivos de robótica assistiva disponíveis para comunicação são, em geral, direcionados ao atendimento de um público específico e não consideram o atendimento das necessidades de um público misto, composto por pessoas com diferentes necessidades. Embora existam modelos versáteis quanto a canais de entrada e saída, não se observa a implementação dos recursos para atendimento de um amplo conjunto de necessidades.

Dado esse contexto, este trabalho propõe um modelo de sistema de Interação Multimodal Humano Computador, embarcado em um robô assistivo, capaz de adaptar o conjunto de meios de interação necessários para a comunicação de acordo com o tipo e o grau de necessidades do usuário, fornecendo informações relevantes por meio dos canais adequados de acordo com cada deficiência. O modelo também inclui um conjunto de recursos destinados a obter informações em fontes diversas e disponibilizar estes resultados aos usuários.

A abordagem proposta enfatiza o gerenciamento de usuários e a adaptação do uso dos canais de interação às suas características. Foi desenvolvido um protótipo com um conjunto de serviços ambientados no contexto de um campus universitário, com capacidade de atender à dúvidas e necessidades sobre este ambiente. Os testes realizados com este protótipo envolveram um grupo de 14 (quatorze) usuários com deficiência e suscitaram atitudes positivas quanto a aspectos de usabilidade como facilidade de uso, utilidade, conforto e eficácia.

Palavras-Chave: Robótica Assistiva. Sistema de Interação Multimodal Adaptativa. Gerenciamento de usuário. Pessoas com deficiência.

ABSTRACT

People facing communication difficulties in their daily activities, because they have special needs to see or hear, are a significant portion of the Brazilian population. Robotic assistive techniques have been developed to support this public in certain activities, with good results being observed. However, assistive robotic devices available for communication are generally oriented to give heed to a specific audience and do not meet the needs of a mixed audience, composed of people with different needs. While there are versatile models in terms of input and output channels, the implementation of the resources in order to attend a wide range of needs is not seen.

Given this context, this work proposes a model of Multimodal Human-Computer Interaction System, embedded in an assistive robot, able to adapt the set of interaction means for communication according to the type and degree of user needs, providing relevant information through appropriate channels according to each disability. The model also includes a set of features to obtain information from various sources and provide these results to users.

The proposed approach emphasizes user management and adaptation of the use of interaction channels to their characteristics. A prototype with a set of services in the context of a university campus and capable of meeting the questions and needs of this environment was developed. The experiments with this prototype involved a group of 14 (fourteen) handicapped users and elicited positive attitudes towards usability aspects as ease of use, utility, comfort and efficiency.

Keywords: Assistive Robotics. Adaptive Multimodal Interactive System. User management. People with disability.

LISTA DE FIGURAS

Figura 1: Distribuição das deficiências na população brasileira.	8
Figura 2: Discrepância entre a quantidade de pessoas com deficiência matriculadas no ensino superior e sua presença na população em geral.	9
Figura 3: Diagrama exemplificativo de interação multimodal	14
Figura 4: Principais blocos de um sistema ASR	19
Figura 5: Modelo genérico de reconhecimento de fala.....	19
Figura 6: Coruja – Sistema de reconhecimento de fala em português-brasileiro	20
Figura 7: Care-O-bot® 4 – um robô de serviço.....	25
Figura 8: Protótipo ALIAS	26
Figura 9: ALIAS – módulos do sistema	26
Figura 10: Protótipo Hobbit, sem e com carcaça.....	27
Figura 11: Hobbit – arquitetura do sistema para observação humana.	28
Figura 12: Arquitetura global do sistema multimodal	29
Figura 13: Solução integrada – robô industrial com interação multimodal adaptativa.....	30
Figura 14: Representação esquemática do sistema integrado.....	30
Figura 15: Arquitetura do sistema	31
Figura 16: Protótipo tutor funcional.....	32
Figura 17: <i>Speech Act Theory</i> – fusão de modalidades de entrada	33
Figura 18: Arquitetura de um sistema multimodal de diálogo para um robô de companhia.	34
Figura 19: Arquitetura do sistema de diálogo.....	36
Figura 20: Arquitetura do modelo proposto	41
Figura 21: Ontologia para acessibilidade do Hefestos	43
Figura 22: Exemplo de documento EMMA gerado por uma interação de voz.....	45
Figura 23: Protótipo funcional.....	49

Figura 24: Imagem da interface gráfica disposta no display <i>touchscreen</i>	52
Figura 25: Documento EMMA de saída	52
Figura 26: Representação gráfica da estrutura de interações	54
Figura 27: XML com a definição da árvore de interações.....	54
Figura 28: Diálogo entre robô e usuário.....	56
Figura 29: Sessão de usuário com deficiência visual completa.....	58

LISTA DE TABELAS

Tabela 1: Comparação dos sistemas multimodais contidos nos trabalhos relacionados.	38
Tabela 2: Exemplo de representação de um serviço de verificação de e-mails.	46
Tabela 3: Modalidades de entradas ativadas conforme perfil de usuário (tipo e grau de deficiência).....	53
Tabela 4: Modalidades de saídas selecionadas conforme o perfil de usuário (tipo e grau de deficiência).....	55
Tabela 5: Tabela de serviços implementados.....	56
Tabela 6: Distribuição das respostas dos participantes do estudo.....	62
Tabela 7: Respostas dos participantes do estudo distribuídas por tema.	63

LISTA DE GRÁFICOS

Gráfico 1: Distribuição das respostas de todos os participantes por eixo temático....	63
Gráfico 2: Distribuição das respostas dos deficientes visuais por eixo temático	65
Gráfico 3: Avaliação da percepção de utilidade pelos participantes com deficiência visual.	65
Gráfico 4: Distribuição das respostas dos deficientes auditivos por eixo temático.....	66
Gráfico 5: Avaliação da facilidade de uso pelos deficientes auditivos.....	67
Gráfico 6: Avaliação da percepção de utilidade pelos deficientes auditivos.....	67
Gráfico 7: Comparação das respostas oferecidas pelos deficientes que adotam LIBRAS e pelos demais participantes do estudo	68
Gráfico 8: Distribuição das respostas dos deficientes parciais (visuais e auditivo) por eixo temático	69
Gráfico 9: Avaliação da percepção de utilidade pelos deficientes parciais.....	69

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
PIPCA	Programa Interdisciplinar de Pós-Graduação em Computação Aplicada
UNISINOS	Universidade do Vale do Rio dos Sinos
RA	Robótica Assistiva
IHC	Interação Humano-Computador
BCI	Brain Computer Interface
IMMHC	Interação Multimodal Humano-Computador
RAF	Reconhecimento Automático de Fala
ASR	Automatic Speech Recognition
HMM	Hidden Markov Models
FRT	Face Recognition Techniques
GUI	Graphical User Interface
AAL	AmbientAssisted Living
TTS	Text to Speech
EBGM	Elastic Bunch Graph Matching
PDBNN	Probabilistic decision-based neural network
WIMP	Windows, Icons, Menus and Pointer
GRI	Gesture Recognition Interface
CTSPN	Continuous-Time Stochastic Petri Net
OpenCV	Open Source Computer Vision
IDE	Integrated Development Environment
API	Application Programming Interface

SUMÁRIO

1 INTRODUÇÃO	8
1.1 MOTIVAÇÃO	11
1.2 QUESTÃO DE PESQUISA	11
1.3 OBJETIVOS	12
1.4 METODOLOGIA	12
1.5 ORGANIZAÇÃO DO DOCUMENTO.....	13
2 EMBASAMENTO TEÓRICO	14
2.1 INTERAÇÃO MULTIMODAL HUMANO-COMPUTADOR.....	14
2.2 ROBÓTICA ASSISTIVA.....	16
2.3 ADAPTATIVIDADE.....	17
2.4 RECONHECIMENTO DE FALA.....	18
2.5 SÍNTESE DE VOZ: GÊNERO.....	21
2.6 RECONHECIMENTO DE FACE	22
2.7 RECONHECIMENTO DE GESTOS.....	23
3 TRABALHOS RELACIONADOS	24
3.1 CARE-O-BOT®	24
3.2 ALIAS	25
3.3 HOBBIT.....	27
3.4 MULTIMODAL ARCHITECTURE TO STRENGTHEN THE INTERACTION OF THE ROBOT IN AMBIENT INTELLIGENCE ENVIRONMENTS	28
3.5 ADAPTIVE MULTIMODAL INTERACTION WITH INDUSTRIAL ROBOT	29
3.6 A NEW ROBOTIC PLATFORM FOR SIGN LANGUAGE TUTORING.....	31
3.7 ADAPTATIVE INPUT INTERPRETATION FOR DIALOGUE MANAGEMENT OF AN AUTONOMOUS ROBOT 33	
3.8 OFFICEMATE – UM ESTUDO DE UM SISTEMA DE APRENDIZAGEM DE DIÁLOGO ONLINE PARA ROBÔS ASSISTIVOS MÓVEIS.....	35
3.9 SMARTKOM – ADAPTIVE AND FLEXIBLE MULTIMODAL ACCESS TO MULTIPLE APPLICATIONS	36
3.10 ANÁLISE DE TRABALHOS RELACIONADOS	37
4 MODELO PROPOSTO	40
4.1 VISÃO GERAL.....	40

4.2	MÓDULO DE GERENCIAMENTO DO USUÁRIO.....	42
4.3	MÓDULO DE PROCESSAMENTO E GERENCIAMENTO DE ENTRADAS	44
4.4	MÓDULO DE CONTROLE DA INTERAÇÃO.....	45
4.5	INTERPRETADOR DE DADOS PARA SAÍDA.....	47
5	ASPECTOS DE IMPLEMENTAÇÃO	48
5.1	ASPECTOS GERAIS.....	48
5.2	COMPONENTES EMPREGADOS	48
5.3	GERENCIAMENTO DE USUÁRIO.....	50
5.4	IMPLEMENTAÇÃO DOS COMPONENTES DE ENTRADA E SAÍDA	51
5.5	IMPLEMENTAÇÃO DO CONTROLE DA INTERAÇÃO	53
5.5.1	Interpretador de dados para saída.....	57
5.6	SESSÕES DE UTILIZAÇÃO.....	57
6	AVALIAÇÃO	59
6.1	METODOLOGIA ADOTADA.....	59
6.2	PERFIL DOS PARTICIPANTES.....	60
6.3	PROCEDIMENTOS DE TESTE	60
6.4	ANÁLISE DOS QUESTIONÁRIOS	61
6.4.1	Deficientes visuais.....	64
6.4.2	Deficientes auditivos.....	66
6.4.3	Deficientes auditivos que empregam LIBRAS	67
6.4.4	Deficientes parciais	68
7	CONCLUSÃO.....	70
7.1	LIMITAÇÕES	71
7.2	TRABALHOS FUTUROS	72
	REFERÊNCIAS.....	73
	APÊNDICE A.....	81
	APÊNDICE B.....	84

1 INTRODUÇÃO

Este trabalho propõe o desenvolvimento de um robô assistivo para pessoas com deficiências visual e auditiva, com o objetivo de facilitar a comunicação humano-computador. O modelo desenvolvido considera os aspectos de interação multimodal para que possa ser utilizado com diferentes usuários, com diferentes necessidades. É dada especial ênfase ao aspecto adaptativo do sistema de interação multimodal, a fim de permitir a seleção de canais de entrada e saída mais adequada à natureza e severidade das deficiências do usuário. Dessa forma, a multimodalidade adaptativa desempenha um papel central na aceitabilidade e usabilidade do robô.

Um importante aspecto da multimodalidade é o uso de múltiplos canais na comunicação entre parceiros de interação, o que, do ponto de vista do robô, são canais de entrada e saída (STEFFEN, 2011). Modalidades de entrada podem ser, por exemplo, fala, gestos de cabeça ou mão, expressões faciais e também toques a partir de uma interface gráfica de usuário (no inglês, *Graphical User Interface* – GUI). Para a saída, fala e outros áudios podem ser usados, assim como saídas em tela, gestos do robô ou mímicas.

A adaptatividade é um dos principais aspectos de um sistema intuitivo. Um sistema adaptativo deve ser capaz de ajustar-se às capacidades e preferências do usuário, a fim de melhorar a comunicação. (STEFFEN, 2011).

O decreto nº 5.296 (BRASIL, 2004) define a pessoa com deficiência como sendo aquela que possui limitação ou incapacidade para o desempenho de atividade, devendo se enquadrar em categorias normativamente definidas. Para fins deste trabalho, importa saber que as deficiências são classificadas em física, auditiva, visual, mental e múltipla (associação de duas ou mais deficiências). O público-alvo deste estudo é formado por pessoas com deficiência visual ou auditiva.

Figura 1: Distribuição das deficiências na população brasileira.



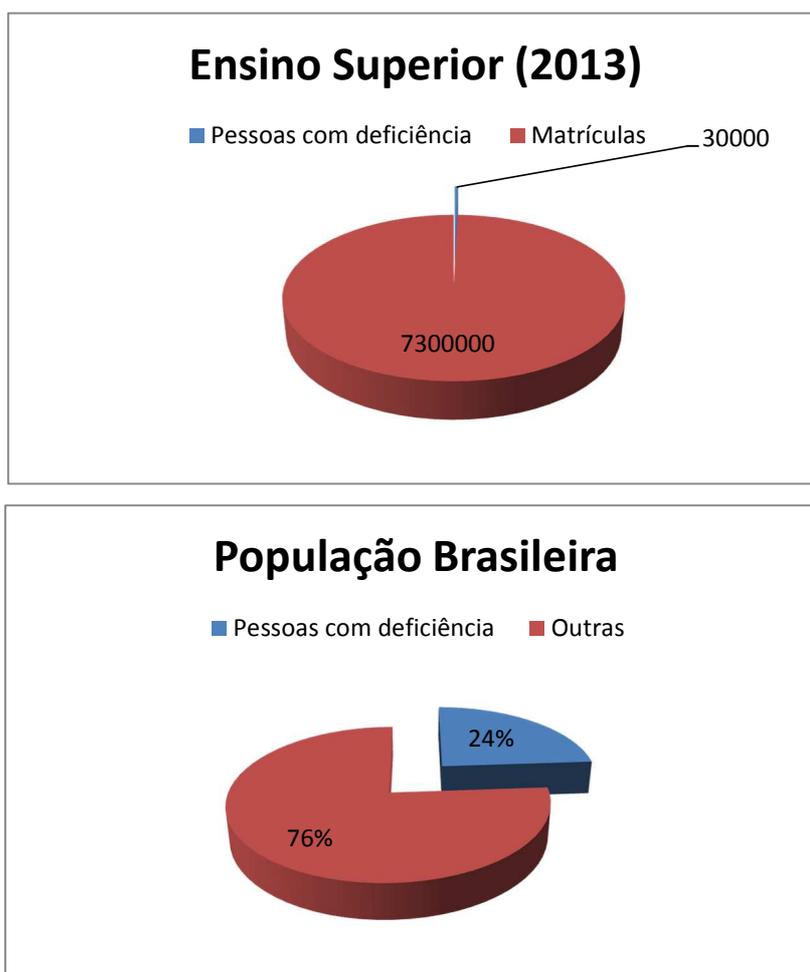
Fonte: Censo Demográfico 2010 (IBGE, 2012)

A metodologia empregada no Censo 2010 para aferição do número de pessoas com deficiência não reflete a conceituação legal acima apontada. Segundo Lenzi (2012), o método do Washington Group, adotado no último censo, permite a identificação de graus de deficiência. Por esse motivo, o Censo de 2010 apresenta resultados significativamente maiores que o Censo de 2000, no qual o enquadramento se dava em razão da percepção subjetiva de incapacidade por parte do entrevistado. De acordo com informações

demográficas do IBGE (2012), constata-se que: a) a deficiência visual é a mais frequente - fig. 1; b) 23,4% possuem pelo menos uma das deficiências investigadas; c) 3,44% e 1,1% dos brasileiros não consegue de modo algum ou possui grande dificuldade para, respectivamente, ver e ouvir; d) 15,3% e 4% da população possui alguma dificuldade, respectivamente, para ver ou ouvir, mesmo com o auxílio de correções como óculos e aparelhos auditivos. Assim, consideradas as dificuldades e impossibilidades para ver e ouvir, é representativo o grupo de pessoas que poderia se beneficiar com a contribuição científica deste trabalho.

As pessoas com deficiência não participam do meio universitário na mesma proporção em que aparecem na população brasileira. De acordo com o Censo da Educação Superior de 2013 (INEP, 2014), das 7,3 milhões matrículas de brasileiros no Ensino Superior no nível de graduação, cerca de 30 mil correspondem a pessoas com deficiência. Segundo o censo de 2010 (IJJJ, 2012), 23,9% da população brasileira apresenta deficiência em algum grau, o que permite considerar que estão sub-representadas nos bancos universitários. A figura 2 destaca visualmente esta diferença entre as proporções dos dois públicos.

Figura 2: Discrepância entre a quantidade de pessoas com deficiência matriculadas no ensino superior e sua presença na população em geral.



Fonte: INEP/MEC e Censo Demográfico/IBGE

Tal situação, contudo, apresenta tendência de melhora, já que houve um incremento de 50% de matrículas desse grupo desde 2010, quando havia pouco mais de 19 mil pessoas com deficiência matriculadas no ensino superior. A aplicação implementada neste trabalho

visa oferecimento de serviços ao público universitário por meio da consulta de informações acadêmicas, informações pertinentes à biblioteca e à localização no campus universitário da Universidade do Vale do Rio dos Sinos.

A plena e efetiva participação e inclusão de pessoas com deficiência na sociedade é um princípio que orienta a Convenção sobre os Direitos das Pessoas com Deficiência (BRASIL, 2009) incorporada à Constituição brasileira. Por esse motivo há necessidade de realização de pesquisa de tecnologias assistivas que priorizem a facilitação, às pessoas com deficiência, do acesso a tecnologias, dispositivos e ajudas técnicas¹ de qualidade. Este contexto geral é um dos motivadores deste trabalho, pois considera-se que o desenvolvimento de um modelo de interação multimodal adaptativa que seja destinado a deficientes com necessidades diversas, tais como visuais e auditivas, pode contribuir para facilitar o acesso à informações e também atender à uma lacuna identificada quanto à falta de modelos mais amplos e que atendam à diversas necessidades ao mesmo tempo. A utilização deste modelo por meio de um robô assistivo, que possa incorporar estes diversos aspectos de interação e também providenciar o acesso à serviços pode se constituir em uma contribuição relevante no conjunto de iniciativas conhecidas com objetivo de facilitar a comunicação humano-computador.

A tecnologia pode ser o meio pelo qual se viabiliza o direito de acesso dos deficientes à informação, à comunicação, à educação, entre outros, ampliando sua autonomia e melhorando sua qualidade de vida. Assim, a tecnologia assistiva pode ser entendida como todo e qualquer produto, serviço ou equipamento cujo objetivo seja melhorar a qualidade de vida de indivíduos com limitações funcionais, sejam elas físicas, sensoriais ou psicológicas (CELESTE et al., 2014). Já a robótica assistiva, segundo Feil-Seifer e Mataric (2005), é aquela que fornece ajuda ou apoio a um usuário humano a partir do uso robôs de reabilitação, cadeiras de roda robóticas e outras formas de ajuda para ampliar a mobilidade, tais como robôs de companhia, robôs educacionais e braços mecânicos para aqueles usuários possuem limitações físicas.

A comunicação humana caracteriza-se pela multimodalidade, uma vez que a compreensão dos conteúdos se dá após análise de informações veiculadas simultaneamente por diversos canais. Olhar, gestos, movimentos, entonação, ritmo: todo esse conjunto é considerado no processo comunicativo. De acordo com Varona, Manresa-Yee e Perales (2008), enriquecer a interação humano-computador (IHC) com esses elementos do comportamento humano natural seria a principal tarefa das interfaces de usuário multimodais. Por essas razões, a criação de interfaces humano-computador não-invasivas e mais naturais, à base de reconhecimento de fala ou técnicas de visão de computador, pode oferecer uma interação com computadores mais fácil e amigável para pessoas com deficiência.

Em razão das limitações impostas aos sentidos, é imperativo que os canais de entrada e saída do sistema multimodal sejam escolhidos em consideração às necessidades e preferências de deficientes auditivos e visuais, público-alvo deste trabalho. De acordo com Akiki, Bandara e Yu (2014) técnicas de desenvolvimento de interface de usuário que se pretendam universais resultam em interfaces rígidas, do tipo “*one design fits all*”, e desprezam a variabilidade de contexto de uso. Interfaces de usuário adaptativas são consideradas uma solução para a variabilidade do contexto, em razão de sua capacidade de adaptar-se automaticamente ao contexto de uso durante a execução.

¹Ajuda técnica é qualquer elemento que facilite a autonomia pessoal ou possibilite o acesso e o uso de meio físico.

1.1 Motivação

A robótica assistiva utiliza-se da interação multimodal para comunicação com o usuário e lhes proporciona mais autonomia, ampliando sua qualidade de vida. Os robôs assistivos em desenvolvimento, entretanto, costumam focar especialmente a população idosa; entretanto, não ponderam de forma específica a gama de dificuldades que esse público pode enfrentar. Assim, não são consideradas como problemáticas para a interação as naturais dificuldades auditivas, visuais e motoras do paciente, tampouco as restrições impostas por distúrbios da fala e motores, consequências da senilidade e também comuns após acidentes vasculares cerebrais.

Quando voltada às pessoas com deficiência, a robótica assistiva é empregada em cadeiras de rodas, próteses, mãos auxiliares, etc. Neste segundo caso, as interações já são específicas para seu propósito, não se adaptando individualmente ao usuário de acordo com suas necessidades e limitações.

Na literatura examinada, a interação multimodal em robôs assistivos restringe-se a seu objetivo tradicional: a comunicação homem-máquina. Sistemas multimodais bem concebidos integrariam modalidades complementares a fim de obter uma mistura altamente sinérgica em que as forças de cada modo fossem aproveitadas e utilizadas para superar fraquezas da outra (OVIATT, 1999). A identificação dos contextos por parte do sistema seria uma maneira de conformar a interação, ofertando o canal de comunicação mais adequado à realidade e à necessidade do usuário, otimizando o contato.

Dentre os robôs assistivos pesquisados, apenas ALIAS (GOETZE; FISCHER; MORITZ, 2012) compreende a *braincomputer interface* – BCI como modalidade de entrada de dados. Por meio dela, pessoas sem capacidade de movimentação dos braços e de verbalização de comandos de voz, interagem sozinhas com o robô, graças aos eletrodos acoplados à cabeça do usuário.

Sabendo que mais de 20% da população brasileira sofre com algum tipo de incapacidade (seja ela física, auditiva, visual e/ou intelectual) e tendo em vista que a robótica assistiva ainda é destinada a públicos específicos, este trabalho tem como objetivo propor um modelo de sistema de Interação Multimodal Humano Computador (IMMHC), embarcado em um robô assistivo, capaz de se adaptar de acordo com a debilitação do usuário e de fornecer informações relevantes a partir de um conjunto de serviços flexível.

1.2 Questão de Pesquisa

Diante do problema exposto, a seguinte questão de pesquisa foi formulada:

Um sistema de interação multimodal embarcado em um robô assistivo que se adapte às necessidades de comunicação de usuários com deficiência visual ou auditiva alcançará a aceitação por parte desses usuários, considerados os principais aspectos de usabilidade (percepção de utilidade, facilidade de utilização, conforto e eficácia)?

1.3 Objetivos

O objetivo principal desta pesquisa é, partindo de estudos sobre interação multimodal adaptativa, propor um modelo que adapte a seleção de canais de interação às necessidades de usuários com deficiência auditiva e visual.

Dessa forma, estabelecem-se os seguintes objetivos específicos para esta pesquisa:

1. Investigar sistemas de interação multimodal adaptativa.
2. Implementar um protótipo de IMMHC embarcado em robô assistivo que se adapte ao usuário de acordo com sua necessidade, com base em informações relativas às suas limitações auditivas ou visuais.
3. Avaliar o protótipo sob o enfoque de usabilidade, coletando informações que colaborem para seu aperfeiçoamento em trabalhos futuros.

1.4 Metodologia

Este trabalho de pesquisa exploratória (KÖCHE, 2011) objetiva a descrição e caracterização de aspectos do problema de pesquisa definido na seção 1.2. Assim, as escolhas metodológicas decorrem do objetivo de avaliar a aceitação dos usuários ao sistema multimodal adaptativo proposto. Dada sua natureza prática, considera-se a pesquisa como sendo de natureza aplicada, uma vez que resulta na elaboração de um modelo de interação multimodal adaptativa capaz de proporcionar a comunicação humano-computador.

A fim de alcançar o embasamento teórico necessário à elaboração do modelo e à implementação do protótipo, realizou-se extensa revisão da literatura pertinente, possibilitada pelo procedimento de levantamento bibliográfico, com o qual foram coletados dados relativos à evolução dos conceitos, técnicas e ferramentas. Caracterizou-se, dessa forma, o estado da arte. Os mesmos procedimentos foram empregados a fim de identificar trabalhos relacionados, cujo escopo se assemelhasse ao desta proposta.

Os testes consistirão na interação dos usuários em cenários previamente delimitados. Aspectos qualitativos serão obtidos por meio da observação direta de cenários (WINCKLER; PIMENTA, 2002), enquanto aspectos quantitativos emergiram da avaliação de questionários e entrevistas estruturadas de usabilidade elaborados de acordo com o Modelo de Aceitação de Tecnologia, como em Costa (2013). A comunidade científica vem usando cenários para validação de sistemas sensíveis ao contexto (DEY; ABOWD; SALBER, 2001), motivo pelo qual se considera adequada a seleção de caso de uso, visto se tratar de um protótipo de multimodalidade adaptativa.

A entrevista/questionário semi-estruturada(o) aplicada aos usuários após livre experimentação do sistema possui perguntas elaboradas em consideração às categorias de análise propostas por Manzini (2004). Quanto à linguagem, cuidados acerca de adequação vocabular, clareza e precisão, uso de jargão e especificidade (não vagueza) das palavras. Já quanto à adequação das perguntas, medidas que controlem seu tamanho, que favoreçam a elaboração mental do entrevistado, que não provoquem impacto emocional, que não

permitam frases manipulativas e que possuam apenas uma única finalidade. Quanto ao roteiro, a fim de manter a atenção dos entrevistados num campo semântico, organizar as perguntas em sequência, utilizando como parâmetro um movimento no sentido das mais fáceis às mais difíceis e empregar blocos temáticos.

1.5 Organização do documento

Este trabalho está dividido em sete capítulos e contém 2 apêndices. O primeiro capítulo contém as considerações iniciais da pesquisa, incluindo justificativas, questão de pesquisa e objetivos, bem como descreve a metodologia aplicada ao trabalho. No capítulo 2, são apresentados conceitos e referências necessários à elaboração do modelo, como interação multimodal humano computador, adaptatividade, reconhecimento de fala, face e gestos e síntese de voz e texto. No capítulo 3, são descritos brevemente estudos na área de robótica assistiva e interação multimodal adaptativa, sendo destacadas lacunas existentes e a contribuição científica do presente trabalho. O capítulo 4 dedica-se à apresentação do modelo proposto e suas justificativas. O capítulo 5 descreve os aspectos de implementação, incluindo decisões quanto a métodos, técnicas e ferramentas aplicados no desenvolvimento do protótipo. O capítulo 6 é inteiramente dedicado aos testes realizados e às estratégias de coleta e análise de dados, apresentando discussão dos resultados obtidos. No capítulo 7 são apresentadas as contribuições, limitações e perspectivas de trabalhos futuros.

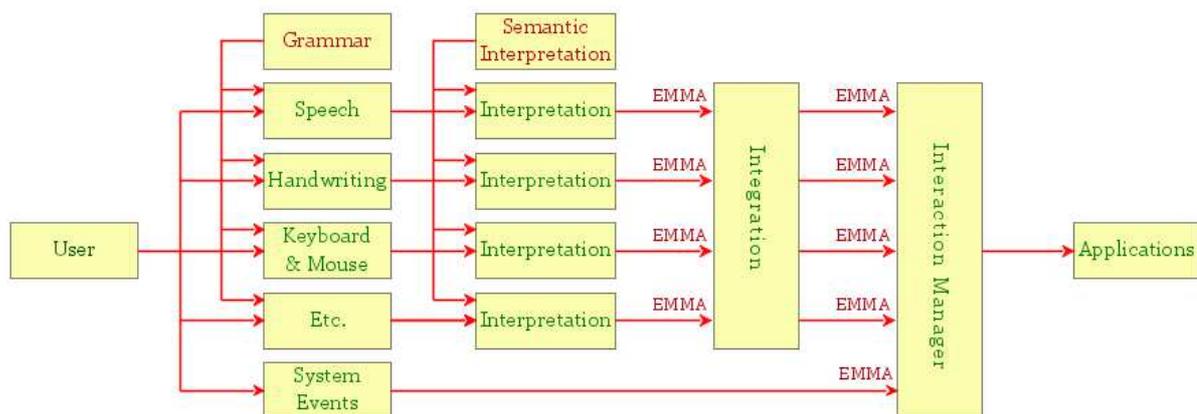
2 EMBASAMENTO TEÓRICO

Neste capítulo são apresentados elementos do embasamento teórico que foi necessário para este trabalho. São inicialmente descritas as áreas que podem ser consideradas como os fundamentos do trabalho aqui proposto, que são as áreas de Interação Multimodal Humano Computador, a Robótica Assistiva e a Adaptatividade. A partir dos recursos destas áreas, foi delimitado o escopo do modelo proposto. Além destes tópicos, são destacados aspectos de técnicas complementares utilizadas nas atividades de interação efetivamente tratadas, tais como o reconhecimento de fala e a síntese da voz, reconhecimento facial e reconhecimento de gestos.

