

UNIVERSIDADE FEDERAL DO PARANÁ

RÚBIA ELIZA DE OLIVEIRA SCHULTZ ASCARI

METODOLOGIA BASEADA EM VISÃO COMPUTACIONAL E APRENDIZADO DE
MÁQUINA PARA GUIAR O PROJETO DE SISTEMAS DE COMUNICAÇÃO
AUMENTATIVA E ALTERNATIVA USANDO INTERAÇÃO GESTUAL PERSONALIZADA

CURITIBA PR

2020

RÚBIA ELIZA DE OLIVEIRA SCHULTZ ASCARI

METODOLOGIA BASEADA EM VISÃO COMPUTACIONAL E APRENDIZADO DE
MÁQUINA PARA GUIAR O PROJETO DE SISTEMAS DE COMUNICAÇÃO
AUMENTATIVA E ALTERNATIVA USANDO INTERAÇÃO GESTUAL PERSONALIZADA

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Ciência da Computação no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Luciano Silva.

Coorientador: Roberto Pereira.

CURITIBA PR

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

A811m Ascari, Rúbia Eliza de Oliveira Schultz
Metodologia baseada em visão