2.1 Interação Multimodal Humano-Computador

O sistema multimodal de interação humano-computador consiste na combinação de múltiplas modalidades ou no uso de mais de um canal independente para a interação entre um usuário e a máquina (GUPTA, 2012). Para Karray e outros (2008), o termo multimodal se refere à combinação de múltiplas modalidades, sendo que nos sistemas multimodais de interação humano-computador, essas modalidades se referem principalmente às formas como esses sistemas respondem às entradas, ou seja, aos canais de comunicação. A definição desses canais é herdada dos tipos de comunicação humana, que são basicamente seus sentidos: visão, audição, tato, olfato e paladar. As possibilidades de interação com a máquina incluem, mas não são limitadas a esses tipos (KARRAY et al., 2008).

Figura 3: Diagrama exemplificativo de interação multimodal



Fonte: Ivan Herman, W3C.

A Figura 3 apresenta um diagrama resumindo o contexto de aspectos relevantes para a Interação Multimodal. Segundo Oviatt (1999), Sistemas de Interação Multimodal Humano Computador (IMMHC) são sistemas capazes de combinar diversos canais de entrada de dados, para que o usuário possa se comunicar com o computador. Esses sistemas têm como principal objetivo facilitar a comunicação e torná-la mais agradável para o usuário.

Esses canais de entrada são geralmente baseados na comunicação entre humanos, que utilizam os sentidos como visão, audição, tato, paladar e olfato, e podem ser feitas através de diversos meios tais como fala, postura, gestos das mãos, gestos da cabeça, movimento dos olhos, expressões faciais e corporais, entre outros. Esses canais podem ser representados e mapeados para estender a interação possível com o uso de dispositivos como sensores, como no caso tela sensível ao toque, sensores cardíacos, sensores de ondas cerebrais e afins.

Um dos primeiros que em seu artigo *“Put That There”* (1980) no qual empregou a interação multimodal por meio de voz e gestos em uma interface gráfica, a fim de criar e movimentar objetos em um ambiente virtual.

O sistema *“Put That There”* utilizava o processamento de voz em paralelo a um mecanismo de apontamento, que permitia aos usuários referenciar e mover objetos bidimensionais exibidos em uma tela de projeção. Diversos sistemas multimodais surgiram desde então, combinando outras entradas, como gestos, movimentos, posição do corpo e comandos de voz.

A interação multimodal pode ser classificada em três diferentes tipos (MARAGOS; POTAMIANOS; GROS, 2008; WASINGER, 2006):

- Interação sequencial – Aquela que recebe uma interação por uma modalidade, no entanto essa modalidade pode mudar com o tempo. Dessa maneira, o usuário pode interagir com uma única modalidade por vez, mudando quando necessário. Esse tipo de interação não precisa de uma integração multimodal. Exemplo dessa interação é um usuário usar o teclado para digitar seu nome e um mouse para selecionar seu gênero em um formulário.
- Interação simultânea – Aquela que recebe como entrada múltiplas modalidades, mas as processa separadamente na ordem em que foram recebidas. Um exemplo dessa interação é uma pessoa acelerar um carro e utilizar o freio de mão. As modalidades são distintas e processadas separadamente/independente.
- Interação composta – Aquela que recebe múltiplas modalidades ao mesmo tempo, mas são processadas como uma única entrada composta. Um exemplo dessa interação é uma aplicação que o usuário fale “Aumente aqui” e com um cursor, desenhe um círculo sobre a área que quer aumentar. O sistema recebe as interações ao mesmo tempo, mas as processa de forma integrada, dando a resposta adequada ao usuário.

A partir dessas interações, reconhecem-se dois diferentes usos da modalidade (MAES; SARASVAT, 2003; MARAGOS; POTAMIANOS; GROS, 2008), a saber:

- Multimodal suplementar: Tais aplicações permitem realizar as interações por meio de todas suas multimodalidades, como se fosse única. Basicamente, a aplicação pode suspender a interação com uma modalidade específica e continuar a interação com outra modalidade.
 - Exemplo:
 - Fala: - Selecione o objeto à esquerda.
 - Gesto: Usuário aponta para esquerda.

- Multimodal complementar: A aplicação faz a interação com uma modalidade e utiliza outra para complementar a interação com a primeira modalidade.
 - Exemplo:
 - Fala: - Selecione aquele objeto.
 - Gesto: Usuário aponta para o objeto desejado.

Thiran, Bourlard e Marques (2010) consideram que os sistemas de interação multimodal: a) são capazes de processar entre dois a três canais de entrada simultaneamente; b) são sistemas para aplicações específicas; e c) têm uma limitação limitada de vocabulário e gramática. Sistemas multimodais de fala e gestos agora processam entradas gestuais complexas, quando comparadas com o simples gesto de apontar, sendo dignos de nota os sistemas que identificam o movimento de caneta ou dos lábios, em associação com a fala (OVIATT; COHEN, 2000; OVIATT et al., 2000). Recentemente, os sistemas multimodais estão incorporando modalidades de gestos manuais em 3D juntamente com o reconhecimento da fala, entretanto ainda estão longe da maturidade quando comparados às tecnologias usadas no reconhecimento das entradas bidimensionais de caneta (THIRAN; BOURLARD; MARQUES, 2010).

2.2 Robótica Assistiva

Segundo Lebec e outros (2013), robótica assistiva (RA) é uma área de pesquisa que tem recebido uma quantidade de atenção crescente desde o início da década de 1970. Os esforços iniciais envolvendo grandes e complexos sistemas robóticos evoluíram para sistemas mais acessíveis e agradáveis do ponto de vista funcional e estético (BROSE et al., 2010).

O objetivo da tecnologia assistiva é desenvolver ajudas técnicas avançadas para promover uma vida independente e para melhorar a qualidade de vida de pessoas com doenças crônicas ou degenerativas em habilidades motoras, sensoriais, comunicativas e/ou cognitivas (ZOLLO; WADA; VAN DER LOOS, 2013). Possui potencial para aumentar a autonomia e a capacidade de seus usuários em casa, no local de trabalho e na comunidade, além de reduzir custos com cuidadores.

O impacto clínico desses avanços é observado quando i) usuários se tornam capazes de se integrar ao mercado de trabalho, ii) diminui a sobrecarga de seus cuidadores e iii) a vida em casa se torna viável por mais tempo, diminuindo a necessidade de estabelecimentos de cuidados ou asilos (BROSE et al., 2010). Benefícios adicionais incluem prevenção de complicações médicas, melhoria da autoimagem e vida satisfatória. Adequar as necessidades únicas de uma pessoa com deficiência ou incapacidade aos serviços fornecidos por dispositivos robóticos impõe-se como um desafio.

Groothuis, Stramigioli e Carloni (2013) elaboraram uma lista de critérios que consideram capturar propriedades fundamentais de adequação de um braço robótico para fins assistivos, critérios esses que podem ser estendidos a outras aplicações. São eles: 1) segurança da interação, associada ao nível de segurança do homem quando interage com o robô; 2) robustez de choque, cobrindo a robustez do sistema robótico frente a choques de alto impacto; 3) controle de posição, sobre a acurácia e repetibilidade das capacidades de posicionamento; 4) energia, armazenagem e reuso de energia; 5) adaptabilidade, que diz

respeito à dinâmica entre desempenho e segurança, face às influências de ambientes ou condições arbitrárias/adversas.

A literatura da área de robótica assistiva propõe diversas taxonomias, sendo que sua análise sistemática conduz a cinco categorias principais (ZOLLO; WADA; VAN DER LOOS, 2013): 1) robôs fisicamente assistivos; 2) robôs socialmente assistivos; 3) interface de usuário e sistemas de controle; 4) sistemas sensoriais e de feedback em robótica assistiva; e 5) casas inteligentes e dispositivos domóticos, que permitem a gestão de todos os recursos habitacionais (automação residencial).

Robôs socialmente assistivos são capazes de perceber e compreender o comportamento e as necessidades humanas, se comunicando com os usuários de uma maneira centrada no humano e respondendo com segurança e eficiência aos comandos recebidos (RANASINGHE et al., 2014). Robôs assistivos utilizam conceitos de interação multimodal humano-computador para comunicação entre o robô e seu usuário.

Avanços na robótica, diminuição dos custos, aumento da capacidade de processamento e da tecnologia sensorial, assim como o aumento da sinergia entre robótica e outras disciplinas médicas e técnicas farão com que seja cada vez mais viável a assistência de robôs em atividades da vida diária, ampliando a integração e participação de pessoas com deficiências ou incapacitadas (ZOLLO; WADA; VAN DER LOOS, 2013).

Projetos de robôs assistivos mais complexos estão em execução há mais de uma década (BAHADORI et al., 2003; GOETZE; FISCHER; MORITZ, 2012; K. ZSIGA, M. RAUHALA, P. RUMEAU, A. TÓTH, T. PILISSY, 2011; POLLACK et al., 2002). Atualmente já podem ser observadas algumas iniciativas em que estes projetos disponibilizaram versões disponíveis no mercado, tais como o *Care-O-bot®* (GRAF; HANS; SCHRAFT, 2004; GRAF et al., 2009; HANS; GRAF; SCHRAFT, 2002), cuja quarta versão foi disponibilizada em 2015 (HÖLN; NEMITZ, 2015).

2.3 Adaptatividade

A adaptação de sistemas interativos descreve mudanças na interface que ocorrem para aumentar a usabilidade ou satisfação do usuário (BEZOLD; MINKER, 2011). O termo adaptação denota personalização que é automaticamente executada pela interface, sem qualquer diretiva explícita do usuário (WELD et al., 2003). O principal objetivo da interface de usuário adaptativa é a plasticidade, que denota a capacidade de a interface de usuário se adaptar, preservando sua usabilidade através de múltiplos contextos de uso (COUTAZ, 2010).

Dey, Abowd e Salbwe (2001) definiram contexto como:

qualquer informação que possa usada para caracterizar de entidades que são consideradas relevantes para a interação entre um usuário e uma aplicação, incluindo o próprio usuário e a aplicação. Contextos são tipicamente: a localização, a identidade e o estado de pessoas, grupos e objetos físicos e computacionais.

Benyon e Murray (1993) consideram que a adaptação automática é um desafio, uma vez que o sistema computacional necessita conter uma representação detalhada e explícita

do usuário (modelo de usuário), de si mesmo (modelo de domínio ou de tarefas) e do sistema de interação (modelo de interação). Esses autores julgam necessário considerar o nível do sistema que será adaptativo e as características do usuário para as quais tais alterações são pretendidas.

Dumas, Solórzano e Signer (2013) afirmam que, dependendo do perfil de um usuário, a interface multimodal pode trocar de uma modalidade para outra, tanto no nível de entrada, quanto no de saída. Assim, é adaptativa a interação multimodal que, ao contemplar as aptidões do usuário, seleciona os canais de comunicação adequados ao contexto. É este o caso deste trabalho, que propõe a adequação dos canais de entrada e saída à deficiência do usuário.

Malinowski e outros (1992) elaboraram importante taxonomia referente à interface adaptativa de usuários. Para a delimitação do escopo, devem ser consideradas informações a respeito do usuário (necessidades e preferências, podendo sendo relevantes ao usuário individual suas características de usuário, habilidades, metas, comportamento, conhecimento e experiência), da aplicação, de aspectos ergonômicos e da interface de usuário, cuja consistência deve ser garantida pela estratégia de adaptação. Segundo esses autores, o principal objetivo da adaptação é apresentar ao usuário uma interface cujo uso seja fácil, eficiente e efetivo. A principal forma de avaliação consiste na aferição da usabilidade, daí a decisão metodológica de submeter os participantes deste estudo a um questionário de usabilidade (ver capítulo 6 – Avaliação).

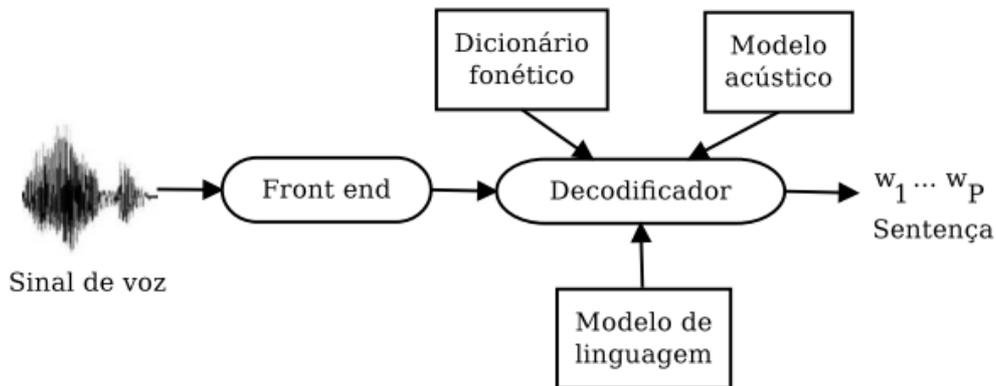
A especificação dos critérios para adaptação consiste na determinação das condições que deveriam impor a adaptação do sistema (MEYER; YAKEMOVIC; HARRIS, 1993), sendo exemplos as seguintes características do usuário, da tarefa ou do ambiente: experiência do usuário com a tarefa, experiência prévia do usuário, aptidão do usuário, preferência do usuário, características demográficas do usuário, complexidade da tarefa, frequência da tarefa, carga de trabalho, condições ambientais. Assim, para os fins deste trabalho, importa ressaltar a possibilidade de adaptação em razão de aptidões do usuário, isto é, da natureza de sua deficiência (visual ou auditiva).

2.4 Reconhecimento de fala

Reconhecimento automático de fala (RAF) ou reconhecimento automático de voz (no inglês, *Automatic Speech Recognition – ASR*) consiste num processo ou técnicas que convertem a fala em palavras sequencias por meio de algoritmos em um computador (LI et al., 2014). A primeira proposta de um modelo para reconhecimento de fala e síntese surgiu em meados de 1930 com *Homer Dudley of Bell Laboratories*; no entanto, a primeira aplicação surgiu apenas em 1952, obra do mesmo laboratório, sendo o sistema capaz de identificar dígitos isolados, falados por um usuário, com base no espectro acústico do fonemas (JUANG; RABINER, 2005).

Um sistema ASR típico adota uma estratégia estatística baseada em modelos ocultos de Markov (Hidden Markov Models – HMM) e é composto por cinco blocos (Fig. 4): *front end*, dicionário fonético, modelo acústico, modelo de linguagem e decodificador.

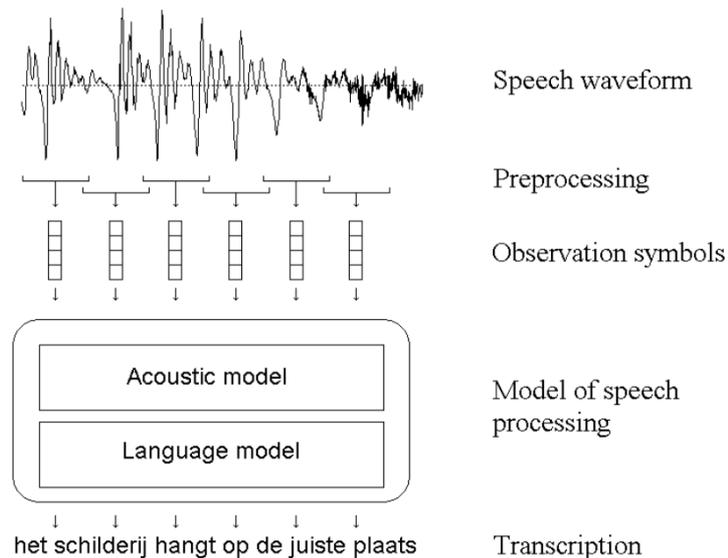
Figura 4: Principais blocos de um sistema ASR



Fonte: BATISTA, 2013

Conforme se depreende da Figura 5, o processo de reconhecimento consiste, basicamente, na captura de ondas acústicas, no pré-processamento esta onda e transformação da onda em sequência de símbolos e vetorização da sequência. O modelo acústico é uma representação do espectro das palavras e determina qual som será produzido a partir de uma determinada *string* de palavras, provendo todas as combinações possíveis para um determinado espectro. O modelo de linguagem, por sua vez, reconhece cada palavra através de um dicionário e utiliza algumas técnicas de mapeamento para prever a próxima palavra subsequente de cada frase ditada (WIGGERS; ROTHKRANTZ, 2003).

Figura 5: Modelo genérico de reconhecimento de fala.

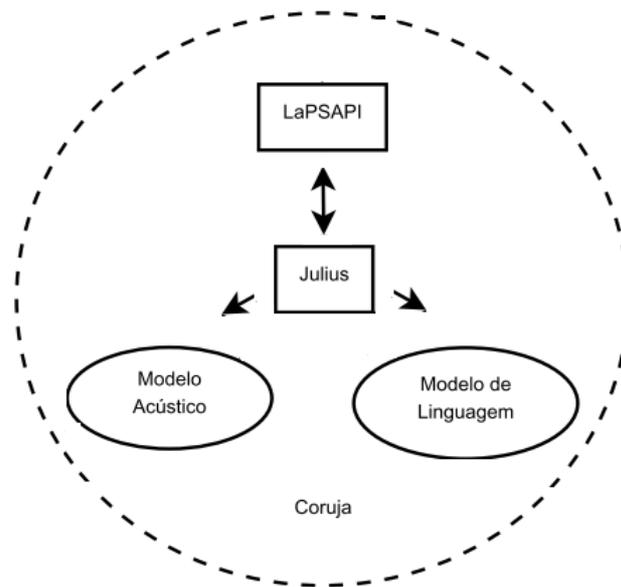


Fonte: Al et al., 2003

Atualmente, há uma grande utilização dessas técnicas de reconhecimento de fala na computação móvel (*Siri – Apple, Cortana – Windows Phone, Google Now – Android*), mediante o processamento de linguagem natural com a finalidade de realizar tarefas simples e fazer buscas na web. Sistemas de reconhecimento de fala ajudam usuários a ter uma melhor interação com o dispositivo e aumentam a acessibilidade de usuários com deficiências ou limitações.

Coruja é uma *engine* de reconhecimento de fala na língua portuguesa (pt-BR) que consiste na utilização do decodificador livre Julius, modelos acústico e de linguagem para reconhecimento de voz em Português Brasileiro e uma API própria, LaPSAPI, para facilitar a utilização do decodificador Julius (BATISTA, 2013), conforme ilustrado na Figura 6. A API pode ser utilizada tanto no sistema operacional Linux (implementações em C++), quanto no Windows (admitindo-se qualquer linguagem compatível com o Microsoft.NET Framework, como C# e Visual Basic).

Figura 6: Coruja – Sistema de reconhecimento de fala em português-brasileiro



Fonte: Adaptado de BATISTA, 2013.

A LaPSAPI permite controle em tempo real do decodificador Julius e da interface de áudio do sistema, permitindo a abstração de detalhes de baixo nível relacionados à *engine*. Isso porque a API consiste basicamente de uma classe principal denominada SREngine, que permite que a aplicação controle aspectos do decodificador Julius, carregando os modelos acústico e de linguagem, iniciando e parando o reconhecimento e recebendo eventos e resultados do processo de reconhecimento de voz.

O maior desafio dentro do projeto de um sistema de ASR é elaborar um modelo acústico, tarefa desempenhada por Oliveira e outros (2011) no bojo do projeto desenvolvido na Universidade Federal do Pará. O treinamento do modelo acústico é feito por meio de uma base de dados com arquivos de áudio e a respectiva transcrição ortográfica. De acordo com Batista (2013), foram empregadas as seguintes bases de dados: LaPSSStory (audiobooks), *corpus* Legislação Brasileira (Constituição e Código de Defesa do Consumidor em áudio), LaPSNews (*corpus* CETENFolha, da Universidade de São Paulo – USP/São Carlos).

Em razão do fato de terem sido empregadas bases de dados públicas cujo áudio é totalmente livre de ruído, as técnicas de *adaptação maximum likelihood linear regression* e *maximum a posteriori* foram aplicadas ao modelo acústico, a fim de diminuir o efeito do descasamento acústico entre os ambientes de treino e teste, diminuindo a taxa de erro por palavras (*word error rate* – WER).

Um modelo de linguagem trígama foi elaborado utilizando 1,6 milhões de frases retiradas das bases CETENFolha, Spoltch, OGI-22, West Point e *audiobooks* e de textos extraídos da internet.

O emprego de um conversor grafema-fone capaz de extrair automaticamente regras de transição de um léxico por meio de técnicas de ML por indução (árvore de decisão) e Naïve Bayes resultou num dicionário fonético de grande vocabulário para o PB com aproximadamente 60 mil palavras (UFPAdic versão 2). O dicionário fonético (modelo léxico) mapeia as palavras em unidades básicas (fones) e vice-versa.

Após o treinamento de todos os modelos estatísticos, um sistema ASR na etapa de teste usa o *front-end* para converter o sinal de entrada em parâmetros e o decodificador para encontrar a melhor sentença. Os modelos acústico e de linguagem podem ser adaptados para melhorar a performance. Aplicações que empreguem vocabulário técnico demandam um modelo de linguagem específico, enquanto o modelo acústico pode ser adaptado a um locutor ou determinado sotaque.

2.5 Síntese de voz: gênero

De acordo com Nass, Moon e Green (1997), a escolha de uma voz em particular dispara no usuário uma série de expectativas associadas ao gênero daquela voz. Vozes masculinas trazem um grande conjunto de expectativas e respostas baseadas em estereótipos sobre homens, enquanto vozes femininas fazem o mesmo quanto ao estereótipo sobre mulheres.

Culturalmente, ainda prevalece o estereótipo de que mulheres desempenham funções de auxílio (FISKE et al., 2002), cabendo aos homens o comando. Considera-se que as mulheres soam menos ameaçadoras e mais calmas (CORSO, 2010) do que os homens e que são presumivelmente melhores do que eles em comunicação e relacionamentos.

Haja vista a tendência a considerar que a voz feminina ajuda na resolução de problemas, enquanto uma voz masculina autoritariamente impõe uma solução, o usuário que se vale da tecnologia provavelmente optaria por uma interface feminina, a fim de manter seu protagonismo na interação. Não por acaso, Siri e Cortana – assistentes pessoais da Apple e da Microsoft, respectivamente – possuem voz feminina.

Embora haja indícios de maior aceitação e proximidade psicológica na interação humano-computador quando robô e usuário são do mesmo gênero (EYSSEL et al., 2012), o efeito “mulheres são maravilhosas” – cunhado por Eagly, Mladinic e Otto (1994) – se manifesta e as mulheres são vistas mais positivamente do que homens.

Segundo Mitchell e outros (2011), a literatura mostra que mulheres fazem mais associações positivas implícitas com o próprio gênero do que com o gênero masculino, enquanto homens tendem à neutralidade. Tanto homens como mulheres expressam preferência pelas mulheres em medidas explícitas. Rudman e Goodwin (2004) confirmam que as mulheres preferem implícita e fortemente o próprio gênero, não se passando o mesmo em relação aos homens.

Em razão do exposto, o pacote de linguagem em português-br Microsoft® Speech TTS Heloisa foi escolhido para que o robô possuísse uma voz feminina, dadas suas finalidades. Os serviços disponíveis no sistema de interação multimodal embarcado em robótica assistiva consistem em prestação de auxílio ao estudante com deficiência, sendo conveniente, segundo a literatura abordada, a escolha de uma voz feminina.

2.6 Reconhecimento de face

De acordo com a pesquisa realizada por Zhao e outros (2003), o trabalho mais remoto sobre reconhecimento facial, de Bruner e Tagiuri, No campo da psicologia, data de 1954, somente surgindo na literatura de engenharia na década seguinte pelas mãos de Bledsoe. As pesquisas sobre reconhecimento mecânico automático de faces começaram com Kelly em 1970, datando de três anos depois o seminal trabalho de Kanade.

É possível classificar os sistemas de técnicas de reconhecimento facial (no inglês, *Face Recognition Techniques – FRT*) em razão de usarem imagens estáticas ou vídeo. Dentro desses grupos, diferenças significativas existem, a depender da aplicação específica. Segundo Zhao e sua equipe, as diferenças consistem em qualidade da imagem, desorganização do fundo, a variabilidade das imagens de um indivíduo que deva ser reconhecido, disponibilidade de um reconhecimento bem- definido ou critério de correspondência, bem como a natureza , o tipo e a quantidade de entrada de um usuário.

O reconhecimento mecânico de faces consiste na identificação ou verificação de uma ou mais pessoas na cena estática ou filmada por meio do uso da base de dados de faces armazenadas. Informações secundárias, como raça, idade, gênero, expressões faciais ou fala, podem ser empregadas para estreitar a busca, aumentando o reconhecimento. A solução para o problema compreende segmentação de faces (detecção facial) em cenas desorganizadas, extração das regiões faciais, reconhecimento ou verificação. Na identificação, o sistema tem como entrada um rosto desconhecido e devolve a determinação da identidade com base no banco de dados de indivíduos conhecidos. Na verificação, o sistema precisa confirmar ou rejeitar a identidade atribuída à face de entrada (ZHAO et al., 2003).

Ainda segundo Zhao e outros, os principais problemas com que um sistema de reconhecimento facial baseado em vídeo lida são:

- a) baixa qualidade do vídeo, já que geralmente a aquisição do vídeo se dá em locais abertos, ou mesmo em locais fechados, mas com condições ruins de captura, o que gera variações de iluminação pose nas imagens. Ainda, oclusões parciais e dissimulações são possíveis;
- b) tamanho das imagens da face, pequenas, em razão das condições de aquisição, afetando a acurácia da segmentação facial, assim como a detecção de pontos e marcações geralmente necessários aos métodos de reconhecimento;
- c) as características de faces e partes do corpo humano, uma vez que as variações dos corpos e faces é muito pequena, razão por que o reconhecimento de indivíduos dentro de uma classe é difícil. É mais fácil detectar e reconhecer faces do que distingui-las, reconhecê-las especificamente.

Podemos relacionar as seguintes técnicas como críticas para o reconhecimento facial em vídeo: segmentação facial e estimação de pose, rastreamento facial e modelagem facial. Originalmente, o reconhecimento facial em vídeo se originou a partir de técnicas baseadas em imagens estáticas. A essa categoria pertencem os seguintes métodos: faces próprias (*eigenfaces*)(SIROVICH; KIRBY, 1987; TURK; PENTLAND, 1991), faces próprias probabilísticas(MOGHADDAM; PENTLAND, 1997), correspondência gráfica elástica (*ElasticBunchGraphMatching – EBG*M) (WISKOTT et al., 1997) e rede neural baseada em

decisão probabilística (*Probabilisticdecision-based neural network - PDBNN*)(LIN; KUNG; LIN, 1997).

A aplicação de rastreamento é uma melhoria em relação a esses métodos, por meio da sintetização de uma vista frontal a partir da estimativa de pose e profundidade dada pelo vídeo. A taxa de reconhecimento pode ser melhorada mediante “votação” probabilística baseada nos resultados do reconhecimento de cada quadro (MCKENNA; GONG, 1997).

Sistemas multimodais oferecem soluções mais amplas à tarefa de identificação, que poderia não ser obtida apenas com o uso das imagens isoladas. É possível associar a análise de características de movimentos corporais e voz, como no sistema descrito por Chowdhury e outros (1999), o qual consiste em um módulo de reconhecimento facial, um módulo de identificação do interlocutor e um módulo de fusão classificadora.

2.7 Reconhecimento de gestos

Gestos são úteis na IHC porque se tratam da forma mais primária e expressiva da comunicação humana (WACHS et al., 2011). Segundo Karam e Schraefel (2005), a maioria das teorias por trás dos gestos como uma técnica de interação computacional decorre das pesquisas multidisciplinares, colhendo conceitos da linguística, antropologia, ciência cognitiva e psicologia.

O conhecimento da classificação das ações gestuais permite um design mais minucioso da linguagem gestual, indicando se o gesto para determinada ação deve ser escolhido entre gestos estáticos de uma mão, bimanuais dinâmicos ou códigos semafóricos, por exemplo (AIGNER et al., 2012).

A fim de superar as limitações da interação WIMP (sigla em inglês para janelas, ícones, menus e ponteiros), é possível optar por detecção de gestos no ar, interfaces multi-toque, tangíveis ou orgânicas (AIGNER et al., 2012). É necessário empregar técnicas de detecção, rastreamento e segmentação de gestos para só então identificar o significado pretendido com o gesto na interação.

São muitos os métodos de reconhecimento de gesto, a seguir arrolados (CHHABRIA; DHARASKAR; THAKARE, 2013): linha *scan* para mão; algoritmo do componente conexo; abordagem híbrida para reconhecimento de gesto e fala (uma combinação entre mapas auto-organizados e modelos markovianos para classificação da trajetória do gesto, além dos seguintes algoritmos: Kohonen, Viterbi e Baum-Welch); algoritmo de alinhamento temporal dinâmico (*dynamic time warping*), avaliador de desempenho e análise de erro.

São exemplos de sistemas multimodais que empregam reconhecimento de gesto: PutThatThere, Cubricon, Xtra, Quickset, RIA withMIND. (CHHABRIA; DHARASKAR; THAKRE, 2012).

3 TRABALHOS RELACIONADOS

Ao apresentar os principais trabalhos relacionados ao tema proposto nesta dissertação, este capítulo aborda o estado da arte, identificando a forma como os robôs assistivos têm sido projetados, tanto sob aspectos de sua arquitetura, quanto em relação aos recursos empregados. São apresentados, também, diversos sistemas de interação multimodal, inclusive adaptativos.

Os trabalhos relacionados seguem duas linhas principais: de implementação de um modelo de robótica assistiva multimodal (FISCHINGER; EINRAMHOF, 2013; MAYER; BECK; PANEK, 2012; REISER et al., 2009) e de implementação de modelos de interação multimodal adaptativa.

A adaptação nestes modelos e exemplos pode ser observada em diversos aspectos, como no nível de saídas multimodais, no gerenciamento do diálogo ou na seleção de serviços. Alguns exemplos descritos para estes casos estão relacionados neste capítulo, para a adaptação da interface (MAAT; PANTIC, 2006; REITHINGER et al., 2003; STEFFEN, 2011), entradas ou saídas (ULUER; AKALIN; KÖSE, 2015), gerenciamento do diálogo (GNJATOVIĆ et al., 2012; STEFFEN, 2011) ou seleção dos serviços (DJAID; SAADIA; RAMDANE-CHERIF, 2012).

Os ajustes gerados pelos processos de adaptação são motivados por aspectos diversos e podem ser observados trabalhos que utilizam para a adaptação as preferências do usuário (MAAT; PANTIC, 2006; MÜLLER; SPRENGER; GROSS, 2014; STEFFEN, 2011), as suas necessidades (ULUER; AKALIN; KÖSE, 2015) ou ainda o contexto (DJAID; SAADIA; RAMDANE-CHERIF, 2012; REITHINGER et al., 2003; STIEFELHAGEN et al., 2007).

Para favorecer aspectos de análise este capítulo está organizado em dois conjuntos de trabalhos, que possuem algumas diferenças de foco entre si, pois um deles trata do estado da arte de robótica assistiva e outro é dedicado aos aspectos de atualização em trabalhos na área de Interação multimodal. Desta forma, são aspectos complementares que contribuem para identificar possibilidades de contribuição levando em conta a robótica assistiva (seções 3.1, 3.2 e 3.3) e aplicações de Interação Multimodal (seções 3.4 a 3.11). Tanto aspectos associados ao simples uso e contexto de aplicação dos trabalhos, como aspectos técnicos e arquiteturais são expostos e analisados.

A seção 3.12 propõe um quadro comparativo a partir do qual é possível identificar lacunas não cobertas pelas pesquisas desenvolvidas pela comunidade acadêmica e das quais decorre a contribuição do presente estudo.

3.1 Care-O-bot®

Care-O-bot é um robô doméstico que foi colocado no mercado pela empresa alemã *Fraunhofer*. Comercialmente, já se encontra em sua quarta versão (HÖLN; NEMITZ, 2015), ilustrada na Figura 7.

Figura 7: Care-O-bot® 4 – um robô de serviço.



Fonte: HÖLN; NEMITZ, 2015

A versão anterior, *Care-O-bot® 3*, é empregada como plataforma robótica em diversos projetos de pesquisa financiados pela União Européia e/ou em curso em universidades daquele continente, como o *SeRoDi* (2015), conduzido pela Universidade de Stuttgart, que serve como apoio ao cuidadores de pacientes em instituições de saúde. Também possui aplicações no setor de serviços, como entrega de pedidos em restaurantes, recepção e serviço de quarto em hotéis, e na indústria, em atividades como encher e esvaziar máquinas, tarefas de transporte, seleção de produtos em prateleiras e despacho de pedidos.

Care-O-Bot® foi concebido para um público com dificuldades motoras e/ou de locomoção – idosos, pessoas debilitadas, doentes crônicos (epilépticos), e pessoas com necessidades especiais permanentes, tais como paraplégicos, ou temporárias, como acidentados. A literatura disponível versa sobre o estado da arte do *Care-O-bot®3* (CONNETTE et al., 2008; GRAF et al., 2009; REISER et al., 2009). O robô possui braço mecânico dotado 7 graus de liberdade, uma garra de 3 dedos com sensores táteis e 7 graus de liberdade, uma bandeja para a entrega dos objetos e um sistema de navegação que lhe permite andar pelos cômodos da casa.

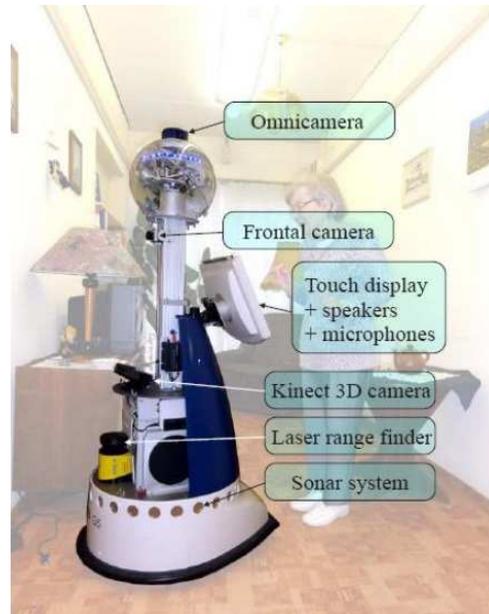
Seu display é *touchscreen*, com interface gráfica para o usuário (do inglês *Graphical User interface* – GUI). Possui visão 3D, graças a duas câmeras posicionadas na cabeça, sendo capaz de identificar e aprender objetos. Realiza reconhecimento de voz, possui sintetizador de fala, além de funções extras, como fazer a limpeza simples, pegar, buscar ou servir coisas.

3.2 ALIAS

ALIAS – Adaptable Ambient Living Assistant (MAYER; BECK; PANEK, 2012) é um assistente robô móvel cujo objetivo é manter constante contato com seus usuários para prevenir o isolamento e estimular a interação social. O protótipo está representado na Figura

8. A interface de usuário possui diferentes modalidades de entrada: os comandos de voz, o Brain Computer Interface - BCI, ou a tela sensível ao toque (GUI).

Figura 8: Protótipo ALIAS

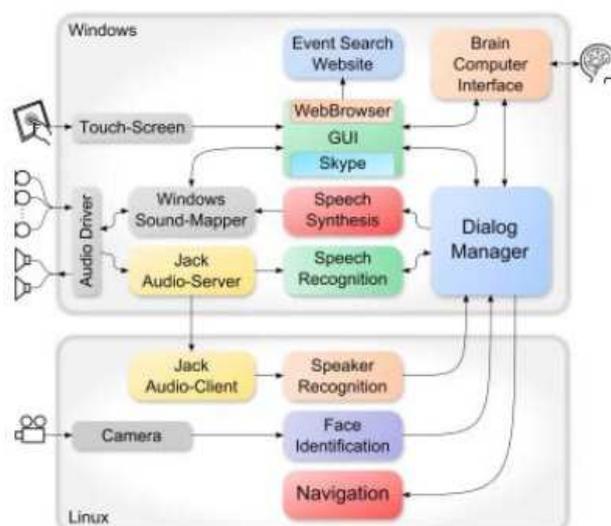


Fonte: REHRL et al., 2012

Um de seus diferenciais em relação aos demais assistentes consiste no BCI, que consiste na captação de estímulos elétricos via eletrodos posicionados sobre a cabeça do paciente que porventura esteja incapacitado de falar e de se mexer (em razão de sequelas de um acidente vascular cerebral, por exemplo). Os estímulos captados são então utilizados para comandar a interface gráfica do display.

Conforme exposto na Figura 9, o robô ALIAS possui módulos de reconhecimento de fala, síntese de fala, reconhecimento facial, navegação, reconhecimento de interlocutor, interface GUI, além do já mencionado BCI.

Figura 9: ALIAS – módulos do sistema



Fonte: MAYER; BECK; PANEK, 2012

A figura 9 identifica os elementos principais de sua arquitetura. O gerenciador de diálogo é um dos mais centrais componentes do robô, já que é responsável pelas decisões que o robô precisa tomar e está ligado a todos os demais módulos, articulando-os, interpretando-os e decidindo quais ações implementar.

3.3 Hobbit

Hobbit (FISCHINGER; EINRAMHOF, 2013; PAPOUTSAKIS et al., 2013b), ilustrado na Figura 10, é um robô assistente que ajuda idosos em suas casas em tarefas como pegar objetos e detectar situações de emergência. Dentre suas principais funcionalidades, destacam-se: detectar queda e ajudar, trazer objetos, limpar o chão, entreter.

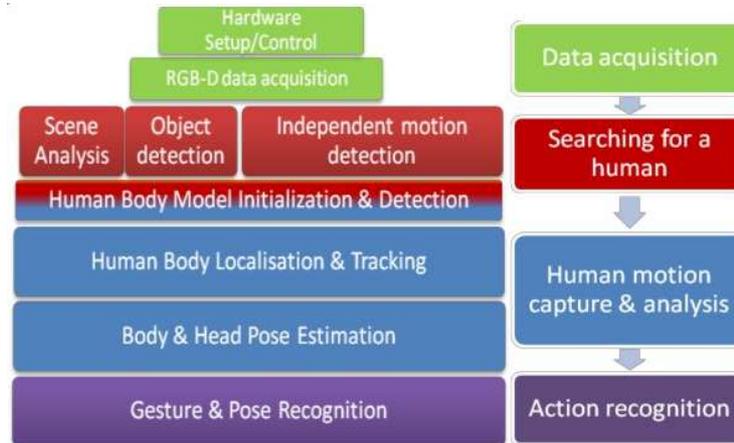
Figura 10: Protótipo Hobbit, sem e com carcaça.



Fonte: FISCHINGER; EINRAMHOF, 2013

A fim de facilitar a chamada do robô a locais específicos da residência, pode ser integrado a um *AmbientAssisted Living - AAL*. Assim, se o usuário pressiona um botão localizado ao lado da cama, por exemplo, faz com que o robô navegue diretamente para aquele ponto do quarto.

Figura 11: Hobbit – arquitetura do sistema para observação humana.



Fonte: PAPOUTSAKIS et al., 2013a

A interface multimodal consiste numa GUI *touch screen*, ASR (*automatic speech recognition*), TTS (*text to speech*), e GRI (*gesture recognition interface*). Um de seus diferenciais é a capacidade de aprender a identificar novos objetos. Na Figura 11, acima, a arquitetura do sistema para a observação humana. Conta com braço mecânico com 5 graus de liberdade e garras com efeito *finray*, à prova de deslize, que permitem que os dedos se ajustem mecanicamente aos objetos, que não escapam, aumentando sua capacidade de pegar coisas sem deixar cair. Hobbit também faz vídeo-chamadas e chamadas de emergência, o que é bastante útil em sua função de prestar assistência.

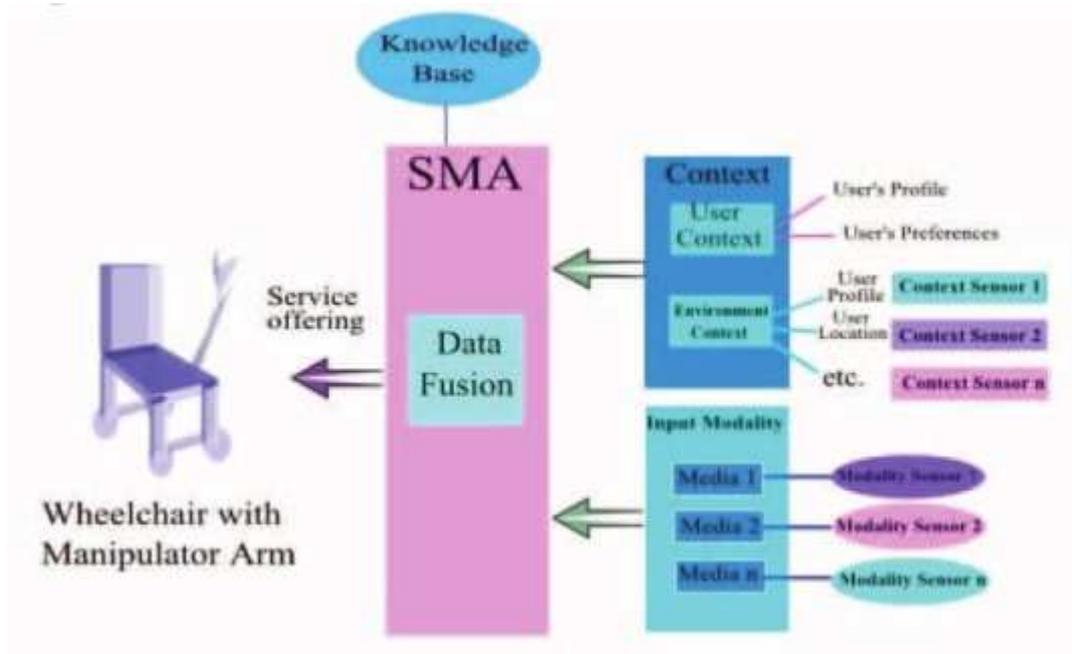
3.4 Multimodal architecture to strengthen the interaction of the robot in ambient intelligence environments

Os autores Djaid, Saadia e Ramdane-Cherif(2012) propõem a multimodalidade adaptativa como meio de prover uma vida confortável e segura para as pessoas dependentes que não podem usar seus corpos para se mover e viver. Infelizmente numa interação homem-computador clássica, o sistema computacional não tem a capacidade de tirar vantagem do ambiente a fim de melhorar as interações. A computação ciente do contexto e presentemente a computação ubíqua (*pervasive*) permitem isso.

A plataforma se baseia nas preferências e necessidades do usuário, assim como um conjunto de processos que coordena a operação a fim de permitir que o robô preste o serviço que mais se adequa ao perfil do usuário e suas preferências e ao contexto ambiental, como sua localização. A aplicação sugerida no artigo é uma cadeira robótica equipada com um braço de manipulação.

A arquitetura do software consiste num sistema multi-agente que considere o contexto e admita entradas de diversas modalidades, como fala, gesto, controle por toque, mouse. Cada agente considera uma ou mais informações do sistema (v. Figura 12). Assim, o sistema implementado observa continuamente o estado das informações dos diferentes agentes e age de acordo com elas. As saídas do sistema decorrem dos contextos do usuário e do ambiente. Os estados das entradas de informação de cada agente são representados numa fila.

Figura 12: Arquitetura global do sistema multimodal



Fonte: DJAID; SAADIA; RAMDANE-CHERIF, 2012

A adaptação deste modelo ocorre a nível de arquitetura, propondo uma arquitetura mais flexível e adaptável a fim de adicionar agentes para diferentes objetivos em um ambiente ubíquo.

A Arquitetura é capaz de suportar os termos das entradas e os valores do contexto, efetuando a fusão e resultando numa saída adequada a cada cenário. Cada agente é modelado por uma rede Petri e gerenciado por um conjunto de regras. As regras determinam quando uma transição pode ser alcançada e assim troca o estado do sistema. Também foi aplicada lógica *fuzzy* em razão de sua habilidade de processar informações de acordo com um leque de variações.

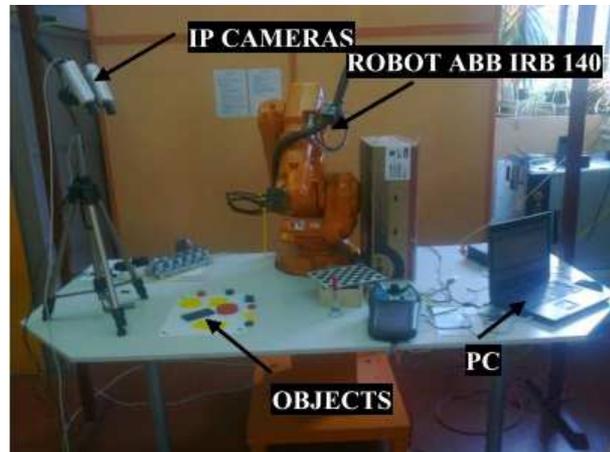
3.5 Adaptive Multimodal Interaction with Industrial Robot

Gnjatović e outros (2012) desenvolveram um sistema de diálogo em linguagem falada natural que gerencia a interação entre o usuário e o robô industrial ABB IRB 140. Tal sistema de diálogo é multimodal, usando três modalidades comunicativas: (i) linguagem falada (ASR e síntese TTS), (ii) reconhecimento visual de figuras e determinação de suas posições, e (iii) texto digitado. Por se tratar de um sistema adaptativo, considera os contextos verbais e espaciais a fim de adaptar seu comportamento no diálogo e de processar comandos do usuário formulados espontaneamente em diferentes formas sintáticas. O robô industrial é capacitado para manipular figuras gráficas, seguindo as instruções do sistema de diálogo.

O cenário de interação reside numa interação terapêutica entre o fonoaudiólogo e a criança com dificuldades de fala. Durante a interação, usuário e sistema compartilham dois contextos relacionados, a saber, o contexto espacial, que consiste num conjunto de figuras coloridas tridimensionais de madeira triangulares, circulares, quadradas e retangulares, e o

verbal, no qual o usuário profere comandos em diferentes formas sintáticas. Na figura 13 é possível observar a configuração do protótipo.

Figura 13: Solução integrada – robô industrial com interação multimodal adaptativa

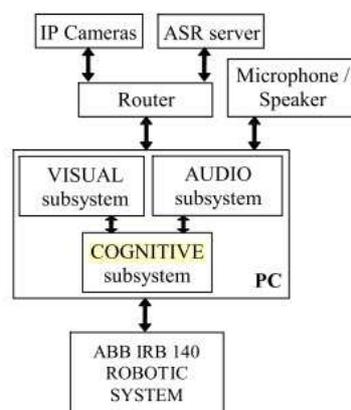


Fonte: GNJATOVIĆ et al., 2012

Quando o usuário não fornece todas as informações necessárias, o sistema aplica uma estratégia de diálogo adaptativa a fim de ajudar o usuário a especificar a informação requerida. Se a criança fala “quadrado amarelo”, então o sistema retorna uma solicitação de especificação: “Como devo interpretar seu comando: quadrado amarelo grosso ou quadrado amarelo fino?”. Após a resposta, o sistema pergunta o que fazer com a figura geométrica: mostrá-la ou movê-la? Assim, sempre que o usuário não formular completamente o comando ou empregar comandos ambivalentes, o sistema toma a iniciativa de guiar o usuário até um comando válido, formulando perguntas que direcionam o diálogo.

Figura 14 representa o sistema multimodal, integrado por quatro subsistemas: o subsistema visual implementa funcionalidades de reconhecimento visual das figuras tridimensionais localizadas sobre uma superfície branca, e determina suas posições; o subsistema de áudio implementa as funcionalidades de ASR e síntese TTS; já o subsistema cognitivo implementa as funcionalidades de processamento de linguagem natural e gerenciamento adaptativo do diálogo; enquanto o subsistema robótico externo é o robô industrial ABB IRB 140, um braço robótico antropomórfico com seis graus de liberdade.

Figura 14: Representação esquemática do sistema integrado.



Fonte: GNJATOVIĆ et al., 2012

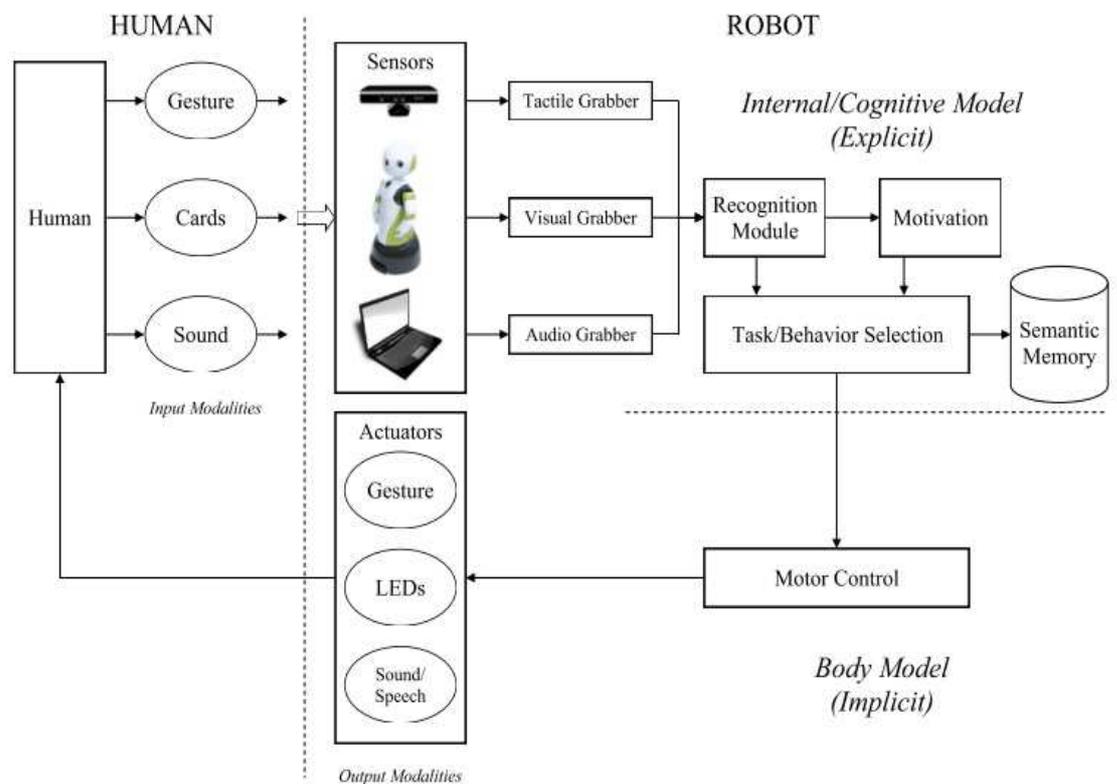
3.6 A New robotic platform for sign language tutoring

Trata-se de uma plataforma robótica para tutoria de língua de sinais (ULUER; AKALIN; KÖSE, 2015). Socialmente interativo, o robô humanoide foi concebido como um companheiro de jogos para crianças com deficiência auditiva, ensinando a língua de sinais turca por meio da língua de sinais e de dicas visuais. A plataforma robótica com cinco dedos *Robovie R3* é usada para expressar um conjunto de palavras (sinais que significam mãe, jogar, primavera, bebê, eu/me, grande, montanha, vir, mesa e preto), combinando movimentos de mão e corpo com expressões faciais. O robô reconhece sinais, dando retornos motores, visuais e vocais.

Estudos pilotos demonstraram que o desenho experimental deveria ser adaptativo, em respeito ao grau de deficiência do participante. Crianças com audição parcial são motivadas e ajudadas com a combinação do emprego de uma voz infantil pelo robô e de cartões coloridos; sendo suficientes apenas os cartões para as que não possuem audição. Esses cartões com desenhos permitem a comunicação com a criança e a manutenção do fluxo interativo de jogo.

A arquitetura do sistema admite entradas perceptíveis por sensores (som, carta e gestos), um modelo interno cognitivo (explícito) com módulo de reconhecimento, motivação e seleção de tarefas ou comportamento e memória semântica – v. Fig. 15. O controle motor, que faz parte do modelo corporal (implícito) dispara atuadores que ativam as modalidades de saída.

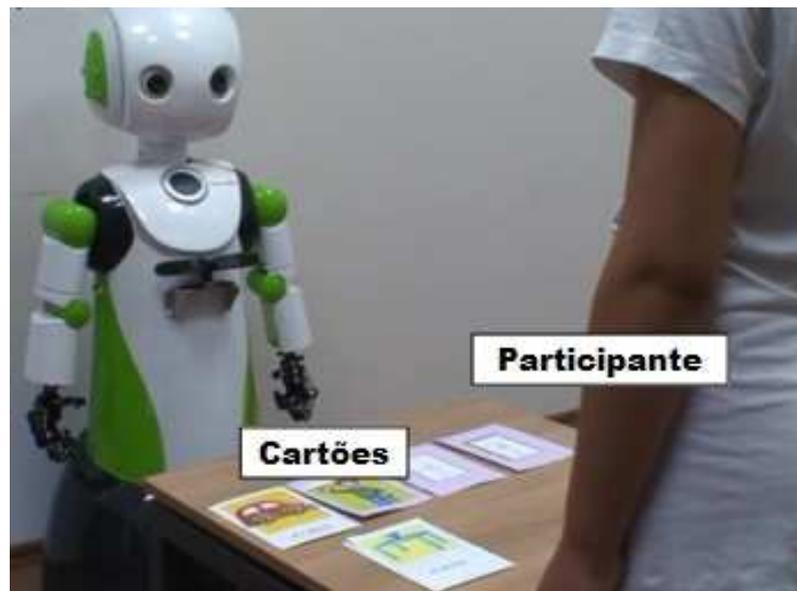
Figura 15: Arquitetura do sistema



Propõe um desenho de cenários de interação adaptativos para cada criança, motivando as crianças a interagir com um companheiro social com o potencial de ser percebido como um igual. Pretende-se que seja usado por professores, terapeutas ou membros familiares como uma parte da educação da língua de sinais, permitindo que a criança melhore sua capacidade de interação, encorajando-a a transmitir o conhecimento aprendido por meio da interação com o robô para sua interação com outros indivíduos.

A novidade do projeto é o fato de empregar um robô humanoide tanto para produzir como reconhecer gestos em jogos interativos para o aprendizado da língua de sinais – v. Fig. 16. A estrutura estabelecida para a interação consiste na obtenção pelo robô de informações relevantes de um humano por meio de diferentes modalidades, como reconhecimento de gestos e sinais, cartas e voz (para indivíduos com audição parcial ou capazes de se expressar mediante vocalização), e seleciona a ação adequada, dando o retorno necessário ao jogador humano. A depender do contexto de jogo, diferentes abordagens são usadas para selecionar a ação apropriada assim que a informação de entrada é reconhecida. Os cenários de jogo propostos e os diferentes níveis de interação são desenhados não apenas para crianças com perda total de audição, como também crianças com perda parcial ou indivíduos com audição normal que queiram aprender a se comunicar por meio da língua de sinais.

Figura 16: Protótipo tutor funcional



Fonte: Adaptado de ULUER; AKALIN; KÖSE, 2015

Robovie R3 foi ligeiramente modificado com a inclusão de graus de liberdade em punhos e nos cinco dedos e uma boca de LED que imita e/ou expressa emoções básicas como felicidade, tristeza ou neutralidade). Em seu peito há uma câmera de profundidade ASUS RGB-D usada para reconhecimento de gestos.

A contribuição do modelo é relevante, pois pretende admitir qualquer língua de sinais. A adaptação, no entanto, é apenas um item mencionado no artigo, não considerado na representação de seu modelo, nem na implementação do protótipo, o qual possui apenas reconhecimento de cartas, síntese de gestos e manifestação de estados emocionais por meio dos LED posicionados na boca.

3.7 Adaptive input interpretation for dialogue management of an autonomous robot

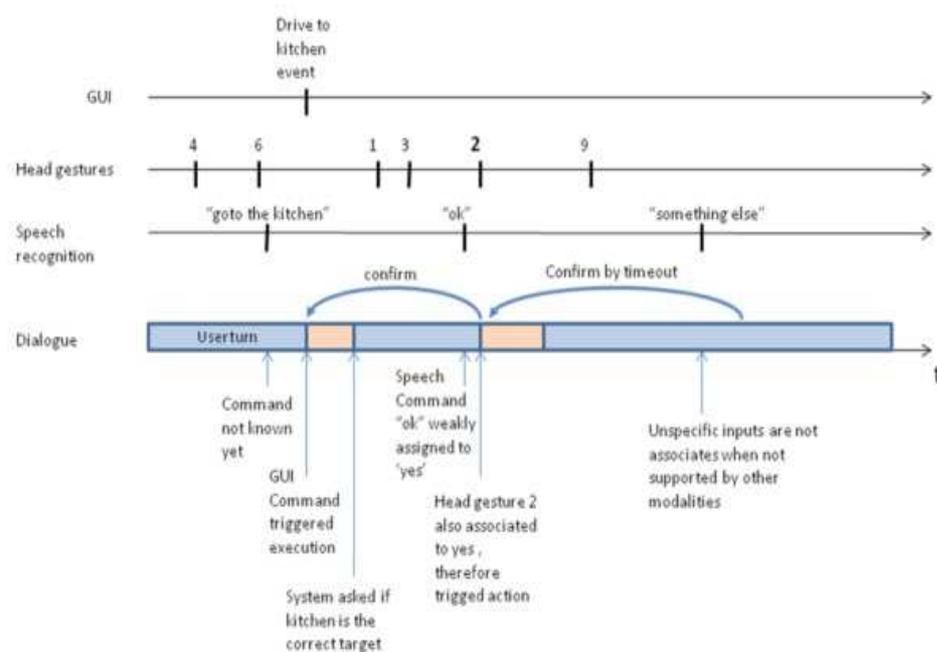
O sistema de interação multimodal adaptativo está embarcado em um robô autônomo de companhia destinado especialmente a idosos com deficiência cognitiva (STEFFEN, 2011). Destaca-se pela fusão de entradas adaptativas e interpretação para entrada multimodal de diálogo, integrando probabilisticamente diversas modalidades de entrada incertas e proporcionando aprendizagem automática de semânticas em entradas anteriormente desconhecidas.

Os autores frisam que o sistema de gerenciamento de diálogo é essencial para a aceitação e usabilidade do robô, enfatizando a comunicação natural e intuitiva. A adaptatividade seria, segundo eles, o principal aspecto de um sistema intuitivo, que deve ajustar-se às capacidades e preferências do usuário, melhorando a comunicação.

O sistema de fusão de entradas pode se adaptar às preferências do usuário (canais de comunicação preferenciais) assim como às especificidades, como jeito pessoal de expressão. Para a interpretação das entradas, significados semânticos são inferidos dos sinais de comunicação do usuário. Isso demanda modelos de relação entre entrada e significado, que podem ser genéricos, mas também adaptados para refletir um usuário específico. Preferências do usuário a respeito de certos canais de comunicação são aprendidas e consideradas, resolvendo ambiguidades ou percepções potencialmente contraditórias.

Um comportamento proativo exige que tempo e situação devem ser escolhidos para abordar o usuário, adaptação que pode ser adotada de acordo com suas preferências. As saídas podem oferecer muitas possibilidades de adaptação, incluindo a escolha de canais de saída, parâmetros como voz e tom de voz, ou mesmo o layout da saída gráfica.

Figura 17: *Speech Act Theory* – fusão de modalidades de entrada



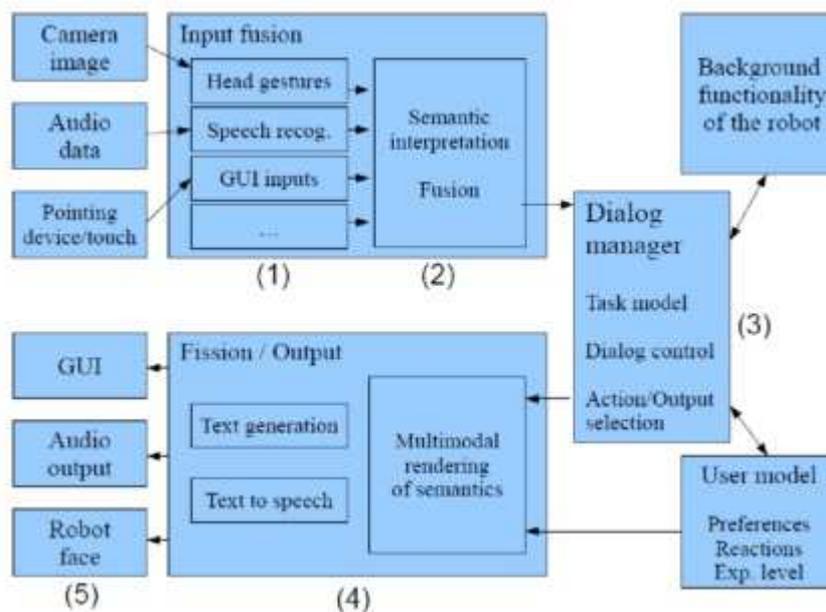
Fonte: STEFFEN, 2011

O desenvolvimento do modelo de gerenciamento de diálogo empregado foi inspirado em *Speech Act Theory*, ilustrado na Figura 17, acima. Quando uma pergunta é feita pelo sistema, para cada ato de entrada representando resposta corresponde um valor de expectativa, o que permite a desambiguação e a escolha da interpretação correta. A confiabilidade indica quando a entrada pode ser admitida/confirmada. A geração de saídas é controlada por uma gramática que cuida de cada ato de saída. Regras gramaticais podem ser ativadas ou suprimidas por meio de condição, que considera a confiabilidade da informação comunicada para confirmar hipóteses implícita ou explicitamente.

Sistema de diálogo multimodal com reconhecimento de voz, entrada GUI via touch screen e gestos de cabeça. Entradas são processadas em paralelo, sendo crucial para a interpretação comandos em paralelo a de sensibilidade ao tempo. Há modelos na literatura que descrevem a combinação de inputs, como o modelo CASE, que classifica as modalidades em a) sequencial ou paralela, B) fusão independente ou combinada. Restam, pois, quatro formas de combinação. Ocorre efeito sinérgico, reduzindo a ambiguidade e aumentando a robustez, quando modalidades independentes são usadas em paralelo.

As entradas semânticas específicas do usuário são aprendidas com base nas interações em curso. O display GUI, bastante confiável, possui uma rotulagem semântica clara. Entretanto, outras modalidades como fala e gestos de cabeça possuem semânticas inespecíficas e são aprendidas quando ocorrem em paralelo com entradas confiáveis ou quando o diálogo pode confirmar sua semântica.

Figura 18: Arquitetura de um sistema multimodal de diálogo para um robô de companhia.



Fonte: STEFFEN, 2011

O foco da pesquisa é a aprendizagem online dos mapeamentos probabilísticos de um usuário específico a partir de detecções específicas de modalidade de entradas semânticas para o sistema de gerenciamento de diálogo. O robô percebe o significado de comandos sem grande entendimento de gramática e sintaxe dos textos de entrada, somente com base em

associações. Esse tipo de modelo de probabilidade substitui a interpretação baseada em gramáticas.

3.8 OfficeMate – Um estudo de um sistema de aprendizagem de diálogo online para robôs assistivos móveis

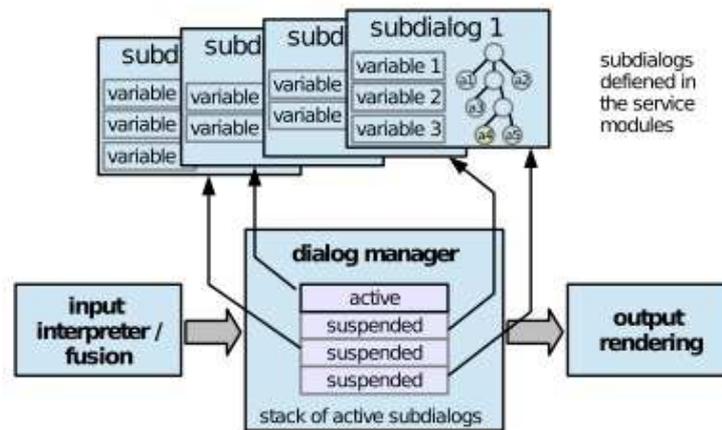
A contribuição do artigo é um planejador de decisão probabilística que implementa a ideia de estratégias de aprendizagem de diálogo online para um robô de serviço móvel em uma interação de longo-termo (MÜLLER; SPRENGER; GROSS, 2014). A experiência e as atitudes dos usuários mudam com o tempo, por isso o robô precisa desenvolver seu comportamento. O sistema de planejamento é parte de um sistema de diálogo multimodal modular e permite uma personalização autônoma dos comportamentos de interação atuais do robô.

Propõe-se à prevenção e assistência a idosos que vivam sozinhos no ambiente doméstico. Esperam o desenvolvimento de um vínculo emocional do usuário com seu robô pessoal com o passar do tempo, o que é reforçado pela habilidade de o sistema se adaptar às preferências e necessidades do usuário. Uma comunicação intuitiva é realizada por uma interface de usuário multimodal que consiste em uma tela sensível ao toque e sensor tátil para carinho. O robô deve aprender as preferências e necessidades do usuário, quais serviços são utilizados em quais situações e quais as atitudes do usuário diante de várias opções que o robô tem em seu comportamento de diálogo. Seleções preferidas são aprendidas pelo robô a fim de aplicar esse conhecimento em interações futuras.

Objetiva uma estrutura para desenvolvimento rápido de aplicações, combinando diálogo multimodal baseado em quadros simples com a capacidade de adaptação opcional sem esforços adicionais de configuração. No estudo, todos os participantes com mais de 4 interações com o robô observaram que ele foi capaz de aprender suas preferências e mudar seu comportamento com o passar do tempo. O robô desenvolveu sequências individuais de sugestão de serviços no menu principal, assim como conjuntos individuais de websites preferidos por cada usuário.

Seu sistema de diálogo, ilustrado na Figura 19, é adaptativo. O gerenciador de diálogo avalia uma árvore de decisão a partir de uma pilha de subdiálogos, cada qual definido por um conjunto de variáveis de estado. Cada subdiálogo é um módulo independente definindo um serviço, cuja configuração se dá principalmente na definição de um estado S (variáveis de entradas de usuário e informações do contexto) e um conjunto de ações disponíveis. Isso se dá pela limitação do conjunto de ações selecionáveis para cada estado S .

Figura 19: Arquitetura do sistema de diálogo.



Fonte: MÜLLER; SPRENGER; GROSS, 2014

À semelhança da implementação proposta no cap. 5 deste trabalho, uma árvore de decisão é elaborada e para cada nó há alternativas que consistem em conjuntos de ações ou sub-árvores. Dessa forma, o gerenciador de diálogo adota um planejamento probabilístico e estratégias de diálogos determinísticas são traçadas, comportando-se nesse caso como uma máquina de estados finita. As sequências de ações consideram o histórico de execução de ações no estado de diálogo. As variações das variáveis podem ser discretas ou valoradas realmente.

3.9 SmartKom – Adaptive and Flexible Multimodal Access to Multiple Applications

É um sistema de interface multimodal genérico no qual o usuário interage com a interface de agente antropomórfico por meio de voz e gestos. É um sistema multimodal destinado a interação com o ambiente doméstico, onde prevê diferentes tipos de canais de entrada para interação do usuário com o sistema. A abordagem uniforme e baseada em conhecimentos do SmartKom (REITHINGER et al., 2003) viabiliza um sistema compreensivo que entende entradas multimodais imprecisas, ambíguas ou incompletas gerando apresentações multimodais coordenadas, coesas e coerentes para três cenários (i) *home* – acompanhante que ajuda a selecionar conteúdo de mídia; (ii) *mobile* – acompanhante de viagem que ajuda na navegação para viagem; e (iii) *public* – acompanhante de comunicação que ajuda a manter o contato e obter informações). São mais de 50 funcionalidades diferentes (tais como seleção de canais de TV, agenda telefônica, assinatura biométrica, seleção de vaga de estacionamento e informação de pontos de interesse), em um total de 14 aplicações (como telefone, e-mail, cinema, TV, mapas e navegação).

A alocação de modalidade adaptada ao usuário foi definida de dois modos, dependendo da atenção do usuário. No modo *lean-forward*, o usuário está prestando atenção ao conteúdo visual do sistema, enquanto no modo *lean-backward* o usuário está executando outra tarefa, vendo televisão, por exemplo.

No cenário Home, o usuário pode dizer ao sistema que não está mais prestando atenção à informação na tela. A informação que o sistema pretende transmitir deve, então, ser expressa verbalmente.

Na consulta “Mostre-me os programas de TV de hoje à noite”, sabendo que a tela é a melhor modalidade para apresentar muitas informações, o sistema apresenta todos os detalhes de forma tabular, deixando para a saída de voz apenas a repetição de detalhes da consulta e um feedback ao usuário sobre a compreensão da entrada.

Quando o usuário ativa o modo *lean-backward*, a resposta à consulta é dada exclusivamente via síntese de voz, mencionando verbalmente apenas os programas mais relevantes. Se o usuário quiser informações adicionais, pode solicitar ao sistema, que fornece os próximos títulos da lista. No display, o agente animado permanece visível para mostrar que o sistema ainda está ativo. No entanto, não exibe informações sobre as consultas efetuados no modo *lean-backward*.

Outra adaptação decorre de estratégias de apresentação pré-definidas pelo desenvolvedor, que decompõe um objetivo de apresentação complexo em tarefas de apresentação. Um exemplo é a estratégia de fissão de mídia, que decide quando uma descrição será feita verbal ou graficamente, baseando-se em restrições na estratégia e na informação no contexto. As modalidades de entrada admitidas são *touchpad*, gestos com caneta e fala. As saídas são síntese de voz e interface gráfica.

3.10 Análise de trabalhos relacionados

Neste capítulo foram relacionadas informações sobre os trabalhos estudados de modo a ampliar o conhecimento do autor sobre o estado da arte e também para a verificação de possíveis lacunas e oportunidades de pesquisa. As três primeiras seções discutem aplicações de robótica assistiva (CONNETTE et al., 2008; FISCHINGER; EINRAMHOF, 2013; MAYER; BECK; PANEK, 2012; PAPOUTSAKIS et al., 2013a; REISER et al., 2009). As demais, analisam modelos ou protótipos multimodais, a maioria adaptativos.

Segundo Meyer, Yakemovic e Harris (1993), certas condições impõem a adaptação do sistema. Os artigos sumariamente descritos nas seções deste capítulo pautam-se por distintos critérios para a adaptação automática de seus sistemas multimodais. O critério de maior incidência é a sensibilidade ao contexto (DJAID; SAADIA; RAMDANE-CHERIF, 2012; GNJATOVIĆ et al., 2012; MÜLLER; SPRENGER; GROSS, 2014), como observado na tabela comparativa exibida na tabela 1.

Tabela 1: Comparação dos sistemas multimodais contidos nos trabalhos relacionados.

Sistemas multimodais adaptativos	Critério de adaptação			Módulos adaptados			
	Necessidade/ aptidão	Preferência	Contexto	Gerenciamento de diálogo	Saídas multimodais	Entradas e saídas	Serviço
Cadeira de rodas com braço manipulativo			X				X
ABB IRB 140			X	X			
Tutor de linguagem de sinais							
Modelo de interpretação adaptativa de entradas		X		X			
OfficeMate			X	X			
Smartkom		X			X		

Fonte: Elaborado pelo autor.

Nesta seção, a adaptatividade de cada trabalho estudado é apresentada. A tabela 1 traz uma comparação entre as principais características referentes a adaptação de cada um deles, relacionadas ao critério de adaptação utilizado e os módulos que serão afetados.

Apesar da previsão de adaptação em razão do contexto, o modelo proposto por Djaid, Saadia e Ramdane- Cherif (2012) é apenas conceitual, não tendo sido de fato implementado, apenas simulada a arquitetura com o uso do conceito de agentes em redes Petri coloridas, estocásticas e temporizadas.

A proposta do robô industrial ABB IRB 140 (GNJATOVIĆ et al., 2012) considera o contexto verbal e espacial para efetuar a adaptação do diálogo. A identificação semântica dos objetos a serem movidos e sua localização no espaço são elementos que permitem ao robô elaborar as informações que transmitirá ao usuário em seu turno de interação, dirigindo-lhe a ação.

O modelo, a ser implementado em um robô real, contido no artigo Adaptive input interpretation for dialogue management of an autonomous robot (STEFFEN, 2011) pontua formas de fusão e interpretação de entradas em um sistema de diálogo multimodal, permitindo a agregação de entradas incertas e fracamente sincronizadas, o aprendizado de entradas anteriormente desconhecidas e o reforço de conhecimentos semânticos incertos de entradas. A adaptação se dá em razão do usuário (preferências e aptidões) a fim de melhorar a comunicação, adequando canais de entrada e saída. O modelo de adaptação probabilístico permite ao sistema de gerenciamento de diálogo a adaptação da fusão de entradas e sua interpretação.

OfficeMate é um sistema de diálogo adaptativo que considera variações de estado no processo de decisão probabilístico que resulta na seleção das ações (MÜLLER; SPRENGER; GROSS, 2014). Dessa forma, para o planejamento e adaptação, o sistema representa o conhecimento do histórico de interações com o usuário. O gerenciador de diálogo elabora um modelo de transição probabilística persistente que representa a probabilidade de se atingir um estado S' dado o estado predecessor S e a ação executada A. Assim, decisões específicas

do usuário e reações são aprendidas assim como restrições internas na sequência de estados, assim o sistema de planejamento não necessita ser configurado previamente. Na prática, contexto e trilhas do usuário são critérios para a adaptação do diálogo e, conseqüentemente, a oferta de serviços.

Em Smartkom, a adaptação da interface depende do nível de atenção do usuário, as quais são decorrem da preferência do usuário, pressupostas no modos de uso *lean-forward* ou *lean-backward* (REITHINGER et al., 2003). A alocação das modalidades adaptadas ao usuário em razão do contexto de uso no cenário doméstico resultam na seleção de canais de saída e adaptação da interação às preferências do usuário.

A principal contribuição deste trabalho é utilizar informações referentes a deficiência do usuário, contidas em seu perfil, para seleção dos canais de entradas e saídas a serem utilizados na interação multimodal. Essa perspectiva não é contemplada em nenhum dos trabalhos estudados. O modelo proposto foi construído visando sua utilização de modo abrangente, não restringindo sua aplicação, e possibilitando uma ampliação de seus componentes no futuro.

4 MODELO PROPOSTO

Este capítulo discorre sobre as justificativas e os componentes do modelo de interação multimodal adaptativa destinado a pessoas com deficiências visuais ou auditivas. A fim de descrever o funcionamento completo do modelo, serão apresentados alguns aspectos gerais e também detalhes do funcionamento de cada um destes módulos.

4.1 Visão geral

O modelo propõe o gerenciamento da interação de um usuário com deficiência e a adaptação das entradas e saídas de dados de acordo com a capacidade de interação deste usuário com os canais de entrada e saída disponíveis. Conforme demonstrado nos trabalhos relacionados (capítulo 3), essa foi uma lacuna identificada nos demais modelos, que atendiam explicitamente um grupo específico de deficientes ou adaptavam seus componentes em razão das preferências dos usuários.

O modelo prevê a utilização desse conjunto de recursos para a utilização de serviços previamente cadastrados. Para o desenvolvimento do modelo, foram consideradas como referência as necessidades de usuários com deficiência auditiva e visual, de acordo com a metodologia empregada pelo IBGE (2012) para esta classificação.

A arquitetura do modelo está representada na notação TAM – *Technical Architecture Modeling* (FMC, 2015). As características da metodologia adotada favorecem a descrição de conjuntos numerosos de elementos e suas interações. Além disso, permite a descrição dos artefatos necessários e suas relações, com clareza e facilidade de identificação.

A arquitetura do modelo se subdivide em 4 módulos principais, conforme a figura 20. Nessa figura, os módulos observam uma sequência lógica de relacionamento, que se inicia no módulo de gerenciamento dos usuários. A ele, segue-se o módulo de tratamento de entradas de dados, depois o módulo de controle de interação e, por fim, o módulo de tratamento das saídas de dados.

De forma preliminar, cumpre salientar que o módulo de gerenciamento de usuários possui como funções principais manter o cadastro de usuários e realizar atividades de identificação do usuário que está iniciando a interação com sistema. O cadastro prévio dos usuários é etapa necessária para a coleta dos dados do perfil de cada um, com o objetivo de posteriormente disponibilizar esta informação nas atividades envolvendo interação.

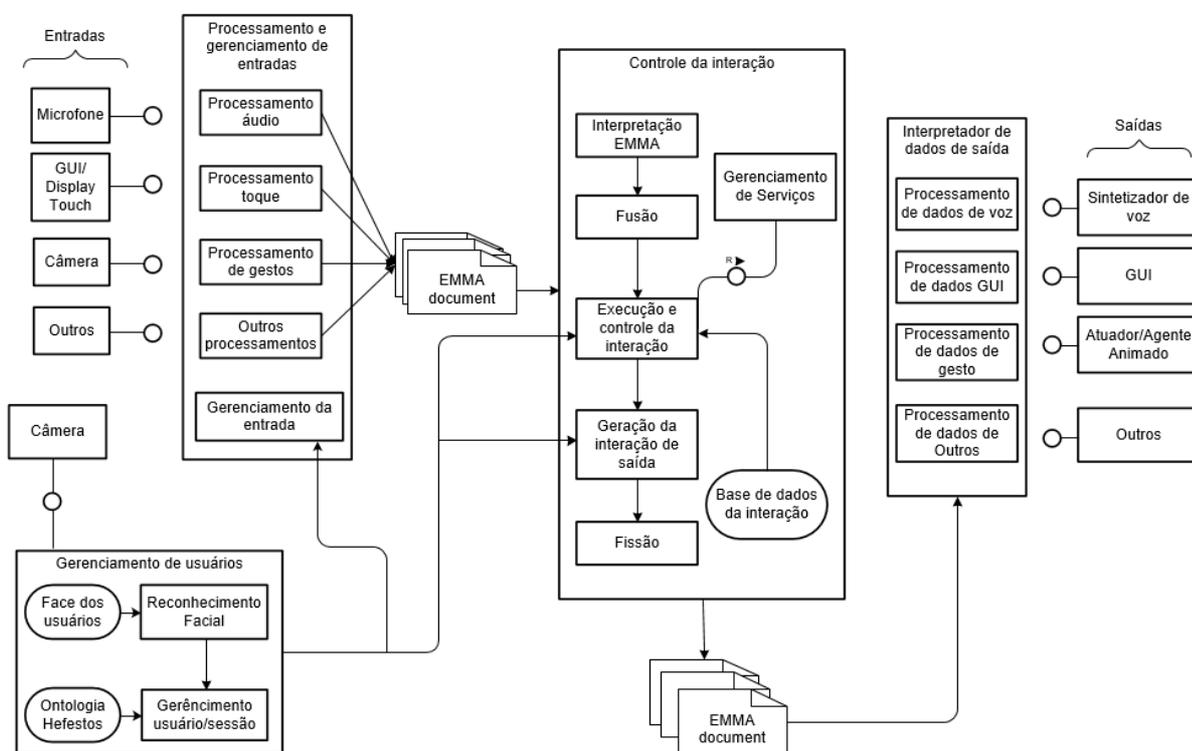
O módulo de processamento e gerenciamento de entradas é responsável por gerenciar a utilização de cada canal de entrada de dados e foi concebido de modo a ser expansível. Novos equipamentos que possibilitem entrada de dados podem ser incluídos, dentro do modelo proposto e adicionados a conjuntos de equipamentos em utilização. Um dos mecanismos adotados para facilitar esta flexibilização foi a definição pelo uso de documentos no formato EMMA para representar os dados obtidos em cada entrada de dados, deste modo mantendo flexível o mecanismo de descrição da interação efetuada pelo usuário.

O módulo de Controle da Interação é responsável por receber e integrar as informações relativas às interações, recebidas nos documentos em formato EMMA. Dessa forma, o módulo executa os tratamentos necessários para a utilização conjunta das diversas

possíveis entradas de dados e, em seguida, gerencia a interação com o usuário, de modo a executar e consumir serviços por ele solicitados. Por fim, esse módulo efetua a criação de documentos, também no formato EMMA, contendo as informações de saída da interação, que são, por sua vez, processadas no último módulo – interpretação de dados para saídas.

O módulo de interpretação de dados de saída realiza a análise dos documentos EMMA recebidos e efetua o encaminhamento das informações necessárias para cada um dos módulos de geração de sinais de saída de dados disponíveis. Esse módulo adota os mesmos conceitos do módulo de entrada de dados, prevendo a inclusão de novos dispositivos sem a alteração do funcionamento do modelo.

Figura 20: Arquitetura do modelo proposto



Fonte: Elaborado pelo autor

A partir da figura 20, podem ser destacados alguns dos diferenciais do modelo proposto, que proporcionam o suporte necessário para o atendimento dos objetivos pretendidos. Um deles é o uso da ontologia Hefestos, empregada no modelo para permitir o detalhamento e armazenamento do perfil dos usuários. Essa abordagem facilita a utilização de elementos de flexibilidade para o posterior gerenciamento das interações, como a adaptação da interação, que pode se beneficiar das informações representadas na ontologia como apoio em atividades de decisão.

Outro elemento a destacar é o gerenciamento de serviços, integrado com o módulo de controle da interação e definido de forma a facilitar a integração de novos serviços, expandindo de forma flexível as capacidades do modelo no atendimento de necessidades dos usuários. Os serviços disponíveis podem ser integrados ao contexto da interação com os usuários a partir de todos os mecanismos de interação disponíveis, tanto para entrada de dados como para geração de dados de resultado a serem exibidos.

A utilização dos documentos EMMA facilita a composição entre os módulos de entrada e saída na sua interação com o módulo de gerenciamento de interação do modelo, desse modo facilitando a sua expansão. A seguir são descritos e analisados em maiores detalhes os módulos do modelo proposto.

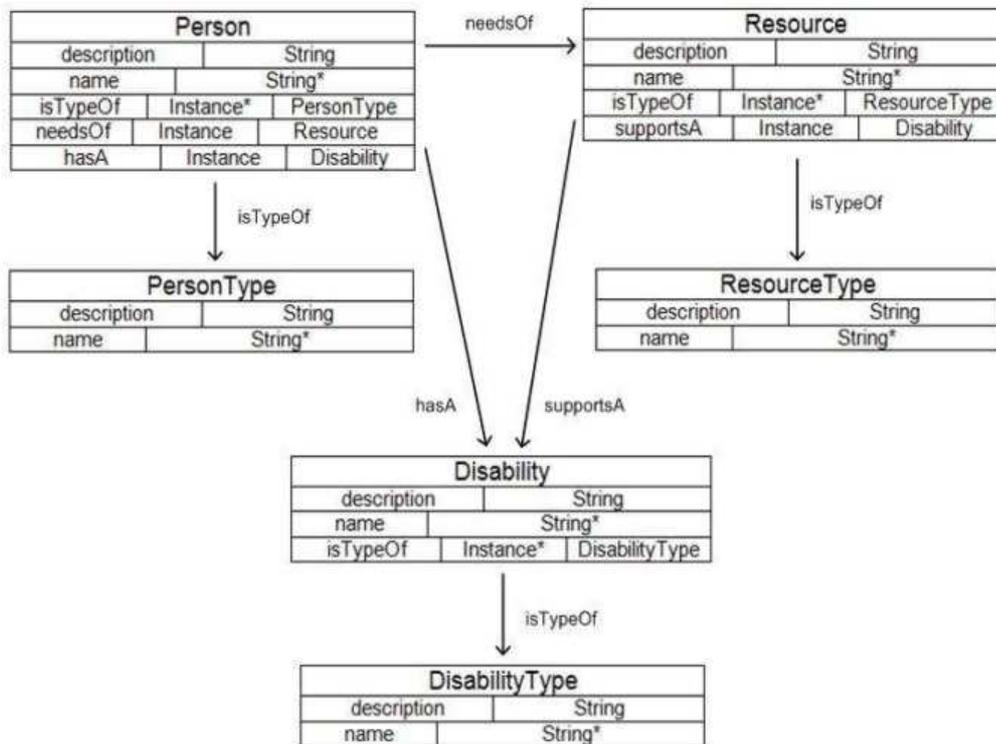
4.2 Módulo de gerenciamento do usuário

Como mencionado anteriormente, este módulo objetiva manter o cadastro dos usuários, apoiar a realização da sua identificação e disponibilizar informações sobre perfil e a sessão de interação em andamento. Dado este cenário, destacam-se como elementos de importância para esse módulo a representação dos dados que podem compor um detalhamento do perfil de cada usuário e dados que apoiem a atividade de identificação.

O modelo prevê a identificação dos usuários por meio de uma interação amigável, com o uso de recursos de identificação biométricos, tais como o reconhecimento facial ou outros. A fim de facilitar essas atividades de interação, uma vez que o modelo será usado em um contexto de robótica assistiva, privilegiou-se a identificação do usuário a partir do reconhecimento facial.

Para representar o perfil dos usuários cadastrados adotou-se uma ontologia voltada à representação de pessoas com deficiência, a ontologia Hefestos (TAVARES, 2011). Empregada com sucesso em projeto voltado para pessoas com deficiência em atividades de apoio à personalização e consulta de dados, está representada na Figura 21. Alguns dos principais elementos da ontologia utilizados neste modelo são os que armazenam informações como chave de identificação, nome, sobrenome, e tipo de necessidade especial do usuário.

Figura 21: Ontologia para acessibilidade do Hefestos



Fonte: TAVARES, 2011

Segundo Tavares (2011), a ontologia Hefestos prevê três classes principais: (1) *Person*; (2) *Resource*; (3) *Disability*. A entidade *Person* representa o perfil da pessoa com deficiência. Esta classe possui o atributo *isTypeOf*, que possibilita categorizar a pessoa. O atributo *hasA* descreve a deficiência da pessoa. O atributo *needsOf* representa o relacionamento com os recursos, que suportam as necessidades da pessoa. A classe *PersonType* é uma subclasse de *Person* que possui atributos como profissionais de saúde, pessoas com deficiência, família e empresas.

Ainda de acordo com esse autor (TAVARES, 2011), a classe *Resource* representa os recursos disponíveis para suportar a acessibilidade do PCD. *ResourceType* é uma subclasse de *Resource*, que possui atributos como tecnologias assistivas, diretrizes, terapias, instituições, construções, etc. A classe *Disability* representa incapacidades, deficiências ou desvantagens de PCDs. A subclasse *DisabilityType* contempla a classificação da deficiência, tal como física, auditiva, visual ou mental.

De forma complementar às informações mantidas na ontologia, uma base de dados mantém imagens da face dos usuários sob a mesma chave de identificação armazenada na ontologia Hefestos. Dessa forma, o reconhecimento facial bem-sucedido permite acessar a chave do usuário, permitindo o carregamento dos dados do usuário da ontologia. Todos os usuários são cadastrados previamente com informações do seu perfil e imagens do rosto.

O processo de identificação do usuário é descrito a seguir. Imagens do usuário são capturadas por meio de uma câmera (disponível no conjunto de equipamentos no módulo entrada), então um algoritmo de reconhecimento facial detecta a face do usuário, comparando-a com aquelas contidas no banco de dados. A correspondência é dada mediante ao algoritmo de reconhecimento facial empregado na aplicação. Detalhes deste algoritmo serão descritos no capítulo de implementação do modelo. Depois de identificado o usuário,

informações de seu perfil mantidas na ontologia do Hefestos são carregadas, permanecendo na memória para utilização pelos próximos módulos e serviços de interação.

Realizado o reconhecimento facial, inicia-se uma sessão deste usuário, a qual será controlada pelo módulo descrito na seção 4.4 deste capítulo. Neste modelo, a inatividade – isto é, a ausência de interação do usuário em um dado lapso temporal – tem como consequência a expiração da sessão, então o modelo retorna ao estado de reconhecimento facial. A fim de se iniciar uma nova sessão, o próximo usuário deve se posicionar em frente à câmera.

4.3 Módulo de processamento e gerenciamento de entradas

O módulo de processamento e gerenciamento das entradas foi definido neste modelo para atuar como o componente que permite a interação com os usuários a partir de um conjunto inicial de equipamentos e recursos, o qual pode ser ampliado mediante a incorporação de novos equipamentos ou recursos de interação.

As entradas do modelo e seu processamento seguem o padrão proposto pela W3C (LARSON et al., 2003) para estruturação da interação multimodal. Neste padrão são consideradas diversos tipos de entrada, cada um através de um dispositivo específico que utiliza um interpretador próprio sendo que as saídas geradas por esses interpretadores são documentos EMMA, que facilitam a integração de várias entradas diferentes.

No modelo proposto, canais de entradas referentes ao reconhecimento de fala, escrita, teclado, dispositivos de toque, imagens, e qualquer outro sensor relevante à aplicação são ativados neste modelo somente se estiverem de acordo com a capacidade de utilização do usuário.

O modelo prevê que cada um dos canais de entrada possuirá um interpretador compatível com os dados que serão recebidos por esse canal. Esse interpretador é responsável pela geração de um documento EMMA, a ser interpretado no módulo de controle. A escolha do padrão EMMA vem ao encontro da proposta de conferir maior flexibilidade ao modelo, facilitando a inclusão de novos equipamentos enquanto mantém um canal de comunicação bem delimitado entre os módulos.

As entradas principais deste modelo são voz, interação por toque e uso de câmera, definidas inicialmente como de maior relevância para suportar as necessidades especiais do público a que o modelo se destina. Também foram considerados os estágios de amadurecimento destas tecnologias e a disponibilidade de equipamentos confiáveis e de custo acessível, bem como de bibliotecas de software para o suporte ao seu uso.

Além destas entradas o modelo prevê um conjunto extensível e flexível, que representa possibilidades de inclusão de novos recursos de captura para entrada de dados. A seguir são descritos aspectos de cada uma destas possibilidades de entrada de dados.

A interação por meio da fala será ativada desde que o usuário tenha capacidade de se expressar verbalmente. Nesse caso, considera-se a situação na qual, durante a interação, o usuário interage verbalmente com o modelo, que utiliza um microfone como canal de entrada de dados. A seguir, o componente interpretador processa a entrada de áudio, utilizando um algoritmo de reconhecimento de fala. O algoritmo utilizado será descrito no capítulo de descrição do protótipo implementado. Após esta etapa cria-se um documento EMMA, que

dispõe textualmente a fala do usuário reconhecida pelo algoritmo de reconhecimento de fala, juntamente com seu grau de confiabilidade. A figura abaixo representa um documento EMMA criado por um canal acústico, com confiabilidade de valor 1.

Figura 22: Exemplo de documento EMMA gerado por uma interação de voz.

```
<emma:emma version="1.0">
<interpretation id="1"
  dialog-turn="dialog-turn1"
  emma:tokens="I want the service number one"
  emma:start="1453348864485"
  emma:end="1453348867485"
  emma:medium="acoustic"
  emma:mode="voice"
  emma:function="dialog"
  emma:verbal="true"
  emma:confidence="1.0"
  emma:uninterpreted="false">
</interpretation>
</emma:emma>
```

Fonte: Elaborado pelo autor.

Para a interação por toque, uma interface gráfica apresenta ao usuário as opções disponíveis para interação naquele momento. O equipamento utilizado para essa forma de interação deve prover capacidade de exibição de dados em um interface gráfica e também possibilitar a indicação/seleção de elementos em uma tela sensível ao toque. O toque no display fornece ao componente interpretador o item selecionado pelo usuário. Caso a área selecionada não possua nenhuma interação referenciada, o interpretador ignora a interação (e.g. uma área em branco na tela). Caso algum item seja selecionado pelo usuário, o interpretador gera um documento EMMA contendo a informação selecionada em formato de texto.

A *câmera* é responsável pela captura de gestos do usuário, os quais podem ser processados de forma complementar a outra modalidade, ou de forma suplementar, paralela, aumentando a robustez e a acurácia do reconhecimento. Gestos ainda podem ser usados como seleção de uma interação, através da representação numérica da interação pela mão do usuário.

A previsão, no modelo, de outros canais de entrada admite como canal de comunicação qualquer tipo de sensor não definido previamente. Essa possibilidade foi incluída no modelo a fim de permitir um escopo maior em sua implementação. Assim como as outras entradas, esse canal está associado a um interpretador que reconhece a interação executada sobre o sensor e cria um documento EMMA representado esta interação.

4.4 Módulo de controle da interação

Como mencionado anteriormente, o modelo prevê a utilização da interação multimodal como forma de indicação de necessidades dos usuários e de facilitação de acesso aos serviços disponíveis. Essa interação objetiva o recebimento de dados do usuário de acordo com suas

possibilidades ou preferências de expressão e, com base nesses dados, identifica o serviço adequado que, ao ser invocado, deve gerar uma saída esperada.

Diversos tipos de interações podem ser considerados no momento da implementação do modelo, tais como simples perguntas e respostas, redes de interação, árvore de interação, diálogo baseado em modelo cognitivo (PRATES, 2015), AIML (WALLACE, 2003) e etc.

Esses serviços podem ser disponibilizados por programas externos relacionados a uma categoria de informação (ou a um conjunto de informações) referente a uma determinada área. A execução desses serviços, utilizando parâmetros relacionados ao próprio serviço e ao usuário, deve retornar dados relevantes ao seu propósito. Citam-se como exemplo dos tipos de serviços considerados: verificação de e-mails do usuário, previsão do tempo, consulta a notas, distância de sua localização até um ponto de interesse, localização de vagas de estacionamento, etc. Não há restrições quanto ao tipo de serviço oferecido ou ao método de implementação. O serviço pode ser uma API, um programa, um webservice ou qualquer outro meio que possa fornecer informações de relevância para seu usuário.

O processo se inicia com o recebimento dos documentos EMMA da interação realizada, quando o controle da interação dá início à interpretação, abrindo o documento e verificando os dados da interação realizada. Nesta etapa, todos os documentos EMMA são interpretados; caso haja mais de um documento, surge a etapa de fusão das interações efetuadas.

Em seguida, a interação é verificada pelo componente de *Execução e controle da interação*. Caso a interação não faça sentido para o sistema, é ignorada. Caso o sistema reconheça como válida a interação, dela deve resultar alguma saída de dados, então dá-se seguimento ao processo de criação de dados para saída. Nesse momento, o módulo de controle da interação pode efetuar a requisição de um serviço, caso algum serviço esteja associado à interação feita pelo usuário.

O serviço tem por objetivo fornecer dados para este módulo, a fim de gerar uma informação válida para o usuário. A Tabela 2 representa o serviço de verificação de e-mails. A interação do usuário, tomada com entrada, consiste na frase “Quantos e-mails não lidos eu tenho na minha caixa?”. A execução e controle da interação irá executar o serviço de Webservice associado para verificação dos e-mails. Com o valor adquirido no resultado do serviço, o módulo de *Execução e controle da interação* irá completar o texto de saída substituindo a variável pela informação adquirida.

Tabela 2: Exemplo de representação de um serviço de verificação de e-mails.

Entrada:	Quantos e-mail não lidos eu tenho na minha caixa?
Saída:	Você possui &variável e-mails não lidos.
Serviço:	Webservice-Email

Fonte: Elaborado pelo autor

Logo, a comunicação entre usuário e sistema é dada em razão da interpretação de documentos EMMA representativos do diálogo, como visto abaixo:

EMMA recebido: Quantos e-mails não lidos eu tenho na minha caixa?

EMMA gerado para síntese de voz: Você possui 5 e-mails não lidos.

Caso ocorra algum erro no serviço, tal como tempo máximo de execução excedido ou sem resposta, o modelo substitui a interação por uma nova mensagem mencionando o erro ocorrido.

Por fim um componente chamado **Geração da interação de saída** reproduz o mesmo processo correspondente às entradas para os componentes de saída. Verificando a deficiência do usuário e seu grau, o componente seleciona todas as saídas de dados pertinentes ao usuário que está utilizando. Essa seleção se dá em razão de um conjunto de regras previamente definidas para cada tipo deficiência e grau. O componente de fissão divide a interação de saída para os canais resultantes e os documentos EMMA são gerados, um para cada interpretador e canal.

4.5 Interpretador de dados para saída

Uma vez gerados os documentos EMMA para cada saída, o interpretador de dados para saída recebe cada um dos documentos EMMA e o destina ao processador correspondente. Se o documento for destinado à síntese de voz, o processamento de dados de voz recebe, interpreta, inicia a síntese de voz, o que resulta numa interação verbal. O mesmo processo é válido para os outros módulos, que recebem o documento relevante a sua saída para ser processado.

A descentralização da interpretação de saída confere ao modelo a flexibilidade necessária à implementação de novos dispositivos de saída, dessa forma a implementação irá depender de um novo interpretador para o respectivo documento EMMA, bem como de um software para externar as informações recebidas como saída.

As saídas previstas no modelo são as de síntese de voz, uma interface gráfica (mesma utilizada para entrada de dados via tela sensível ao toque), um atuador ou agente para representação de gestos e quaisquer outras saídas necessárias à implementação do modelo. Uma vez gerada e apresentada para o usuário a saída, o sistema retorna para o estado inicial, no qual aguarda uma nova entrada do usuário, fechando-se assim o ciclo da interação com o usuário previsto pelo modelo.

5 ASPECTOS DE IMPLEMENTAÇÃO

Para permitir avaliação do modelo proposto, foi necessário dividir o trabalho em duas etapas. Na primeira, visando os requisitos necessários à utilização do conceito de robô assistivo e objetivando sua interação com o ser humano, criou-se um robô humanoide, com traços que remetem ao corpo humano. A segunda consiste na implementação de parte do modelo sobre o hardware utilizado no robô.

5.1 Aspectos gerais

Durante a primeira etapa, destinada à definição de aspectos de hardware e componentes de software a serem empregados, vários conceitos foram considerados e avaliados para o desenvolvimento da estrutura do robô. Seu esqueleto é feito de alumínio e fibra de vidro e as medidas de tórax, cintura e comprimento das pernas correspondem às de um voluntário de 1,70m. Foram pesquisados modelos tridimensionais licenciados pela Creative Commons (INMOOV, 2015; XLROBOT.COM, 2015) e que pudessem ser empregados dentro das finalidades previstas pela licença e em atendimento a esta proposta (uso não comercial e sujeitos à adaptação).

No desenvolvimento do braço e mão, utilizou-se o projeto *InMoov*, (LANGEVIN, 2015), que contém arquivos STLs de modelos tridimensionais prontos para serem impressos em impressora 3D. Para o rosto, foi escolhido o *X1 Opensource Female Robot Companion*, disponibilizado pela XLRobots.com (2016). O modelo 3D da cabeça e da face foram adequados a fim de acoplar uma câmera em seu olho e um pescoço em sua base.

O design do rosto do robô apresenta traços semelhantes à face humana, e foi estrategicamente escolhido a fim de evitar o vale da estranheza (HANSON, 2006). O conceito, elaborado por Mori, sugere que a familiaridade de um robô aumenta de acordo com a semelhança que guarda com a aparência humana, até que certo ponto é atingido e o robô, dadas suas imperfeições, passa a ser repulsivo. Esta queda súbita é referida como o vale da estranheza – *uncanny valey* (SHIMADA et al., 2007).

Para implementação do protótipo do modelo multimodal, decidiu-se por um sistema de interação que fornece dados relacionados ao aluno no escopo acadêmico. Ou seja, um robô com multimodalidade embarcada capaz de fornecer informações sobre o campus e sobre o aluno que está consultando essas informações. Com o protótipo, pretende-se validar a adaptação das entradas e saídas de um sistema multimodal para usuários com deficiência visual ou auditiva. Algumas funcionalidades, tais como a livre movimentação do robô, não foram implementadas, devido ao foco do modelo a ser testado.

5.2 Componentes empregados

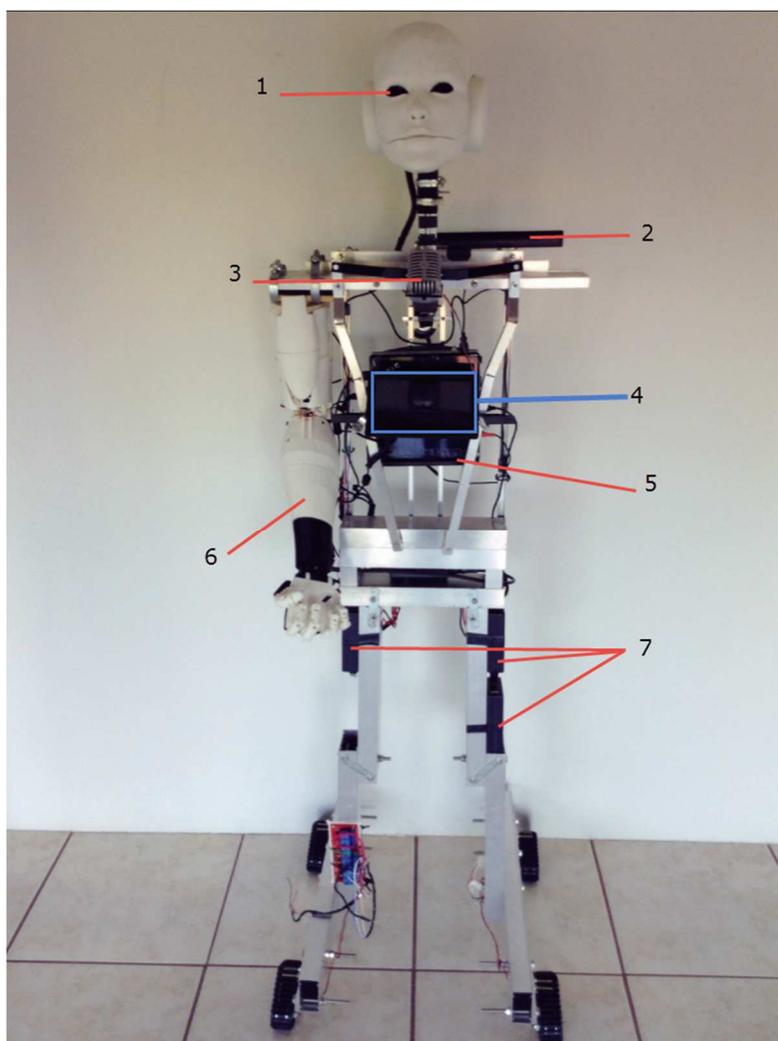
Na montagem do protótipo foram empregados os seguintes hardwares: um microfone omnidirecional destinado ao reconhecimento de fala; um display *touchscreen* modelo CLi07TS

para entrada e saída de dados da interface gráfica; uma câmera USB de 1.2 megapixels, para reconhecimento de usuário; e uma caixa de som para saída de áudio por síntese de voz.

A mão foi construída utilizando 5 servo-motores modelo Futaba S3003, os quais atuam para abrir e fechar cada um dos dedos. Para o controle, utilizou-se um controlador de servo motores sem marca, conhecido comercialmente como *Servo Motor Controller* e disponível em lojas de micro controladores e robótica.

A mão robótica funciona como um atuador e está suspensa em um braço imóvel. Seu emprego está associado a informações numéricas complementares de saída, sendo disparado quando o turno de interação do robô contém dados numéricos de 1 a 5, como nas saídas dos serviços de biblioteca (livros em empréstimo, em atraso ou reservas).

Figura 23: Protótipo funcional.



Fonte: Elaborado pelo autor.

A fim de tornar a interação por voz mais amigável aos usuários com deficiência visual parcial, um *VU Meter (Volume Unit)* foi construído e acoplado à boca, a fim de simular a movimentação da boca por meio da emissão de luz, i.e., o VU Meter ilumina a boca quando alguma palavra é dita pela síntese de voz. Por fim, um GPS modelo GlobalSat BU-353 foi disponibilizado para um serviço específico de distância e pontos de localização. Foi utilizado

um conjunto de 3 baterias de lítio de 12 Volts, modelo YSN-12980, o qual serviu como fonte de alimentação para os servo-motores e o *VU Meter*.

Um computador ZOTAC mini-ITX, com processador Atom D525 (1,8 GHz) e 3 GB de memória RAM, será responsável pela execução do software. Devido ao vasto escopo do trabalho, optou-se pela não implementação de método de processamento de gestos como uma entrada no protótipo.

Na Figura 23, os componentes que integram o protótipo estão numerados de 1 a 7. São eles: 1) câmera; 2) saída de áudio; 3) microfone; 4) display; 5) Zotac; 6) servo-motores; e 7) baterias.

5.3 Gerenciamento de usuário

Foi implementado um módulo de gerenciamento do usuário conforme descrição do modelo e em atendimento ao escopo do protótipo desenvolvido. Foram considerados, portanto, o gerenciamento das informações do usuário pela ontologia Hefestos e o reconhecimento de face e associação da face ao usuário reconhecido.

Para o reconhecimento e gerenciamento de face, dois métodos referentes ao assunto foram empregados. Enquanto o sistema está ocioso aguardando o posicionamento do usuário, executa-se o algoritmo de detecção de face proposto pelos pesquisadores Paul Viola and Michael Jones (2001), o qual emprega padrões de faces pré-treinadas para detectar a face humana em uma imagem. Adotada a biblioteca OpenCV, que já tem esses padrões pré-classificados prontos para o uso.

Detectado o rosto, um segundo algoritmo é iniciado. Esse algoritmo é conhecido como *Eigenfaces*, também implementado pela OpenCV. Seu emprego requer o treinamento da face de cada usuário. Essa etapa objetiva o reconhecimento do usuário em frente ao robô, após o que será possível identificar o usuário cadastrado e recuperar informações sobre ele, inclusive aquelas que se referem aos requisitos para a interação, em observância às suas necessidades e limitações.

Paralelamente aos algoritmos, foi implementada uma tabela de apoio contendo o nome, sobrenome, matrícula e deficiência do usuário. As deficiências catalogadas são: visual total, visual parcial, auditiva total e auditiva parcial.

O objetivo é detectar a presença de um usuário em frente ao robô, e uma vez detectado, iniciar o reconhecimento facial e a interação. O sistema permanece ocioso até uma pessoa parar em frente do robô. A face do usuário é identificada e o algoritmo de reconhecimento de face é iniciado. Se o usuário é reconhecido, seus dados mantidos na tabela de usuários são carregados para memória e o módulo responsável pela interação é ativado, já que um usuário entrou em interação com o sistema ao ser reconhecido. A interação é iniciada por parte do sistema para o usuário. As informações do usuário ficam disponíveis na memória até o usuário se despedir, concluindo a interação, ou se o protótipo constatar inatividade por lapso temporal superior a 1 minuto.

Todos os usuários são previamente cadastrados e suas deficiências vão definir quais entradas e quais saídas serão ativadas para interação com o robô. O módulo de gerenciamento de usuário não controla a adaptatividade do modelo, realizada durante o Processo e gerenciamento de entradas e durante o Controle da Interação. No entanto, o gerenciamento provê as informações necessárias para o sistema adaptar suas entradas e saídas de acordo com cada usuário.

5.4 Implementação dos componentes de Entrada e saída

A implementação do módulo de gerenciamento de entradas e também do módulo de gerenciamento de saída descritos no modelo foi realizada de acordo com o escopo do protótipo desenvolvido. Implica os seguintes componentes: na entrada, há o componente de entrada de dados por voz, o componente de aquisição de imagens por câmera e o componente para interação por toque; na saída estão presentes o componentes para controle da mão robótica, o componente de saída de dados por síntese de áudio e o componente de vídeo por interface gráfica. Todos os documentos gerados no protótipo são derivados do EMMA, sendo uma representação simplificada de sua implementação.

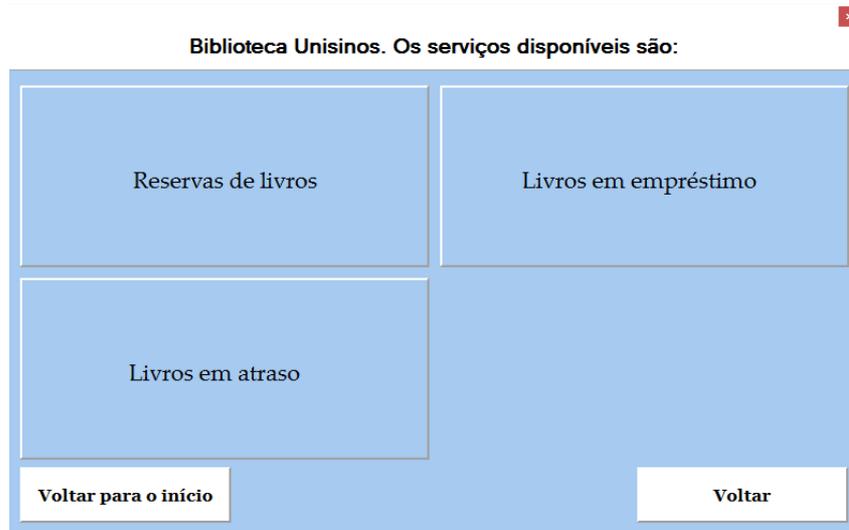
O microfone é empregado como canal de entrada para voz, sendo essa a principal entrada do usuário com deficiência visual. O áudio capturado é processado por meio da API *Coruja 0.2*, desenvolvida pela Universidade Federal do Pará. (BATISTA, 2013), a qual efetua o reconhecimento de voz em língua portuguesa brasileira (pt-br). Testes previamente realizados com esta API, demonstraram a possibilidade de sua utilização no protótipo. No entanto, a API – combinada aos requisitos de hardware do robô – não resulta em um processamento de métodos de linguagem natural plenamente satisfatório.

Por esse motivo, a API empregou um dicionário com 40 palavras e expressões linguísticas previamente definidas para o usuário utilizar durante a interação. Como um recurso de proteção, a fim de evitar o reconhecimento errôneo de palavras na presença de ruídos ambientais, palavras foneticamente semelhantes (como faltas e fraldas) foram adicionadas, (e.g. Faltas e Fraldas) ao dicionário.

Assim, o software de reconhecimento de fala executa em tempo real o reconhecimento da fala, utilizando palavras e expressões pertinentes ao software e de uso comum na língua portuguesa. Logo, quando há uma interação, o módulo de processamento de áudio reconhece a palavra ou expressão e fornece um valor entre 0 a 1, que define a confiabilidade do reconhecimento, criando um documento que contém essas duas informações. Esse documento é posteriormente processado pelo Controle da Interação.

O Display *touchscreen* é a segunda entrada disponível, a principal para usuários com deficiência auditiva. Um software foi criado especificamente para ser executado no Display, que tem tela de 7 polegadas. No display são apresentadas opções que o usuário deve pressionar. Caso o usuário pressione alguma área delimitada por alguma interação, o software do display gera um documento contendo o nome da interação selecionada. A Figura 24 reflete a imagem vista pelo usuário no *display*.

Figura 24: Imagem da interface gráfica disposta no display *touchscreen*



Fonte: Elaborado pelo autor.

O Display também é empregado como saída para usuários com deficiência auditiva ou visual parcial. Para isso, o módulo de interação envia um documento com todas as informações que o Display deve organizar na tela e apresentar ao usuário. O documento segue um mesmo padrão, com marcações do que deve ser visualizado. Abaixo, o documento gerado para a figura 25.

Figura 25: Documento EMMA de saída

```
<titulo> Biblioteca Unisinos. Os serviços disponíveis são: </titulo>
<opcao1> Reservas de livros </opcao1>
<opcao2> Livros em empréstimo </opcao2>
<opcao3> Livros em atraso </opcao3>
<voltar> x </voltar>
```

Fonte: Elaborado pelo autor.

O display exibe até 4 botões referentes a interações disponíveis. Caso a interação seja a última disponível, descrita como um nodo folha da árvore de interações (v. seção 5.4), uma nova marcação é enviada: trata-se do *resultado*. Nesse caso, um novo campo cobre a área dos botões e apresenta o resultado final da interação.

A síntese de voz é o principal canal de saída para os deficientes visuais. Optou-se por utilizar a Microsoft Speech API devido à facilidade de implementação e por se tratar de um protótipo *off-line*. Essa API conta com vozes em língua portuguesa-brasileira, tendo sido escolhida uma voz feminina devido à sua aceitação em sistemas automatizados ser superior à voz masculina (v. seção 2.5). A interação gera um documento contendo unicamente a frase a ser sintetizada. O interpretador lê o documento e ativa a API, verbalizando a saída de áudio.

A mão robótica é a terceira saída disponível na implementação, que consiste em movimentação de até 5 dedos na representação de números. A mão robótica é usada em usuários com alguma capacidade visual, a fim de suplementar a mesma informação fornecida no display ou por síntese de voz. Foi desenvolvida com software específico para o atuador,

sendo empregados comandos seriais para controle dos servo-motores dos dedos. A mão, que permanece inicialmente fechada, é ativada se um documento, contendo um número maior que zero e menor ou igual a cinco, é enviado pela interação. O software abre os dedos da mão, representando o número recebido na *string*, que contém o seguinte padrão #P<numero_da_porta_servo><sinal_TTL>\r\n. Visto que o sinal TTL, necessário para abrir e fechar cada dedo, varia de acordo com a instalação do respectivo servo motor, o comando para representar o número 5 é o seguinte: #1P989#2P750#3P750#4P1044#5P1344T100\r\n.

Dentro do módulo de Processamento e gerenciamento de entradas, há um componente importante para a implementação, denominado Gerenciamento da Entrada. Este componente é responsável por ativar e desativar os canais de entradas para cada tipo de usuário.

Tabela 3: Modalidades de entradas ativadas conforme perfil de usuário (tipo e grau de deficiência)

Deficiência:	Entradas	
	Reconhecimento de fala	Display Touch
Auditiva_Parcial	X	X
Auditiva_Total		X
Visual_Parcial	X	X
Visual_Total	X	

Fonte: Elaborado pelo autor.

O Gerenciamento da Entrada utiliza as informações referentes ao perfil do usuário fornecidas pelo Gerenciamento de Usuários e mantidas na memória. Com um conjunto de regras baseado nas quatro categorias de deficiências pré-definidas, o Gerenciamento da Entrada seleciona como entrada o Reconhecimento de Fala e/ou Display. É possível verificar essa seleção na tabela 3, acima.

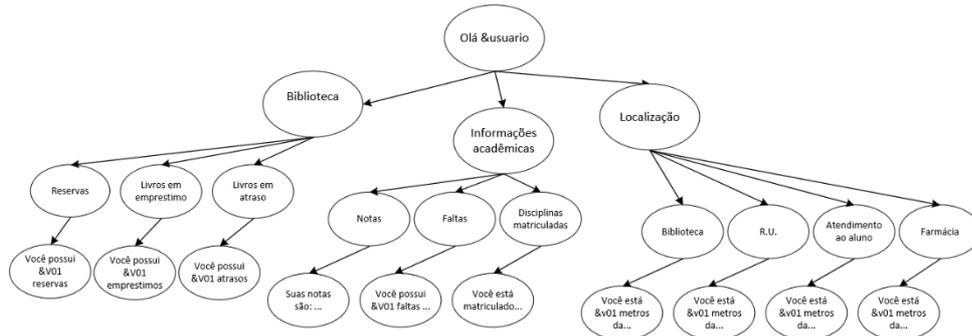
5.5 Implementação do Controle da Interação

Definimos interação, para fins de descrição deste contexto de implementação, como sendo o processo de o usuário se comunicar com o protótipo e o protótipo fornecer retorno como resposta. O módulo de controle da interação tem por objetivo gerenciar esse processo que envolve verificar a requisição do usuário, avaliar o seu perfil, invocar o serviço desejado e informar os dados requisitados.

Neste módulo, considera-se o processo de interação em três grandes etapas. Na primeira, estão envolvidas as atividades de receber o documento gerado em uma comunicação proveniente de um canal de entrada e validar a existência da disponibilidade da informação no atual contexto. Na segunda etapa, estão previstas atividades de análise da estrutura de interações e criação de uma resposta para o usuário. Finalmente, a terceira etapa consiste na seleção do canal de saída para o usuário que está utilizando o sistema e geração de um ou mais documentos para os canais de saída. No decorrer desta seção serão detalhadas as etapas acima referidas.

A previsão de um conjunto de interações com perguntas passíveis de serem feitas pelo usuário e possíveis respostas do robô resultou numa estrutura de árvore de interação com entradas e saídas. A estrutura definida segue o padrão de uma árvore, com raiz, nodos pais e filhos, conforme a Figura 26. Apesar do modelo considerar recursos mais flexíveis como possibilidade (tal como a ontologia Hefestos e sua ampliação), na implementação do protótipo foi definida especificamente com objetivo de otimização do processo de implementação e teste, uma estrutura similar porém mais simples.

Figura 26: Representação gráfica da estrutura de interações



Fonte: Elaborado pelo autor

Esta estrutura descrita na figura 27 pretende resumir as possibilidades de interação em agrupamentos gerais, como pode ser observado em relação ao primeiro nível da árvore, onde estão localizados aspectos de serviços disponíveis (biblioteca, informações acadêmicas, localização). Associados a cada serviço estão opções detalhadas e exemplos (não exaustivos) de perguntas possíveis de serem realizadas.

Essa árvore de interação é armazenada e representada em um XML, que mantém a estrutura apresentada na figura 27.

Figura 27: XML com a definição da árvore de interações.

```
<?xml version="1.0" standalone="yes"?>
<DATAPACKET Version="2.0">
  <METADATA>
    <FIELDS>
      <FIELD attrname="ID" fieldtype="i4"/>
      <FIELD attrname="Input_Audio" fieldtype="string" WIDTH="200"/>
      <FIELD attrname="Input_Display" fieldtype="string" WIDTH="200"/>
      <FIELD attrname="Output_Audio" fieldtype="string" WIDTH="200"/>
      <FIELD attrname="Output_Display" fieldtype="string" WIDTH="200"/>
      <FIELD attrname="Output_Gesto" fieldtype="string" WIDTH="30"/>
      <FIELD attrname="Filhos" fieldtype="string" WIDTH="200"/>
      <FIELD attrname="Servico" fieldtype="string" WIDTH="50"/>
      <FIELD attrname="Parametros" fieldtype="string" WIDTH="50"/>
      <FIELD attrname="Timeout" fieldtype="i4"/>
      <FIELD attrname="Descricao" fieldtype="string" WIDTH="100"/>
      <FIELD attrname="InteracaoInterna" fieldtype="string" WIDTH="10"/>
    </FIELDS>
  </METADATA>
</DATAPACKET>
```

Fonte: Elaborado pelo autor

A seguir, seguem comentários sobre cada um dos campos exibidos na figura 27. Na figura 27, o campo *ID* refere-se a um identificador único para cada nodo da árvore. *Input_Audio* é uma *string* com a sentença ou palavra que corresponde à interação com aquele nodo, proveniente do reconhecimento de fala. O campo *Input_Display* é uma *string* com a interação correspondente ao nodo cujo valor é proveniente do display. *Output_Audio* é uma

string com as informações que serão verbalizadas pelo sintetizador de voz. *Output_Display* é o campo referente à *string* que será apresentada no *display*. *Output_Gestos* é o valor numérico que será representado pela mão robótica.

Ainda na figura 27, o campo *Filhos* identifica todos os nodos filhos do nodo atual. *Servico* é o campo que mantém o nome do programa a ser executado durante a interação. *Parametros* são todos os parâmetros referentes ao serviço. *Timeout* é o tempo máximo de espera para que o serviço seja finalizado. *Descricao* contém a descrição da interação de cada nodo. *InteracaoInterna* campo utilizado a fim de realizar uma interação dentro da navegação da árvore (e.g. usuário: - Voltar para o início; o estado da árvore volta para o início)

Quando há uma interação feita pelo usuário, o módulo de *Controle da interação* recebe o documento com a interação feita e de onde ela veio (pelo reconhecimento de fala ou pelo *Display*). Nesse momento, ocorre a interpretação do documento e da interação feita pelo componente de *Interpretação*. Em seguida, o componente *Execução e controle da interação* verifica se algum dos nodos filhos possuem essa interação no seu campo de *Entrada*. Se a interação veio do reconhecimento de fala, o componente procura no campo *Input_Audio*, se a interação veio pelo *Display*, o componente procura no campo *Input_Display*. Caso positivo, o componente verifica se algum serviço está associado ao nodo, o executa e espera o valor ser retornado. Mais detalhes sobre o serviço serão expostos na próxima seção.

O próximo componente a ser descrito é o componente Geração da interação de saída, o qual, graças a um conjunto de regras, seleciona as saídas pertinentes para o usuário e captura a informação dos campos *Output*. Em seguida, os valores gerados pelo serviço são associados às interações de saída e os documentos são criados. É possível descrever essa seleção de saída utilizando a Tabela 4.

Tabela 4: Modalidades de saídas selecionadas conforme o perfil de usuário (tipo e grau de deficiência)

	Saídas:		
	Síntese de voz	Display	Gestos
Deficiência:			
Auditiva_Parcial	X	X	X
Auditiva_Total		X	X
Visual_Parcial	X	X	X
Visual_Total	X		

Fonte: Elaborado pelo autor.

A seguir são comentados os aspectos do componente de Gerenciamento de serviços. A interação possibilita o usuário a navegar por toda a estrutura da árvore, mas todas as informações fornecidas pelo controle da interação são estáticas, ou seja, independente do usuário que está usando, toda informação será constante. Para solucionar essa situação, utilizam-se serviços nas interações. Os serviços são programas que retornam alguma informação sobre um assunto referente ao usuário. Dessa forma, quando o usuário interage com um nodo que tem associado um serviço, a interação de saída será personalizada.

Como exemplo, na figura 28, é possível observar um trecho de um diálogo gerado pelo protótipo.

Figura 28: Diálogo entre robô e usuário

Robô: Olá, Edward. Diga o serviço desejado
 Usuário: Biblioteca
 Robô: Biblioteca. Diga o que você deseja saber. Reservas, livros em empréstimo ou livros em atraso.
 Usuário: Livros em atraso
 Robô: Você possui 2 livros em atraso. São eles: Vinte mil léguas submarinas e Da Terra à Lua.

Fonte: Elaborado pelo autor.

No exemplo descrito na figura 28, o robô pergunta o serviço desejado e o usuário responde “Biblioteca”. O nodo da estrutura de dados interna que representa o serviço da biblioteca não possui nenhum serviço previamente selecionado, logo a segunda frase do robô será igual para todos os usuários. Já a interação do usuário indicando a frase “Livros em atraso” contém uma associação para um serviço descrito na estrutura interna, que será invocado pelo robô, o qual retornará a quantidade de livros em atraso e o nome de cada livro atrasado. Neste exemplo, caso o usuário fosse categorizado como com deficiência visual parcial, a mão robótica iria representar o número 2 em conjunto com a fala.

O serviço associado ao nodo “Livros em atraso” se chama “Biblioteca.exe” e seu parâmetro é “Atrasos”. O programa é executado com o parâmetro mantido no nodo e o serviço tem livre acesso aos dados associados ao número de matrícula do usuário, chave essa fornecida pelo módulo de Gerenciamento de usuário e mantida na memória. Assim, o programa gera um documento com os exatos valores de quantidade de livros em atraso e nomes dos livros. Este mesmo recurso pode ser usado para acessar, por exemplo, com conjunto de serviços web ou de APIs, portanto com uma ampla flexibilidade de acesso aos serviços desejados em outras situações. O contexto da implementação do protótipo foi reduzido neste aspecto, por conta de restrições de tempo disponível, mas de modo a manter o cenário para uma futura expansão.

Visto que o protótipo objetiva o atendimento de usuários universitários, todos os serviços listados na Tabela 5, abaixo, referem-se à vida no campus.

Tabela 5: Tabela de serviços implementados

Programa	Parâmetro	Utilidade
Biblioteca.exe	Reservas	Informações referentes às suas reservas de livros.
Biblioteca.exe	Empréstimos	Informações sobre livros emprestados em seu nome.
Biblioteca.exe	Atrasos	Informações referentes a livros em atraso em seu nome.
DiscFaltas.exe	Disciplinas	Informa o nome das disciplinas matriculadas neste semestre.
DiscFaltas.exe	Notas	Informa as notas finais de cada disciplina.
DiscFaltas.exe	Faltas	Indica as faltas do aluno em cada disciplina matriculada.
GPS.exe	RU	Distância até o RU
GPS.exe	Atendimento Aluno	Distância até a Central de Atendimento ao aluno

GPS.exe	Farmacia	Distância até a farmácia
GPS.exe	Biblioteca	Distância até a biblioteca

Fonte: Elaborado pelo autor.

Todo nodo que utiliza um serviço para prover informações possui um *timeout* em milissegundos referente ao tempo máximo de execução esperado. Caso exceda o tempo, componente Execução e controle da interação fornece uma mensagem de erro para o fluxo de execução do modelo, que é apresentado para o usuário.

5.5.1 Interpretador de dados para saída

Este módulo apenas abre os documentos gerados pelo módulo de controle e interação e cria uma saída para canal correspondente. Se um dos documentos for destinado a saída de dados por voz, o conteúdo do documento é aberto e a *string* dados é verbalizada, utilizando MS Speech Synthesis da Microsoft. Se o documento for para interface gráfica, o *processamento de dados GUI* monta a tela com título e botões para o usuário. Se o documento se destinar a gestos, o processamento de dados de gesto abre o documento, verifica se o dado é numérico e envia um comando serial para a porta do controlador dos servos da mão, gesticulando o valor de 1 a 5.

5.6 Sessões de utilização

Numa sessão típica de utilização, o usuário precisa ser cadastrado previamente pelo administrador. Utilizando dois programas específicos para essa função, o administrador informa os dados referentes ao perfil do usuário, tais como matrícula, nome, sobrenome e deficiência. Após este cadastro, o usuário posiciona-se em frente ao robô para iniciar o cadastramento facial, no qual, cerca de 10 fotos do rosto são capturadas e associadas ao número de matrícula fornecido.

Após o cadastro, o usuário já é reconhecido pelo robô, que inicia uma interação com uma mensagem de boas-vindas e solicitando o serviço desejado pelo usuário. Os canais de entrada e saída serão ativados de acordo com a deficiência do usuário, que iniciará sua interação por fala ou toque. Após a utilização do sistema, o usuário se despede do sistema ou se ausenta da presença do robô, que finalizará a sessão iniciada para aquele usuário. Na Figura 29, um usuário com deficiência visual completa se comunica com o protótipo. Como é possível notar, o display está ativo, mas nenhuma informação é exibida na tela e toda a interação é verbalizada.

Figura 29: Sessão de usuário com deficiência visual completa



Fonte: Elaborado pelo autor.

6 AVALIAÇÃO

Este capítulo retoma a metodologia de avaliação prevista na seção 1.4 deste trabalho, cuja aplicação permite verificar aspectos de usabilidade, como facilidade de uso, percepção de utilidade e satisfação, considerando as modalidades selecionadas adaptativamente para cada tipo de usuário durante a execução do software.

Após os testes, os usuários participantes avaliaram o sistema de interação multimodal adaptativo embarcado em robótica assistiva implementado por meio da realização de questionários estruturado (12 questões com respostas no padrão Likert e 1 questão que admite sim ou não) e semi-estruturado (5 questões subjetivas). As notas de campo do pesquisador são fonte de informações que permitirão a interpretação e análise dos dados obtidos por meio dos questionários.

6.1 Metodologia adotada

O objetivo dessa avaliação é verificar aspectos de usabilidade, como facilidade de uso, percepção de utilidade e satisfação, considerando as modalidades selecionadas adaptativamente para cada tipo de usuário durante a execução do protótipo.

Segundo Malinowski (MALINOWSKI et al., 1992), a principal forma de avaliação de um sistema multimodal adaptativo consiste na aferição da usabilidade, daí a decisão metodológica de submeter os participantes deste estudo a um questionário de usabilidade. Os participantes foram submetidos ao final dos testes a um questionário estruturado com 13 questões e a um semi-estruturado, que permite respostas objetivas, com 5 questões.

O Modelo de Aceitação de Tecnologia – TAM (do inglês, *Technology Acceptance Model*), tradicionalmente empregado na avaliação de aceitação de novas tecnologias (TAVARES et al., 2015) foi empregado a fim de embasar a elaboração das questões. Esse modelo mede a satisfação do usuário por meio da percepção de utilidade e de facilidade de uso, sendo que a primeira implica que a tecnologia possibilita ao usuário o desenvolvimento de um trabalho melhor, enquanto a segunda avalia se a tecnologia pode ser empregada com um mínimo de esforço. 5 questões sobre facilidade de uso do sistema e 5 questões sobre percepção de utilidade foram incluídas no questionário.

De acordo com a norma ISO 9241-11, usabilidade é "a capacidade de um produto ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso" (GONÇALVES, 2004). A fim de verificar a eficácia do processo de interação e a satisfação do usuário em termos conforto e aceitabilidade, foram elaboradas outras 3 questões objetivas, a fim de quantificar opiniões expressas subjetivamente pelos usuários.

Todas as questões do questionário estruturado (exceto a última, que prevê resposta sim ou não por tratar da eficácia do sistema) adotam como padrão de resposta a escala de Likert, um tipo de escala de resposta psicométrica comumente adotada em questionários por permitir aos respondentes a especificação de seu nível de concordância, que vai da concordância plena à discordância total, contemplando estados parciais e a neutralidade – nem concorda/nem discorda. Uma vantagem do uso desta escala é que ela fornece direções

sobre a atitude do respondente em relação a cada afirmação.(SILVA; PIMENTEL; SOARES, 2012).

Notas de campo tomadas pelo pesquisador durante a fase de testes e após a realização dos questionários também serviram como fonte para a avaliação dos resultados (seção 6.4).

6.2 Perfil dos participantes

Quatorze voluntários participaram dos testes após assinarem termo de consentimento livre e esclarecido aprovado pelo Comitê de Ética em Pesquisa da Universidade do Vale do Rio dos Sinos – UNISINOS.. O material sobre este termo encontra-se no apêndice A deste trabalho. Os voluntários foram divididos em 3 (três) grupos, resultando em 17 (dezesete) usuários-participantes, uma vez que 3 (três) voluntários com deficiência parcial submeteram-se também a testes no qual simularam deficiência total, procedimento semelhante ao adotado recentemente por Uluer, Akalin e Köse (2015) em experimentos com o robô tutor da língua de sinais turca.

Para os participantes com deficiência auditiva, a capacidade de leitura em língua portuguesa foi requisito de seleção. Dos 6 (seis) participantes com deficiência auditiva, 4 (quatro) apresentavam escolaridade inferior ao nível médio e tinham a língua brasileira de sinais como língua materna (primeira língua), diferentemente das demais, que possuíam pelo menos o ensino médio completo e pleno domínio (proficiência) da língua portuguesa. Os outros 2 (dois) deficientes auditivos eram proficientes em língua portuguesa, não faziam uso da língua de sinais e possuíam escolaridade igual ou superior ao nível médio. Desses, 1 (um) possui deficiência auditiva severa e outro apresenta comprometimento parcial da audição.

A faixa etária dos participantes varia entre 14 (quatorze) e 60 (sessenta) anos e sua seleção considerou critérios empregados pelo IBGE no Censo 2010 para definição de deficiência permanente visual ou auditiva e grau de severidade: a autopercepção de dificuldade em enxergar ou ouvir, mesmo com o uso de facilitadores como óculos, lentes de contato e aparelho auditivo. Assim, é severa a deficiência visual ou auditiva de quem declara ter grande dificuldade ou não conseguir ver ou ouvir (IJJJ, 2012).

Componentes demográficos impactaram o resultado da avaliação dos participantes com deficiência auditiva, como se verá na análise elaborada adiante (subseção 6.4.3); estatisticamente, foi insignificante a correlação entre escolaridade, idade e forma de avaliação dos participantes com deficiência visual.

6.3 Procedimentos de teste

14 (quatorze) voluntários com diferentes tipos de deficiência participaram dos testes, divididos em 3 (três) etapas. Após a leitura e assinatura dos termos de consentimento livre e esclarecidos aprovados pelo Comitê de Ética em Pesquisa da Universidade do Vale do Rio

dos Sinos – Unisinos, os participantes foram cadastrados. Submeteram-se a procedimento de reconhecimento de face, mediante a captura de 8 imagens, e prestaram informações para seu perfil, como tipo de deficiência, grau e nome completo.

A primeira etapa contemplou a interação de 8 (oito) usuários cadastrados sob o perfil “deficiência visual completa”. 2 (dois) desses usuários realizaram o procedimento de forma simulada, como indicado na seção anterior, e foram selecionados para participar da terceira etapa, descrita adiante. Num segundo momento, o protótipo foi testado com 6 (seis) usuários cadastrados sob o perfil “deficiência auditiva completa”, sendo que 1 (um) deles realizou esse teste de forma simulada e foi voluntário também na terceira etapa.

A terceira etapa consistiu no reenquadramento de perfil de 3 usuários, a fim de viabilizar uma experiência de interação multimodal suplementar, associando modalidades de canais na entrada e na saída (áudio e display). Participaram dessa experiência 2 (dois) usuários com deficiências visuais parciais – visão monocular, isto é, visão completamente comprometida num dos olhos, somada a perdas significativas, mas não impeditivas da capacidade de leitura no outro olho – e 1 usuário com deficiência auditiva parcial – comprometimento da audição em 80% num ouvido e 50% no outro.

Na primeira bateria de testes, usuários com deficiência visual receberam breves explicações sobre as funcionalidades do sistema e foram esclarecidos sobre o uso de comandos simples de interação por voz. Caso o usuário não entendesse o que foi falado pelo robô, poderia usar o comando de voz “repetir”, então o robô repetiria a última informação; se perdido, poderia recorrer ao comando “preciso de ajuda”, que o robô mencionaria todas as interações disponíveis para o usuário naquele momento; “Voltar para o início” implicaria o recomeço da interação, enquanto os comandos “Obrigado” e “Voltar” restauraria o menu imediatamente anterior.

A abordagem do segundo grupo de testadores, com deficiência auditiva, deu-se por meio de um intérprete, que esclareceu os participantes quanto ao termo de consentimento e transmitiu as informações descritas acima por meio da linguagem brasileira de sinais – LIBRAS. O terceiro grupo era formado por pessoas com deficiências menos severas, as quais participaram das etapas 1 e 2 simulando maior severidade. Como na interação anterior restringiram-se ao canal de interação adequado à natureza de sua deficiência, lhes foi informado que nesta fase a interação mediante fala e display estariam ambas ativas, sendo livre o uso de qualquer modalidade de acordo com o interesse, a conveniência ou o que julgasse mais adequado ao contexto de uso.

Após as devidas instruções, cada usuário utilizou livremente o robô por até 15 minutos. Depois da exploração inicial, deveria – sem ajuda externa do pesquisador – interagir com a intenção de obter as informações relativas a livros em atraso (cenário).

6.4 Análise dos questionários

O padrão de respostas de todos os grupos foi consolidado e consta na Tabela 6. Em todas as questões, o percentual de participantes que respondeu as alternativas “Concordo parcialmente” ou “Concordo plenamente” foi de, pelo menos, 70%. Esta visão geral está apresentada como um ponto de partida para uma análise mais detalhada, que será

apresentada em seguida, nessa seção. Mas inicialmente cabe destacar o aspecto positivo observado com este padrão de respostas.

Tabela 6: Distribuição das respostas dos participantes do estudo

<u>Facilidade de uso</u>	Discordo plenamente	Discordo parcialmente	Não concordo nem discordo	Concordo parcialmente	Concordo plenamente
1 - A aplicação é fácil de usar	0,0%	11,8%	0,0%	29,4%	58,8%
2 - A aplicação é fácil de entender	0,0%	5,9%	0,0%	41,2%	52,9%
3 - Com pequeno esforço posso encontrar minhas informações acadêmicas	5,9%	0,0%	0,0%	23,5%	70,6%
4 - Com pequeno esforço posso encontrar minhas informações sobre minha relação com a Biblioteca	5,9%	0,0%	17,6%	11,8%	64,7%
5 - As opções são claras e objetivas.	5,9%	0,0%	5,9%	11,8%	76,5%
<u>Percepção de utilidade</u>					
5 - As opções são claras e objetivas.	5,9%	0,0%	5,9%	11,8%	76,5%
6 - As opções de interação são relevantes	0,0%	0,0%	0,0%	47,1%	52,9%
7 - O robô facilita acessar com meus dados acadêmicos	0,0%	5,9%	11,8%	23,5%	58,8%
8 - O robô facilita a obtenção de informações referente ao campus	0,0%	11,8%	17,6%	23,5%	47,1%
9 - O robô facilita encontrar pontos de interesse no campus	0,0%	11,8%	17,6%	52,9%	17,6%
10 - Se eu precisasse dessas informações, eu usaria o robô	0,0%	0,0%	0,0%	23,5%	76,5%
<u>Conforto e eficácia</u>					
11 - Senti-me confortável ao interagir livremente com a aplicação	0,0%	11,8%	0,0%	11,8%	76,5%
12 - Senti-me confortável ao desempenhar a tarefa solicitada pelo pesquisador	11,8%	11,8%	0,0%	11,8%	64,7%
13 - Consegui desempenhar a tarefa solicitada	Não	Sim			
	11,8%	88,2%			

Fonte: Elaborada pelo autor.

É possível resumir as informações contidas na tabela 6 acima se realizado o agrupamento temático das questões. O resultado é exibido na tabela 7. As questões 1 a 5 foram reunidas sob o rótulo “Facilidade de Uso”, de 6 a 10, como “Percepção de Utilidade” e 11 a 13 como conforto e eficácia, tendo sido incluídos os percentuais referentes às respostas à questão 13 nos dois extremos da tabela (“não” em “discordo plenamente” e “sim” em “concordo plenamente”).

Esta tabela 7 permite que sejam observados com maiores detalhes os resultados em cada uma das dimensões, o que motivou a geração dos elementos seguintes para apoio da avaliação.

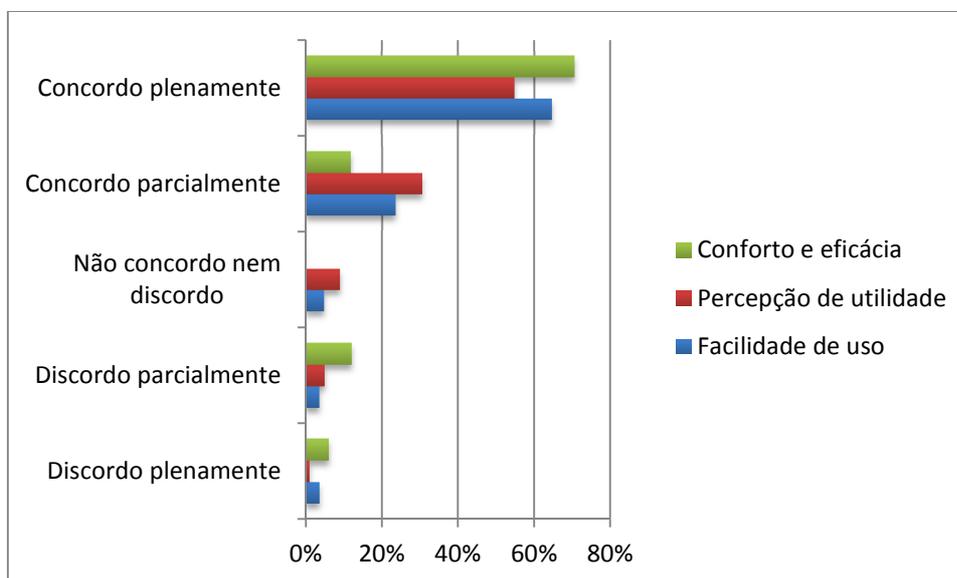
Tabela 7: Respostas dos participantes do estudo distribuídas por tema.

Tema	Discordo plenamente	Discordo parcialmente	Não concordo nem discordo	Concordo parcialmente	Concordo plenamente
Facilidade de uso	3,5%	3,5%	4,7%	23,5%	64,7%
Percepção de utilidade	1,0%	4,9%	8,8%	30,4%	54,9%
Conforto e eficácia	5,9%	11,8%	0,0%	11,8%	70,6%

Fonte: Elaborada pelo autor.

O Gráfico 1, abaixo, representa a distribuição temática das respostas de todos os participantes do estudo. Nele pode ser observado que existe uma grande tendência a uma avaliação positiva do protótipo.

Gráfico 1: Distribuição das respostas de todos os participantes por eixo temático



Fonte: Elaborado pelo autor.

As próximas subseções, que introduzem os resultados de cada grupo e trechos de respostas do questionário subjetivo, permitirão conclusões mais específicas sobre aspectos positivos e também sobre limitações do trabalho, em razão das quais são sugeridas melhorias em trabalhos futuros (ver seção 7.2).

6.4.1 Deficientes visuais

Dada a ausência de entonação, um usuário sugeriu que o robô assinalasse com um bipe o início do turno do usuário, indicando o momento a partir do qual deve falar. Três dos seis participantes que sugeriram melhoria recomendaram que o sistema fosse ajustado de forma a contemplar seleção de item de menu antes de finda a descrição das opções.

Seguem exemplos de trechos destas observações feitas pelos participantes da avaliação:

Usuário 6: “Possibilidade de interromper quando já está familiarizado com as opções”.

Usuário 7: “Quando o robô inicia a fala das opções, devemos esperá-lo terminar para seguir, porém seria bom poder continuar sem esperar ele terminar”.

Usuário 8: “Acredito que poderia haver uma opção para parar ou voltar imediatamente ao invés de esperar toda a informação ser lida”.

Em razão da dificuldade de identificação das faces com óculos escuros, acessório comumente utilizado por pessoas com deficiência visual, foi indicada a necessidade de aprimoramento da detecção de faces. Outras sugestões foram a possibilidade de inclusão de um componente de fone de ouvido, a fim de manter a privacidade das informações, e um leitor de tela, como o Talk Back.

62,5% dos participantes consideraram o serviço de localização como a funcionalidade mais útil, especialmente devido à forma como a informação foi transmitida (x metros, à sua direita, esquerda, para frente, para trás). Contudo, um usuário discordou do emprego da distância em metros, já que, dada a dificuldade de abstração das pessoas, pessoas cegas costumam utilizar outras referências para se localizar, como portas, corredores, degraus e calçadas.

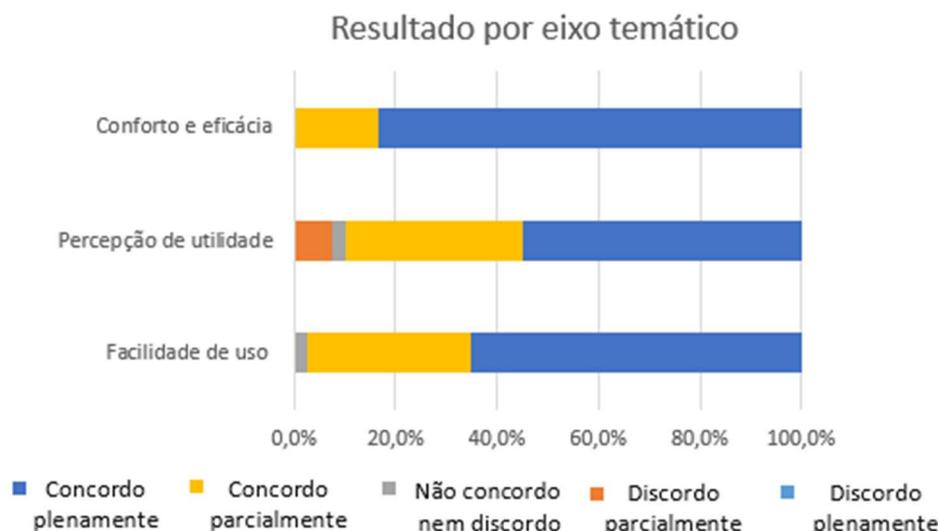
Seguem exemplos de trechos destas observações feitas pelos participantes da avaliação:

Usuário 3: As informações de humanos não são precisas. Dizem vai "reto", enquanto o conceito de reta compreende ponto A, onde se está, e ponto B, onde se quer ir e cujo caminho se desconhece. O robô emprega linguagem acessível, compatível e objetivo. Voz feminina é mais solidária, mais sensível de acordo com a experiência diária, em que as mulheres oferecem muito mais ajuda.

Usuário 4: Em locais grandes as pessoas se perdem. O robô usa como referência a direita e esquerda do usuário, enquanto isso não é claro na comunicação com as pessoas no dia a dia.

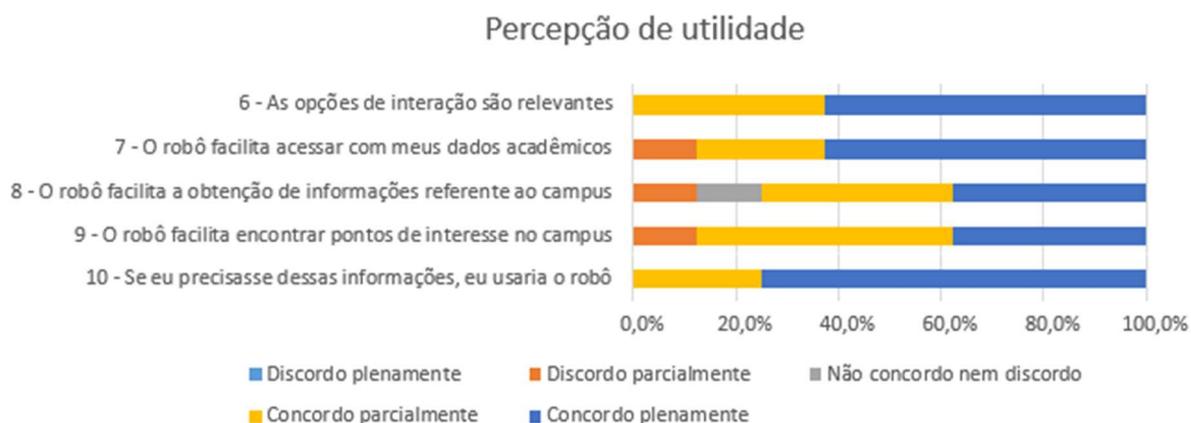
Usuário 7: Aumentar os detalhes. Exemplo: 5 metros à direita, na 2ª porta.

O gráfico a seguir (gráfico 2) demonstra a boa aceitação do protótipo pelos deficientes visuais.

Gráfico 2: Distribuição das respostas dos deficientes visuais por eixo temático

Fonte: Elaborado pelo autor.

Considerados os três eixos temáticos nos quais as perguntas estavam divididas, percebe-se que a percepção de utilidade do sistema foi avaliada menos positivamente. Abaixo, no gráfico 3, segue o resumo da distribuição das respostas nas questões 6 a 10.

Gráfico 3: Avaliação da percepção de utilidade pelos participantes com deficiência visual.

Fonte: Elaborado pelo autor.

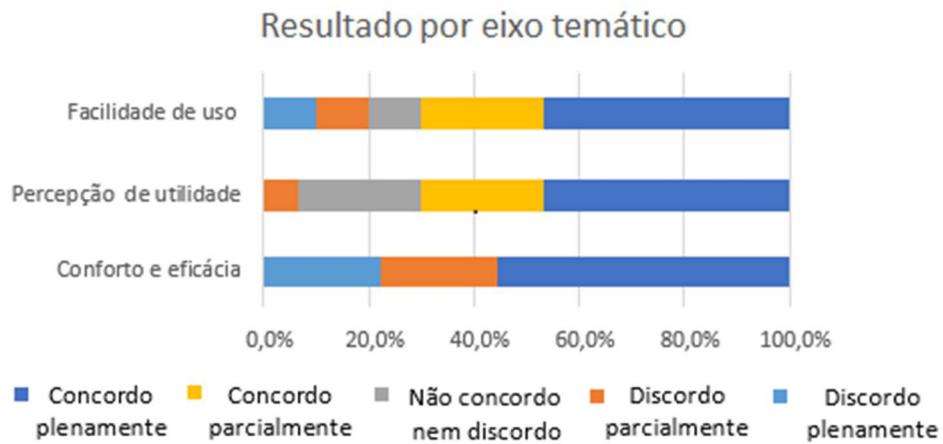
As notas de campo sugerem que o baixo tempo de teste do usuário 5 (inferior a 5 minutos) pode estar relacionado com suas respostas parcialmente discordantes nas questões 7, 8 e 9, visto que explorou apenas o serviço de biblioteca. Trata-se do único participante deste grupo que forneceu respostas discordantes.

6.4.2 Deficientes auditivos

As respostas deste grupo, menos homogêneas do que as respostas do primeiro, podem ser correlacionadas a fatores demográficos, o que será explorado na próxima subseção. No Gráfico 4, abaixo, percebe-se uma maior variação de respostas na escala Likert.

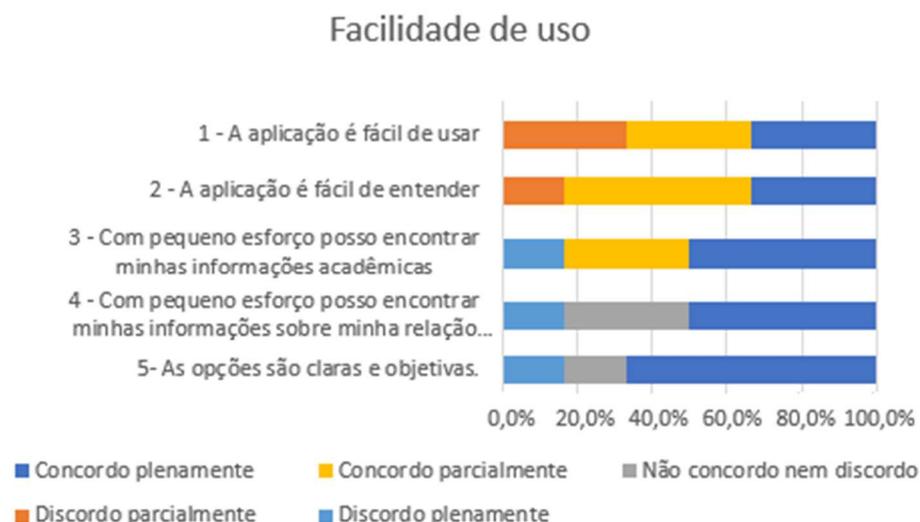
Gráfico 4: Distribuição das respostas dos deficientes auditivos por eixo temático.

Fonte: Elaborado pelo autor.



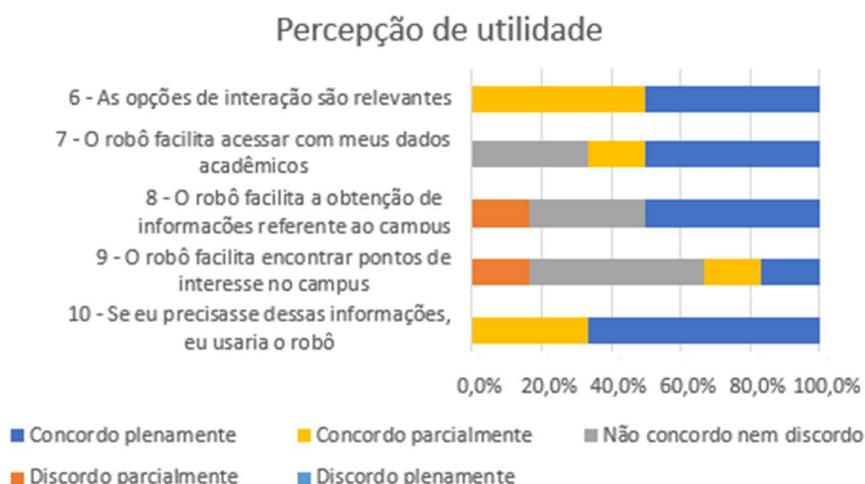
Fonte: Elaborado pelo autor.

O Gráfico 5 demonstra o padrão de respostas do grupo nas questões sobre facilidade de uso. Dois dos seis participantes discordaram parcialmente da afirmação 1 “A aplicação é fácil de usar”. Um deles não considerou a aplicação fácil de entender, discordando parcialmente da afirmativa 2. As afirmativas 3, 4 e 5 receberam uma discordância plena cada. Dois usuários responderam de forma neutra à pergunta 4. Notas de campo sugerem que a resposta de um deles pode estar relacionada ao fato de não terem conseguido realizar sem auxílio a tarefa proposta no caso de uso, que consistia na consulta aos livros em atraso e que a dificuldade em interpretar a afirmação possa ter gerado dúvidas quanto à resposta mais adequada, preferindo o participante abster-se.

Gráfico 5: Avaliação da facilidade de uso pelos deficientes auditivos.

Fonte: Elaborado pelo autor.

Quanto à percepção de utilidade, o gráfico 6, abaixo, sugere menor concordância com as afirmativas 8 e 9.

Gráfico 6: Avaliação da percepção de utilidade pelos deficientes auditivos.

Fonte: Elaborado pelo autor.

6.4.3 Deficientes auditivos que empregam LIBRAS

O Gráfico 7 traz a comparação das respostas oferecidas pelos deficientes que adotam LIBRAS e pelos demais participantes do estudo, inclusive deficientes auditivos cuja língua materna é o português. Foi possível constatar que as respostas dos 4 deficientes auditivos que empregam LIBRAS escapam ao padrão de resposta dos demais grupos e dos deficientes auditivos que não fazem uso da língua de sinais. Dois dos seis participantes com deficiência

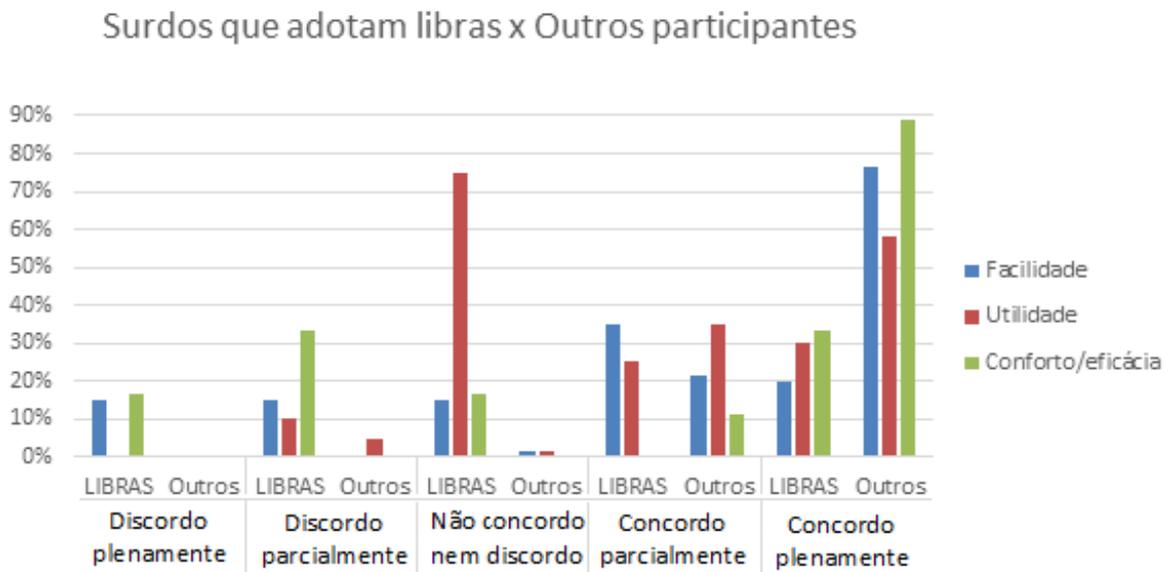
auditiva fazem uso do português na comunicação diária: uma senhora de 50 anos que sofreu perda auditiva progressiva em razão de problemas degenerativos e hoje possui deficiência severa; e um jovem de 26 anos com perda bilateral em razão de meningite na infância.

A partir da análise detida das questões com menor aceitação, verificou-se forte correlação entre esses resultados e as respostas de 2 usuários que não conseguiram desempenhar a tarefa solicitada. As anotações de campo revelam que esses usuários apresentaram dificuldades durante a interação e demonstraram sinais de desconforto e impaciência, o que redundou em uma experiência menos satisfatória com o robô.

Ambos são deficientes auditivos completos (severos, de acordo com a terminologia do IBGE) com escolaridade até Ensino Médio e sua língua materna é a linguagem brasileira de sinais. O relato da intérprete a respeito desses sujeitos e suas respostas subjetivas no questionário qualitativo demonstram que a falta de domínio pleno da língua portuguesa foi o principal obstáculo à interação.

Uma limitação do protótipo reside na ausência de previsão de uma interação adequada a usuários com dificuldades e limitações em língua portuguesa. Ambos participantes reconheceram limitações linguísticas e um deles recomendou o emprego da língua de sinais em substituição à língua portuguesa na aplicação destinada a pessoas com deficiência auditiva severa de nascença.

Gráfico 7: Comparação das respostas oferecidas pelos deficientes que adotam LIBRAS e pelos demais participantes do estudo



Fonte: Elaborado pelo autor.

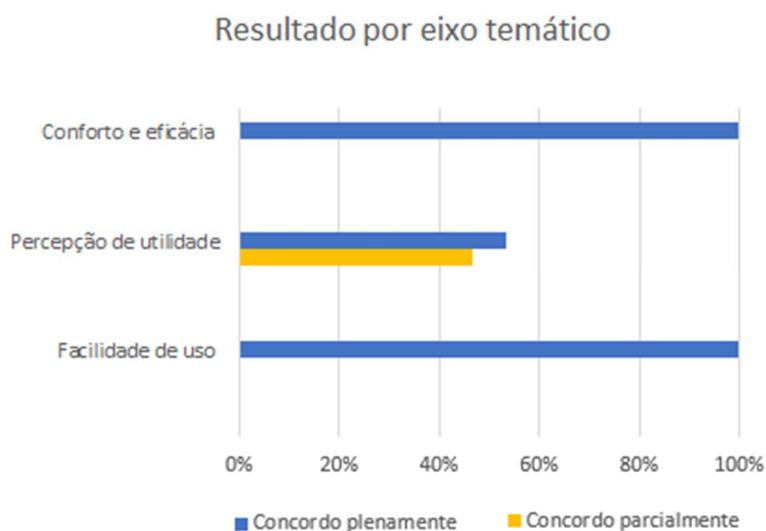
6.4.4 Deficientes parciais

Dentre os participantes dos três grupos, os deficientes parciais mistos foram os que avaliaram o protótipo de forma mais favorável, tendo a liberdade de explorar todas as formas de

interação de acordo com seu interesse e conveniência, adotando a modalidade que julgasse mais adequada ao contexto de uso.

No Gráfico 8, abaixo, o resultado do grupo em cada eixo temático.

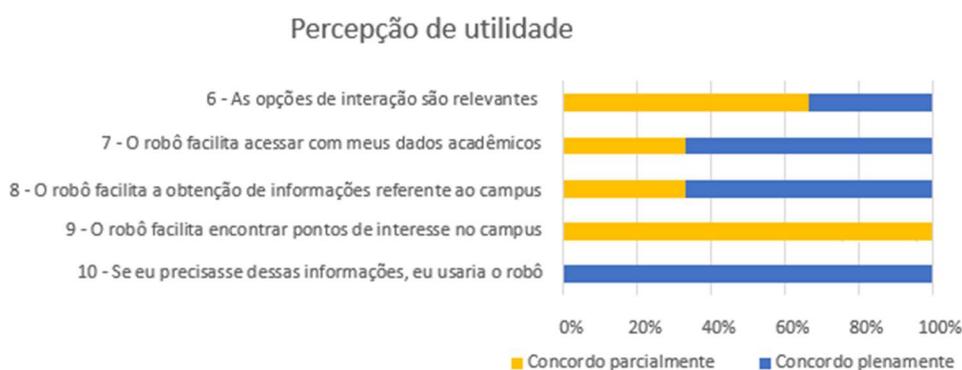
Gráfico 8: Distribuição das respostas dos deficientes parciais (visuais e auditivo) por eixo temático



Fonte: Elaborado pelo autor.

A percepção de utilidade foi a única categoria a receber respostas parcialmente concordantes. No Gráfico 9, a seguir, a distribuição das respostas nas questões 6 a 10.

Gráfico 9: Avaliação da percepção de utilidade pelos deficientes parciais.



Fonte: Elaborado pelo autor.

Neste caso, observa-se que os resultados foram positivos e apontam para uma aceitação do uso do protótipo.

A partir dos dados e relatos apresentados, observa-se um resultado positivo, de forma geral, quanto aos objetivos propostos de ampliar a percepção de facilidade de uso dos usuários com deficiência a partir da adaptação de canais de entrada e saída de dados adequados. Além disso, foram constatadas limitações e possibilidades de melhorias. Estas serão resumidas na seção seguinte onde serão apresentadas as conclusões e trabalhos futuros.

7 CONCLUSÃO

O estudo da interação multimodal é área do conhecimento em evidência, considerando a proliferação de trabalhos em que são abordadas as técnicas de reconhecimento de voz, face e gestos em diferentes aplicações, além da investigação de recursos mais adequados ao processamento das interações. Aplicações de interação multimodal proporcionam ao usuário uma experiência mais satisfatória e amigável, próxima à comunicação entre humanos. Este estudo buscou investigar recursos para possibilitar a realização de interação humano-computador de uma forma mais adequada ao usuário, considerando-se suas limitações e necessidades.

Realizou-se um estudo introdutório sobre aspectos gerais da população com deficiência na sociedade brasileira, inclusive no meio universitário. Daí emergiu a motivação para abordagem deste tema, buscando-se uma contribuição adequada às necessidades das pessoas com deficiência auditiva e visual.

A fim de nivelar conhecimentos prévios necessários à compreensão do modelo que se pretendia propor, foram abordados os conceitos de interação multimodal, adaptatividade e robótica assistiva, bem como o estado da arte e principais técnicas de entrada e saída de dados. Esses pressupostos (background) eram necessários à compreensão e análise de trabalhos relacionados que exploravam robótica assistiva e interação multimodal adaptativa.

A revisão de literatura e o conhecimento do atual estado da arte demonstraram as principais estratégias empregadas na adaptação da interação humano-computador, evidenciando o diferencial deste trabalho, qual seja, a adaptação dos canais (modalidades) de entrada e saída de acordo com o perfil do usuário. Com base nisso, o modelo de um sistema de interação multimodal adaptativo embarcado em um robô assistivo foi proposto e um protótipo foi implementado.

O diferencial do modelo proposto, em relação a outros trabalhos encontrados na literatura, é utilizar informações referentes à deficiência do usuário como critério para seleção automática das entradas e saídas de um sistema multimodal. Diferentemente dos trabalhos analisados, o modelo tem como proposta selecionar os canais de entrada e saída de acordo com a capacidade do usuário de utilizar aqueles canais em específico. Juntamente com a seleção adequada dos canais, o modelo não prevê restrições para a implementação de algum tipo específico de interação ou serviço, abrindo assim a possibilidade de exploração de um nicho cuja demanda é reprimida e para o qual múltiplas aplicações podem ser desenvolvidas.

O modelo propõe perfis de usuários que sejam usados na seleção de um conjunto específico de entradas e saídas coerente com as limitações e necessidades do usuário em particular. Assim, o usuário com surdez, por enquadrar-se nesse perfil já delimitado, poderia comunicar-se com o sistema por meio de gestos, interface GUI ou *touchscreen*. As saídas seriam adequadas às suas limitações, o que impede o sistema de apresentar resultados por meio de síntese de voz, por exemplo. Identificaram-se, portanto, aspectos que proporcionem uma melhor interação entre o sistema e o público usuário com deficiência, com vistas à sua integração no ambiente universitário, caso de ambientação explorado no protótipo implementado. Com a implementação do protótipo utilizando a robótica num contexto de serviços acadêmicos, foi possível avaliar a boa aceitação por parte dos usuários, sendo

descritas na próxima seção (7.1) as limitações identificadas. A fim de superá-las, são propostos trabalhos futuros na seção 7.2, a qual também abrange iniciativas de ampliação e melhoria do modelo.

7.1 Limitações

Apesar de a literatura em Ciências da Computação ser majoritariamente em língua inglesa, a língua portuguesa foi adotada para o protótipo a fim de torná-lo mais acessível ao público a que se destina – a população com deficiência auditiva e visual. Um sistema de reconhecimento e síntese em língua inglesa poderia comprometer os resultados de avaliação, enviesando o estudo, já que o público disponível para testes seria provavelmente mais escolarizado e com maior experiência prévia em recursos tecnológicos. Esse recorte amostral seria pouco representativo da realidade nacional, sendo mais fidedigna a amostra se empregada a língua portuguesa.

Essa escolha restringiu as possibilidades de escolha de uma API de reconhecimento de fala, impactando o método de interação selecionado. A API utilizada, Coruja (BATISTA, 2013) é totalmente voltada ao idioma português, entretanto ainda está em versão beta, o que demandou o emprego de um dicionário de palavras restrito para o usuário final. Essa limitação impede o processamento de linguagem natural, embora facilite as interações de modo padronizado por parte do usuário, uma vez que há um script bastante claro, aumentando a satisfação quanto à facilidade de uso do sistema.

A implementação não considerou que o público com deficiência visual costuma ser ágil na seleção do objeto de seu interesse, visto que familiarizado com recursos de acessibilidade e leitores de tela, como Dolphin, JAWS and Window-Eye (YE; LI; LI, 2014). Talk Back (fornecido pela Google e pré-instalado em celulares com sistema operacional Android) e o VoiceOver (disponível em dispositivos da Apple e em celulares com sistema iOS) são bastante populares entre esse público, porque são recursos de acessibilidade presentes nas configurações de fábrica dos aparelhos – não há necessidade de instalar um aplicativo adicional.

Como empregam essas tecnologias em seu dia a dia, possuem ouvidos aguçados e atentos a opções de voz, elevando sua capacidade de julgamento no momento da interação. Os participantes com deficiência visual iniciavam seu turno de interação antes mesmo de o robô concluir a leitura dos menus em seu turno, e sua fala não era identificada pelo sistema, pois a saída de áudio implica a desativação da entrada de áudio. Foi observada impaciência nesses usuários e amplamente recomendada, como mudança no sistema a fim de facilitar o uso, a possibilidade de interromper o robô, dando o comando desejado no momento em que o usuário se sentisse esclarecido para escolher a opção de menu desejada e não somente após o fim do turno do robô. Também foi sugerido que os recursos “Preciso de Ajuda”, “Voltar” e “Resetar” estivessem disponíveis a qualquer tempo, não apenas durante o turno do usuário.

A interação dos participantes com deficiência auditiva que empregam LIBRAS foi prejudicada em razão das palavras em língua portuguesa escolhidas para os ícones (botões)

dos menus propostos. Pessoas com deficiência auditiva costumam apresentar restrições linguísticas em português (GUARINELLO et al., 2009), então a adoção da interface gráfica deveria contemplar essas limitações por meio de uma seleção de palavras em língua portuguesa que sejam simples, claras e objetivas.

Ainda quanto a esse público, restou clara a preferência pela língua de sinais como alternativa ao português. O modelo atenderia de forma mais satisfatória os usuários que se comunicam preferencialmente por meio LIBRAS se reconhecesse esses gestos como entrada. Embora o modelo preveja o reconhecimento e processamento de gestos entre as modalidades de entrada, essa funcionalidade não foi implementada no protótipo. Na saída, os gestos se restringiram à comunicação de dados numéricos (como número de livros em empréstimo ou atraso), expressos pela mão robótica do protótipo em seu turno de interação. A avaliação demonstrou que essa utilização restrita afeta a usabilidade, que seria melhorada em termos de facilidade de uso e satisfação (conforto e eficácia) se todas as informações comunicadas ao usuário por meio de LIBRAS.

7.2 Trabalhos Futuros

Ante as limitações observadas na seção anterior e como forma de colaborar com a continuidade do trabalho nesta linha de investigação, listamos a seguir possíveis formas de superar as necessidades detectadas e melhorar o modelo e o protótipo.

Principais sugestões de trabalhos futuros:

- 1) Empregar ou desenvolver método de reconhecimento de fala em Português-Brasileiro que permita interação em linguagem natural.
- 2) Implementar método de reconhecimento de gestos em LIBRAS, à semelhança do que consta na arquitetura do robô assistivo tutor da língua de sinais turca (DEBEVC; KOSEC; HOLZINGER, 2011)
- 3) Incluir no modelo a previsão de um agente animado, implementando a interface gráfica com um agente animado que aja à semelhança do intérprete virtual Hugo, do software comercial Hand Talk (PRESS_RELEASE, 2012), executando gestos em LIBRAS (modalidade de saída).
- 4) Adaptar o modelo a fim de permitir o processamento de entrada de áudio durante o turno de saída de áudio do robô.
- 5) Incluir no modelo um módulo de identificação da deficiência para cadastramento automático do usuário no tipo de perfil mais adequado às suas aptidões, habilidades e limitações, identificadas em turnos breves de interação.
- 6) Adaptar a seleção de serviços disponíveis ao usuário por meio da sensibilidade ao contexto (DJAID; SAADIA; RAMDANE-CHERIF, 2012) e gerenciamento de trilhas (TAVARES, 2011).

REFERÊNCIAS

- AIGNER, R. et al. **Understanding Mid-Air Hand Gestures : A Study of Human Preferences in Usage of Gesture Types for HCI. Microsoft Research Technical Report MSR-TR-2012-111.** Redmond, USA: [s.n.]. Disponível em: <<http://research.microsoft.com/apps/pubs/?id=175454>>.
- AKIKI, P. A.; BANDARA, A. K.; YU, Y. Adaptive Model-Driven User Interface Development Systems. **ACM Computing Surveys**, v. 47, n. 1, p. 1–33, 1 maio 2014.
- BAHADORI, S. et al. **RoboCare: an Integrated Robotic System for the Domestic Care of the Elderly.** Ambient Intelligence. **Anais...**Pisa, Italia: 2003
- BATISTA, P. Avanços em Reconhecimento de Fala para Português Brasileiro e Aplicações: Ditado no LibreOffice e Unidade de Resposta Audível com Asterisk. p. 95, 2013.
- BENYON, D.; MURRAY, D. **Developing adaptive systems to fit individual aptitudes** Proceedings of the 1st international conference on Intelligent user interfaces - IUI '93. **Anais...**New York, New York, USA: ACM Press, 1993 Disponível em: <<http://portal.acm.org/citation.cfm?doid=169891.169925>>
- BEZOLD, M.; MINKER, W. **Adaptive Multimodal Interactive Systems.** [s.l: s.n.].
- BOLT, R. A. **Put that there.** Proceedings of the 7th annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '80. **Anais...**New York, New York, USA: ACM Press, 1980
- BRASIL. Decreto nº. 5.296, de 2 de dezembro de. . 2004.
- BRASIL. Decreto nº. 6.949, de 25 de agosto de. . 2009.
- BROSE, S. W. et al. The Role of Assistive Robotics in the Lives of Persons with Disability. **American Journal of Physical Medicine & Rehabilitation**, v. 89, n. 6, p. 509–521, jun. 2010.
- CELESTE, W. C. et al. Robótica assistiva. In: **Robótica móvel. Roseli Aparecida Francelin Romero et al (Org.).** [s.l: s.n.].
- CHHABRIA, S. A.; DHARASKAR, R. V.; THAKARE, V. M. **Intelligent Gesture Recognition to Design more Efficient & Intelligent Multimodal System** 6th International Conference on Emerging Trends in Engineering and Technology. **Anais...**IEEE, 2013 Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6754817>>
- CHHABRIA, S. A.; DHARASKAR, R. V.; THAKRE, V. M. Multimodal Interpretation Gesture Recognition System: A Review. **International Journal of Computer Applications**, v. 2012, p. 35–39, 2012.
- CHOUDHURY, T. et al. **Multimodal Person Recognition using Unconstrained Audio and Video** International Conference on Audio and Video-Based Person Authentication Proceedings. **Anais...**1999
- CONNETTE, C. P. et al. **The mobility concept of Care-O-bot 3** 39th International Symposium on Robotics. **Anais...**Seoul, Korea: 2008
- CORSO, R. A. **Are Consumers More Responsive to Male or Female Voices in Advertisements?** Reston, VA, USA: [s.n.]. Disponível em: <<http://media.theharrispoll.com/documents/Harris-Interactive-Poll-Research-Media-Advertising-2010-03.pdf>>.

COSTA, H. J. DE M. **Um Modelo de Arquitetura Para o Turismo Ubíquo Utilizando Dispositivos Móveis.** [s.l: s.n.].

COUTAZ, J. **User interface plasticity** Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems - EICS '10. **Anais...**New York, New York, USA: ACM Press, 2010Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-77955106658&partnerID=tZOtx3y1>>

DEBEVC, M.; KOSEC, P.; HOLZINGER, A. Improving multimodal web accessibility for deaf people: Sign language interpreter module. **Multimedia Tools and Applications**, v. 54, n. 1, p. 181–199, 2011.

DEY, A.; ABOWD, G.; SALBER, D. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. **Human-Computer Interaction**, v. 16, n. 2, p. 97–166, 1 dez. 2001.

DIAS, C. **Estudo de caso : idéias importantes e referências.** [s.l: s.n.]. Disponível em: <http://www.reocities.com/claudiaad/case_study.pdf>.

DINIZ, E. H. et al. **Abordagens Epistemológicas em Pesquisas Qualitativas: Além do Positivismo nas Pesquisas na Área de Sistemas de Informação**XXX Encontro Anual da ANPAD. **Anais...**Salvador: 2006

DJAID, N. T.; SAADIA, N.; RAMDANE-CHERIF, A. **Multimodal architecture to strengthen the interaction of the robot in ambient intelligence environments**Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12. **Anais...**New York, USA: ACM, 2012

DUMAS, B.; SOLÓRZANO, M.; SIGNER, B. **Design guidelines for adaptive multimodal mobile input solutions**Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13. **Anais...**New York, New York, USA: ACM Press, 2013Disponível em: <<http://dl.acm.org/citation.cfm?doid=2493190.2493227>>

EAGLY, A. H. .; MLADINIC, A.; OTTO, S. Cognitive and Affective Bases of Attitudes toward Social Groups and Social Policies. **Journal of Experimental Social Psychology**, v. 30, p. 113–137, 1994.

EYSSEL, F. et al. **“If you sound like me, you must be more human”**Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12. **Anais...**New York, New York, USA: ACM, 2012Disponível em: <<http://dl.acm.org/citation.cfm?doid=2157689.2157717>>

FEIL-SEIFER, D.; MATARIC, M. J. **Socially Assistive Robotics.**9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005. **Anais...**IEEE, 2005Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1501143>>. Acesso em: 29 set. 2014

FISCHINGER, D.; EINRAMHOF, P. **Hobbit-The Mutual Care Robot**Assistance and Service Robotics in a Human Environment. **Anais...**Japan: 2013Disponível em: <http://users.ics.forth.gr/~argyros/mypapers/2013_11_asrob_hobbit.pdf>. Acesso em: 1 out. 2014

FISKE, S. T. et al. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. **Journal of personality and social psychology**, v. 82, n. 6, p. 878–902, 2002.

FMC. **Fundamental Modeling Concepts - TAM: the SAP way combining FMC and UML.**

Disponível em: <<http://www.fmc-modeling.org/fmc-and-tam>>. Acesso em: 12 dez. 2015.

GNJATOVIĆ, M. et al. **Adaptive multimodal interaction with industrial robot** 10th Jubilee International Symposium on Intelligent Systems and Informatics. **Anais...** Subotica, Serbia: IEEE, 2012

GOETZE, S.; FISCHER, S.; MORITZ, N. **Multimodal human-machine interaction for service robots in home-care environments** Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments. **Anais...** Stroudsburg, PA, USA: Association for Computational Linguistics, 2012 Disponível em: <<http://dl.acm.org/citation.cfm?id=2392867>>. Acesso em: 1 out. 2014

GRAF, B. et al. **Robotic home assistant Care-O-bot® 3 - product vision and innovation platform** IEEE Workshop on Advanced Robotics and its Social Impacts. **Anais...** IEEE, nov. 2009 Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5587059>>. Acesso em: 1 out. 2014

GRAF, B.; HANS, M.; SCHRAFT, R. D. Care-O-bot II—Development of a Next Generation Robotic Home Assistant. **Autonomous Robots**, v. 16, n. 2, p. 193–205, mar. 2004.

GROOTHUIS, S. S.; STRAMIGIOLI, S.; CARLONI, R. Lending a helping hand: Toward novel assistive robotic arms. **IEEE Robotics and Automation Magazine**, v. 20, n. March, p. 20–29, 2013.

GUARINELLO, A. C. et al. Surdez e letramentos: pesquisa com Surdos universitários de Curitiba e Florianópolis. **Revista Brasileira de Educação Especial**, v. 15, n. 1, p. 99–120, 2009.

GUPTA, R. Human Computer Interaction – A Modern Overview. **International Journal for Computer Technology & Applications**, v. 3, n. 5, p. 1736–1740, 2012.

HANS, M.; GRAF, B.; SCHRAFT, R. **Robotic Home Assistant Care-O-bot: Past-present-future** International Workshop on Robot and Human Interactive Communication. **Anais...** Berlin: IEEE, 2002 Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1045652>. Acesso em: 1 out. 2014

HANSON, D. **Exploring the Aesthetic Range for Humanoid Robots** Cognitive Science Workshop on Android Science , Proceedings of the. **Anais...** 2006 Disponível em: <<http://www.androidscience.com/proceedings2006/6Hanson2006ExploringTheAesthetic.pdf>>

HÖLN, R.; NEMITZ, F. **Robots as multifunctional gentlemen**. Stuttgart: [s.n.]. Disponível em: <http://www.care-o-bot.de/content/dam/careobot/en/documents/pressreleases/2015_01_15_Care-O-bot_4_en_final.pdf>.

HOPPEN, N.; LAPOINTE, L.; MOREAU, E. Um guia para avaliação de artigos de pesquisas em sistemas de informação. **REAd**, v. 2, n. 2, p. 1–34, 1996.

HOPPEN, N.; MEIRELLES, F. S. Sistemas de informação: um panorama da pesquisa científica entre 1990 e 2003. **Revista de Administração de Empresas**, v. 45, n. 1, p. 24–35, mar. 2005.

IBGE. **Censo Demográfico 2010: Características gerais da população, religião e pessoas com deficiência**. Rio de Janeiro: [s.n.].

IJJJ. **Censo Demográfico 2010: Características gerais da população , religião e pessoas com deficiência**. Rio de Janeiro: IBGE, 2012.

- INEP. **Matrículas no ensino superior crescem 3,8%**. Disponível em: <http://portal.inep.gov.br/visualizar/-/asset_publisher/6AhJ/content/matriculas-no-ensino-superior-crescem-3-8>.
- JUANG, B. H.; RABINER, L. R. Automatic Speech Recognition – A Brief History of the Technology Development. **Elsevier Encyclopedia of Language and Linguistics**, p. 1–24, 2005.
- K. ZSIGA, M. RAUHALA, P. RUMEAU, A. TÓTH, T. PILISSY, G. F. **User requirements for home-care robots** 6th ISPO Central European. **Anais...**2011
- KARAM, M.; SCHRAEFEL, M. C. **A Taxonomy of Gestures in Human Computer Interactions** **Technical Report, Electronics and Computer Science**. Southampton, UK: [s.n.]. Disponível em: <<http://eprints.soton.ac.uk/261149/>>.
- KARRAY, F. et al. Human-Computer Interaction: Overview on State of the Art. **International Journal on Smart Sensing and Intelligent Systems**, v. 1, n. 1, p. 137–159, 2008.
- KÖCHE, J. C. **Fundamentos de metodologia científica: teoria da ciência e iniciação à pesquisa**. Petrópolis: Vozes, 2011.
- LANGVIN, G. **InMoov Project**. Disponível em: <<http://www.inmoov.fr/project/>>. Acesso em: 5 abr. 2015.
- LARSON, J. A. et al. **W3C Multimodal Interaction Framework** (J. A. Larson, T. V. Raman, D. Raggett, Eds.) **Components**. [s.l.: s.n.]. Disponível em: <<http://www.w3.org/TR/2003/NOTE?mimi?framework?20030506/>>.
- LEBEC, O. et al. **High level functions for the intuitive use of an assistive robot**. International Conference on Rehabilitation Robotics. **Anais...**IEEE, 2013
- LENZI, M. B. **Os dados sobre Deficiência nos Censos Demográficos Brasileiros** XVIII Encontro Nacional de Estudos Populacionais - transformações na população brasileira: complexidades, incertezas e perspectivas. **Anais...**Águas de Lindóia: ABEP, 2012 Disponível em: <[http://www.abep.nepo.unicamp.br/xviii/anais/files/ST35\[845\]ABEP2012.pdf](http://www.abep.nepo.unicamp.br/xviii/anais/files/ST35[845]ABEP2012.pdf)>
- LI, J. et al. An overview of noise-robust automatic speech recognition. **IEEE Transactions on Audio, Speech and Language Processing**, v. 22, n. 4, p. 745–777, 2014.
- LIN, S. H.; KUNG, S. Y.; LIN, L. J. Face recognition/detection by probabilistic decision-based neural network. **IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council**, v. 8, n. 1, p. 114–132, 1997.
- MAAT, L.; PANTIC, M. Gaze-X: Adaptive, Affective, Multimodal Interface for Single-User Office Scenarios. **Artificial Intelligence for Human Computing**, p. 251–271, 2006.
- MAES, S. H.; SARASVAT, V. (EDS.). **Multimodal Interaction Requirements** **W3C**. [s.l.: s.n.]. Disponível em: <<http://www.w3.org/TR/2003/NOTE-mimi-reqs-20030108/>>. Acesso em: 18 nov. 2015.
- MALINOWSKI, U. et al. A Taxonomy of Adaptive User Interfaces. In: MONK, A. F.; DIAPER, D.; HARRISON, M. D. (Eds.). **People and Computers VII: Proceedings of the British Computer Society HCI '92 Conference, September 1992**. Cambridge: Cambridge University Press, 1992. p. 391–414.
- MANZINI, E. J. **Entrevista semi-estruturada: análise de objetivos e de roteiros** Seminário internacional sobre pesquisa e estudos qualitativos. **Anais...**Bauru: Universidade de São Carlos, 2004 Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Entrevista+semi->

estruturada:+an?lise+de+objetivos+e+de+roteiros#0>

MARAGOS, P.; POTAMIANOS, A.; GROS, P. (EDS.). **Multimodal Processing and Interaction: Audio, Video, Text**. New York, USA: Springer, 2008.

MAYER, P.; BECK, C.; PANEK, P. **Examples of multimodal user interfaces for socially assistive robots in Ambient Assisted Living environments**3rd International Conference on Cognitive Infocommunications (CogInfoCom). **Anais...IEEE**, dez. 2012Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6422014>>

MCKENNA, S.; GONG, S. **Non-intrusive person authentication for access control by visual tracking and face recognition**1st International Conference on Audio- and Video-Based Biometric Person Authentication. **Anais...London, UK: Springer-Verlag**, 1997

MEYER, B.; YAKEMOVIC, K. C. B.; HARRIS, M. **Issues in practical application of an adaptive interface**Proceedings of the 1st international conference on Intelligent user interfaces - IUI '93. **Anais...New York, New York, USA: ACM Press**, 1993Disponível em: <<http://dl.acm.org/citation.cfm?id=170018>>

MITCHELL, W. J. et al. Does social desirability bias favor humans? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management. **Computers in Human Behavior**, v. 27, n. 1, p. 402–412, 2011.

MOGHADDAM, B.; PENTLAND, A. Probabilistic visual learning for object representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 696–710, 1997.

MÜLLER, S.; SPRENGER, S.; GROSS, H.-M. OfficeMate: A Study of an Online Learning Dialog System for Mobile Assistive Robots. **ADAPTIVE 2014, The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications**, n. Adaptive, p. 104–110, 2014.

MYERS, M. D.; KLEIN, H. K. A set of principles for conducting and evaluating interpretative field studies in Information Systems. **MIS Quarterly**, v. 23, n. 1, p. 67–93, 1999.

NASS, C.; MOON, Y.; GREEN, N. Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. **Journal of Applied Social Psychology**, v. 27, n. 10, p. 864–876, 1997.

OLIVEIRA, R. et al. **Recursos para desenvolvimento de aplicativos com suporte a reconhecimento de voz para desktop e sistemas embarcados**12th International Conference on Computational Processing of the Portuguese Language (PROPOR). **Anais...Coimbra, Portugal: 2011**Disponível em: <<https://www.laps.ufpa.br/marcusdias/sitelaps/images/publicacao/2011/paper2.pdf>>

OVIATT, S. Ten myths of multimodal interaction. **Communications of the ACM**, v. 42, n. 11, p. 74–81, 1 nov. 1999.

OVIATT, S. et al. Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions. **Human-Computer Interaction**, v. 15, n. 4, p. 263–322, 2000.

OVIATT, S.; COHEN, P. Perceptual user interfaces: multimodal interfaces that process what comes naturally. **Communications of the ACM**, v. 43, n. 3, p. 45–53, 2000.

PAPOUTSAKIS, K. et al. **Developing visual competencies for socially assistive robots : the HOBBIT approach**Workshop on Robotics Assistive Environments (RasEnv 2013), in conjunction with PETRA. **Anais...2013a**

- PAPOUTSAKIS, K. et al. **Developing visual competencies for socially assistive robots : the HOBBIT approach** Workshop on Robotics Assistive Environments (RasEnv 2013), in conjunction with PETRA. **Anais...**Rhodes, Greece: 2013b
- POLLACK, M. E. et al. **Pearl : A Mobile Robotic Assistant for the Elderly**. Workshop on Automation as Caregiver: the Role of Intelligent Technology in Elder Care (AAAI). **Anais...**2002 Disponível em: <<http://www.cs.cmu.edu/~nurse-bot/web/papers/umich/aaai02wkshp.pdf>>
- PRATES, J. S. **Gerenciamento de diálogo baseado em modelo cognitivo para sistema de interação multimodal voltado a tratamento de perguntas e respostas**. [s.l.] Universidade do Vale do Rio dos Sinos, 2015.
- PRESS_RELEASE. **Hand Talk : Inovação e acessibilidade digital em Libras para milhões de surdos** . Maceió: [s.n.]. Disponível em: <<http://www.handtalk.me/files/media-kit-handtalk.zip>>.
- RANASINGHE, R. et al. **Smart hoist: An assistive robot to aid carers** 13th International Conference on Control ,Automation, Robotics & Vision. **Anais...**Marina Bay Sands, Singapoure: IEEE, 2014 Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7064501>>
- REHRL, T. et al. **The Ambient Adaptable Living Assistant is Meeting its Users** Proceedings of the AAL Forum 2012. **Anais...**Eindhoven, Netherlands: Eurecom, 2012 Disponível em: <http://www.aal.fraunhofer.de/publications/2012_Rehrl_RTb_12_The_Ambient_Adaptable_Living_Assistant_is_Meeting_its_Users_AAL-FORUM12.pdf>
- REISER, U. et al. **Care-O-bot® 3 - Creating a product vision for service robot applications by integrating design and technology** IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS. **Anais...**St. Louis, USA: IEEE, 2009
- REITHINGER, N. et al. **SmartKom** Proceedings of the 5th international conference on Multimodal interfaces - ICMI '03. **Anais...**New York, New York, USA: ACM Press, 2003 Disponível em: <<http://portal.acm.org/citation.cfm?doid=958432.958454>>
- RUDMAN, L. A.; GOODWIN, S. A. Gender Differences in Automatic In-Group Bias: Why Do Women Like Women More Than Men Like Men? **Journal of Personality and Social Psychology**, v. 87, n. 4, p. 494–509, 2004.
- SeRoDi**. Disponível em: <www.serodi.de>. Acesso em: 6 jan. 2015.
- SHIMADA, M. et al. **Uncanny Valley of Androids and Its Lateral Inhibition Hypothesis** RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. **Anais...**IEEE, 2007 Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4415112>. Acesso em: 1 out. 2014
- SIMONI, C. A. C.; BARANAUSKAS, M. C. C. **Pesquisa qualitativa em sistemas de informação** MIS Quarterly. [s.l.: s.n.]. Disponível em: <<http://www.ic.unicamp.br/~reltech/2003/03-02.pdf>>.
- SIROVICH, L.; KIRBY, M. Low-dimensional procedure for the characterization of human faces. **Journal of the Optical Society of America. A, Optics and image science**, v. 4, n. 3, p. 519–524, 1987.
- STEFFEN, M. **Adaptative input interpretation for dialogue management of an autonomous robot** 5th Companion Able Workshop. **Anais...**Brussels, Belgium: 2011

- STIEFELHAGEN, R. et al. Enabling Multimodal Human – Robot Interaction for the Karlsruhe Humanoid Robot. **IEEE Transactions on Robotics**, v. 23, n. 5, p. 840–851, 2007.
- TAVARES, J. E. DA R. **Hefestos: um Modelo para Suporte à Acessibilidade Ubíqua**. [s.l.] Unisinos, 2011.
- THIRAN, J.-P.; BOURLARD, H.; MARQUES, F. (EDS.). **Multimodal signal processing: theory and applications for human-computer interaction**. 1. ed. Oxford, UK: Academic Press, 2010.
- TURK, M. A.; PENTLAND, A. P. **Face recognition using eigenfaces**Computer Society Conference on Computer Vision and Pattern Recognition. **Anais...**Maui, HI: IEEE, 1991Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=139758>>
- ULUER, P.; AKALIN, N.; KÖSE, H. A New Robotic Platform for Sign Language Tutoring. **International Journal of Social Robotics**, v. 7, n. 5, p. 571–585, 23 nov. 2015.
- VARONA, J.; MANRESA-YEE, C.; PERALES, F. J. Hands-free vision-based interface for computer accessibility. **Journal of Network and Computer Applications**, v. 31, n. 4, p. 357–374, 2008.
- VIOLA, P.; JONES, M. **Rapid object detection using a boosted cascade of simple features**IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. **Anais...**IEEE Comput. Soc, 2001Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=990517>>
- WACHS, J. P. et al. Vision-based hand-gesture applications. **Communications of the ACM**, v. 54, n. 2, p. 60, 2011.
- WALLACE, R. **The elements of AIML style**. [s.l.] ALICE A. I. Foundation, Inc., 2003.
- WALSHAM, G. **The emergence of interpretivism in IS research**Information Systems Research, 1995.
- WASINGER, R. **Multimodal Interaction with Mobile Devices: Fusing a Broad Spectrum of Modality Combinations**. [s.l.] Saarbrücken: Universität des Saarlandes, 2006.
- WELD, D. S. et al. **Automatically personalizing user interfaces**IJCAI International Joint Conference on Artificial Intelligence. **Anais...**2003
- WIGGERS, I. P.; ROTHKRANTZ, L. J. M. Automatic speech recognition using hidden Markov models. In: **Course IN4012TU - Real-time AI & Automatische Spraakherkenning**. Delft, Netherlands: [s.n.]. p. 70.
- WILSON, C. E. Triangulation. **interactions**, v. 13, n. 6, p. 46–47; 63, 1 nov. 2006.
- WINCKLER, M.; PIMENTA, M. Avaliação de Usabilidade de sites Web. In: **Escola de Informática da SBC-Sul (ERI2002)**. [s.l.: s.n.]. p. 85–137.
- WISKOTT, L. et al. Face recognition by elastic bunch graph matching. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 775–779, 1997.
- YE, M.; LI, P.; LI, Q. **VIPReader: A Light News Reader for the Visually Impaired Person**2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). **Anais...**IEEE, ago. 2014Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6927636>>
- YIN, R. K. **Estudo de caso: planejamento e métodos**. 2nd. ed ed. Porto Alegre: Bookman, 2001. v. XXXIII
- ZHAO, W. et al. Face recognition: A literature survey. **ACM Computing Surveys**, v. 35, n. 4,

p. 399–458, 2003.

ZOLLO, L.; WADA, K.; VAN DER LOOS, H. F. Special Issue on Assistive Robotics. **IEEE Robotics and Automation Magazine**, p. 16–19, 2013.

ZSIGA, K. et al. **User requirements of socially active robots** Poster at European Federation for research in Rehabilitation. **Anais...**Riva Del Garda: 2011

APÊNDICE A

1) Aprovação do projeto nº 15/183



UNIVERSIDADE DO VALE DO RIO DOS SINOS
Unidade de Pesquisa e Pós-Graduação (UAAPPG)
Comitê de Ética em Pesquisa (CEP)

Versão agosto/2013

UNIDADE DE PESQUISA E PÓS-GRADUAÇÃO
COMITÊ DE ÉTICA EM PESQUISA
RESOLUÇÃO 165/2015

O Comitê de Ética em Pesquisa da Universidade do Vale do Rio dos Sinos – UNISINOS analisou o projeto:

Projeto: Nº CEP 15/183 **Versão do Projeto:** 28/10/2015 **Versão do TCLE:** 28/10/2015

Coordenador:
Mestrando Edward Simon John (PPG em Computação Aplicada)

Título: Interação multimodal adaptativa embarcada em robótica assistiva para comunicação com pessoas com deficiência.

Parecer: O projeto foi APROVADO, por estar adequado ética e metodologicamente, conforme os preceitos da Resolução 466/12 do Conselho Nacional de Saúde.

O pesquisador deverá encaminhar relatório anual sobre o andamento do projeto, conforme o previsto na Resolução CNS 466/12, item XI.2, letra d. Somente poderão ser utilizados os Termos de Consentimento onde conste a aprovação do CEP/UNISINOS.

São Leopoldo, 28 de outubro de 2015.


Prof. Dr. José Roque Junges
Coordenador do CEP/UNISINOS

2) Termo de Consentimento Livre e Esclarecido



UNIVERSIDADE DO VALE DO RIO DOS SINOS
Unidade Acadêmica de Pesquisa e Pós-Graduação
Comitê de Ética em Pesquisa

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – TCLE

Você está sendo convidado(a) a participar como **voluntário** em uma pesquisa desenvolvida por Edward Simon John, mestrando do Programa de Computação Aplicada, ênfase Inteligência Artificial, da Universidade do Vale do Rio dos Sinos – Unisinos. A dissertação “Interação Multimodal Adaptativa embarcada em robótica assistiva para comunicação com pessoas com deficiência”, orientada pelo Prof. Dr. Sandro José Rigo, objetiva o desenvolvimento de tecnologia para inclusão de pessoas com deficiência no ambiente universitário, facilitando sua comunicação e permitindo seu acesso a informações de forma autônoma, sem auxílio ou dependência de outra pessoa. Uma das justificativas dessa pesquisa é o fato de a população universitária com deficiência ser bastante inferior à população brasileira com deficiência, o que pode ser um sinal de que a Universidade precisa ser mais acessível e inclusiva.

Participando como voluntário nessa pesquisa, você interagirá com um robô, que lhe apresentará opções de serviços relacionados à sua vida acadêmica, como consulta de disciplinas, graus e faltas. Você se comunicará com ele por meio da fala ou do toque em um monitor (display), sem que qualquer desconforto seja gerado para você. Após a interação com o robô, você responderá questões sobre facilidade de uso do sistema, desempenho do sistema, nível de acessibilidade, relevância dos recursos sugeridos, adequação da solução às necessidades reais do usuário e avaliação geral. As respostas serão dadas de acordo com a escala de Likert, podendo ser uma das cinco seguintes: discorda totalmente, discorda, nem concorda/nem discorda, concorda, concorda totalmente. Você terá um campo livre no qual pode incluir observações, expressar críticas, sugestões e/ou opiniões.

Essa pesquisa envolve risco e desconforto mínimos para você, uma vez que são empregados canais adequados à sua deficiência, seja ela visual, seja auditiva. Você se beneficia de sua participação graças ao acesso, de forma autônoma, a informações relevantes ao contexto acadêmico, como suas faltas por disciplina e graus alcançados em cada matéria. Essa pesquisa possibilita geração de conhecimento sem afetar seu bem-estar, podendo auxiliar futuramente outras pessoas com deficiência graças ao aperfeiçoamento tecnológico obtido e ao aprimoramento de técnicas de interação multimodal humano com deficiência-computador.

Você será esclarecido(a) sobre a pesquisa em qualquer aspecto que desejar, sendo livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não acarretará qualquer penalidade ou perda de benefícios.

O pesquisador irá tratar a sua identidade com padrões profissionais de sigilo e os resultados da pesquisa serão enviados para você, permanecendo confidenciais. Seu nome ou o material que indique a sua participação não será liberado sem a sua permissão. Você não será identificado(a) em nenhuma publicação que possa resultar deste estudo.

Você desde já está ciente da possibilidade de ser fotografado ou filmado, consentindo que o pesquisador utilize esses materiais, desde que preservada sua identidade e aplicada tarja sobre seu rosto. Você será consultado previamente sobre sua disponibilidade para realização dos testes, a serem agendados. Você autoriza o acesso a suas informações acadêmicas na Universidade do Vale do Rio dos Sinos – Unisinos, tais como disciplinas

Av. Unisinos, 950 Caixa Postal 275 CEP 93022-000 São Leopoldo Rio Grande do Sul, Brasil
Fone: (51) 3591-1198 ou ramal 2198 Fax: (51) 3590-8118 <http://www.unisinos.br>

CEP - UNISINOS
VERSÃO APROVADA
Em: 29/10/15
.....

cursadas, graus e faltas. A participação no estudo não acarretará custos para você e não será ofertada nenhuma compensação financeira.

Declaro de forma autônoma, consciente, livre e esclarecida que concordo em participar desse estudo. Recebi uma cópia assinada deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas. Se deficiente visual, oferecido o esclarecimento por meio de áudio, minha manifestação de vontade livre e esclarecida será gravada em vídeo na presença de testemunha que firmará este TCLE.

Nome	Assinatura do Participante	Data
------	----------------------------	------

Nome	Assinatura do Pesquisador	Data
------	---------------------------	------

Nome	Assinatura da Testemunha	Data
------	--------------------------	------

CEP - UNISINOS
 VERSÃO APROVADA
 Em: 29 / 10 / 11

.....


APÊNDICE B

Questionários, avaliação pelos usuários e notas de campo do pesquisador.

1) Questionário estruturado

ATENÇÃO:

As informações recebidas (inclusive sua identidade) serão mantidas confidenciais. Este questionário possui duas etapas, sendo a primeira para fins de identificação e a segunda contendo questões de avaliação de usabilidade do robô RITA.

Por favor, escreva em LETRA DE FORMA e MAIÚSCULA, para facilitar nossa leitura.

Caso cometa algum erro, não se preocupe: basta colocar um traço sobre o dado incorreto.

PARTE 1 – Fale um pouco sobre você

Número de identificação: [preenchimento pelo pesquisador] Idade:

Formação: () até Ensino Médio () Graduação () Pós-graduação

Área de Formação: () Exatas () Humanas () Biológicas

Você é uma pessoa com deficiência: () Não () Sim

() Auditiva () Visual

() Completa () Parcial

PARTE 2 – Questionário de Usabilidade

Facilidade de Uso	Discordo plenamente	Discordo Parcialmente	Nem concordo, nem discordo	Concordo Parcialmente	Concordo plenamente
1. Aplicação é fácil de entender					
2. Aplicação é fácil de usar					
3. Com pequeno esforço posso encontrar informações acadêmicas					
4. Com pequeno esforço posso encontrar informações sobre minha relação com a Biblioteca					
5. As opções são claras e objetivas					
Percepção de Utilidade					
6. As opções de interação são relevantes					
7. O robô facilita acessar meus dados acadêmicos					
8. O robô facilita a obtenção de informações referentes ao campus					
9. O robô facilita encontrar pontos de interesse no campus					
10. Se eu precisasse dessas informações, eu usaria o robô					
Conforto/Eficácia					
11. Senti-me confortável ao interagir livremente com a aplicação.					
12. Senti-me confortável ao desempenhar a tarefa solicitada pelo pesquisador.					
13. Consegui desempenhar a tarefa solicitada	Não ()	Sim ()			

2) Respostas dos participantes, organizados em 3 grupos, ao questionário estruturado

Deficientes Visuais Completos	Idade	Formação	Área	Questões objetivas													
				1	2	3	4	5	6	7	8	9	10	11	12	13	
Usuário 1	53	Ensino Médio	-	4	4	5	3	5	5	2	5	5	5	5	5	5	SIM
Usuário 2	50	Graduação Incompleta	Humanas	4	4	5	5	5	5	5	4	4	5	5	5	5	SIM
Usuário 3	60	Graduação	Humanas	4	4	4	4	4	4	5	3	4	5	4	4	4	SIM
Usuário 4	44	Ensino Médio	-	5	5	5	5	5	5	5	4	4	5	5	5	5	SIM
Usuário 5	22	Graduação Incompleta	Humanas	5	5	4	4	4	4	4	2	2	4	5	5	5	SIM
Usuário 6	45	Pós-graduação	Humanas	5	4	5	5	5	5	5	5	5	4	4	4	4	SIM
Usuário 7*	31	Graduação	Exatas	5	5	5	5	5	4	4	4	4	4	5	5	5	SIM
Usuário 8*	32	Graduação	Exatas	5	5	5	5	5	5	5	5	5	4	5	5	5	SIM

* Usuário realizou o experimento de forma simulada

Deficiência Auditiva Completa	Idade	Formação	Área	Questões objetivas												
				1	2	3	4	5	6	7	8	9	10	11	12	13
Usuário A	14	Ensino Médio		4	4	4	3	3	5	3	5	3	5	5	3	SIM
Usuário B	11	Ensino Médio		4	4	5	5	5	5	5	3	3	5	5	3	SIM
Usuário C	19	Ensino Médio		2	4	4	3	5	4	4	2	2	4	2	2	NÃO
Usuário D	15	Ensino Médio		2	2	1	1	1	4	3	3	3	4	2	2	NÃO
Usuário E	51	Ensino Médio		5	5	5	5	5	5	5	5	5	5	5	5	SIM
Usuário F*	26	Graduação Incompleta	Humanas	5	5	5	5	5	4	5	5	4	5	5	5	SIM

* Usuário realizou o experimento de forma simulada

Deficientes Parciais	Deficiência	Idade	Formação	Área	Questões objetivas													
					1	2	3	4	5	6	7	8	9	10	11	12	13	
Usuário 7	Visual	32	Graduação	Exatas	5	5	5	5	5	5	5	5	5	4	5	5	5	SIM
Usuário 8	Visual	31	Graduação	Exatas	5	5	5	5	5	4	4	4	4	5	5	5	5	SIM
Usuário G	Auditiva	26	Graduação Incompleta	Humanas	5	5	5	5	5	4	5	5	4	5	5	5	5	SIM

3) Questionário semi-estruturado e respostas

1 - Em algum momento você se sentiu perdido, sem saber o que fazer? Se sim, quando isso ocorreu?

Usuário 1: Sim, ela (Rita) se perdia e não sabia o que fazer na minha primeira utilização"

Usuário 2: Não, fui bem orientado.

Usuário 3 a 8: Não

Usuário A: Não.

Usuário B: Sim, porque tive dificuldades de compreensão da língua portuguesa.

Usuário C: Sim, porque tive dificuldades de compreensão da língua portuguesa.

Usuário D: Sim, dificuldade em vencer as barreiras impostas pela língua portuguesa.

Usuários E e F: Não.

2 - Recomendaria alguma mudança no sistema para facilitar seu uso?

Usuário 1: Não.

Usuário 2: Um som que avisasse quando o usuário pode falar, em razão da falta de entonação do robô

Usuário 3: Aperfeiçoar o reconhecimento de face. Não aceitar interação de outra voz. Reconhecimento de digital, como a SIRI.

Usuário 4: (em branco)

Usuário 5: Componente de fone de ouvido para manter a privacidade das informações. Leitor de tela e teclado por meio do aplicativo Talk Back, o que facilitaria a utilização em local ruidoso.

Usuário 6: Possibilidade de interromper quando já está familiarizado com as opções. Implementar mais serviços que atendam a variadas demandas.

Usuário 7: Quando o robô inicia a fala das opções, devemos esperar ele terminar para seguir, porém seria bom poder continuar sem esperar ele terminar

Usuário 8: Acredito que poderia haver uma opção para parar ou voltar imediatamente ao invés de esperar toda a informação ser lida.

Usuários A, B, C, E, F, G: (em branco)

Usuário D: Uso de sinais como método de entrada e saída de dados, isto é, em toda a interação.

3 - Alguma funcionalidade ou recurso do sistema lhe pareceu mais útil?

Usuário 1: Serviço disciplinas. Informações concentradas de diversos setores

Usuário 2: Localização para cegos.

Usuário 3: Localização, porque as orientações de humanos não são precisas (vai "reto", enquanto o conceito de reta compreende ponto A, onde se está, e ponto B, onde se quer ir e cujo caminho se desconhece). Linguajar acessível, compatível e objetivo. Voz feminina é mais solitária, mais sensível de acordo com a experiência diária, em que as mulheres oferecem muito mais ajuda.

Usuário 4: Notas

Usuário 5: Localização

Usuário 6: Localização.

Usuário 7: O serviço de localização. Sugestão: aumentar os detalhes. Exemplo: 5metros à direita, 2ª porta.

Usuário 8: (em branco)

Usuários: A, B, C, D, F e G: (em branco)

Usuário E: Biblioteca.

4 - Alguma funcionalidade ou recurso não despertou seu interesse?

Usuário 1,2, 7 e 8: (em branco)

Usuário 3: Mão robótica.

Usuário 4: Todas interessantes.

Usuário 5: Localização, pela forma como a informação é dada (em metros)

Usuário 6: Mão robótica.

Usuários A, B, C, D, E e F: (em branco)

5- Quais mudanças tornariam o robô mais útil?

Usuário 1: Bom funcionamento (não apresentar erros).

Usuário 2: Detectar usuário com chapéu e óculos.

Usuário 3: Locomover-se com a pessoa para os destinos. Ampliação dos recursos e funcionalidades.

Usuário 4: Se ele identificasse melhor as palavras.

Usuário 5: Compatibilizar o referencial, uma vez que cegos possuem dificuldade de abstração em metros. As pessoas não costumam dar informações confiáveis com esse parâmetro. Considero mais fácil informação que diga corredor 2, 2ª porta à direita.

Usuário 6: Dispor o robô em local de fácil acesso aos cegos.

Usuário 7: Acredito que o robô atende bem os recursos existentes. Alguns recursos poderiam apresentar divisões (notas).

Usuário 8: O recurso de localização pode ser melhor explorado e pode ser melhorado. Um giroscópio já poderia ser mais útil.

Usuário A: (em branco)

Usuário B: Ajustes na parte elétrica

[Nota do autor: durante o teste, o circuito elétrico do braço ficou inoperante]

Usuário C : Se ele desempenhasse tarefas que ajudassem o usuário, como levar até locais, servir comidas, aprender outras coisas

Usuário D: Emprego da língua de sinais.

Usuário E: Que caminhasse até o ponto de interesse.

Usuário F: Fone de ouvido para deficientes parciais

4) Notas de campo

Usuário 2: Iniciação orientada no começo. Passou a se soltar e a interagir de forma independente e confiante. Testou funções voltar, resetar e ajuda. Postura física relaxada. Problemas operacionais de localização.

Usuário 3: Apesar de orientado, empregou Linguagem natural no começo da interação. Adaptou-se. Baixa confiabilidade do sistema em reservas x resetar, faltas x voltar.

Usuário 4: Interagiu de forma espontânea e descontraída, elevando seu nível de confiança durante os testes. Sorriu ao tocar nos dedos do robô quando indicaram números. Baixa confiabilidade na identificação de faltas (fraldas). Localização ok. Sistema reinicializado uma vez por um erro inesperado na consulta à biblioteca.

Usuário 5: Usuário realizou o teste rapidamente, sem testar todas as funcionalidades

Usuário 6: Usuário usava um tom de voz baixo, dificultando a captação do som pelo robô. Iniciou a interação de forma mais tensa, relaxando durante os testes.

Usuário A enfrentou problemas operacionais do robô (screenshot), mas entendeu a lógica da aplicação.

Usuário B apresentou um desempenho mais desinibido.

Usuários C e D estavam visivelmente pouco confortáveis com a aplicação, principalmente em razão das palavras empregadas no display, que consideravam difíceis.

Relato da intérprete de LIBRAS: Os participantes sentiram falta da ajuda da mão robótica, mas encontraram com facilidade o número de faltas. Acham que falta o apoio da LIBRAS. Um dos alunos gostaria que a tela fosse grande, os outros acham que seria difícil implantar uma tela grande por que seria cara ou ficaria pesada. Um dos alunos gostaria que o robô agradecesse.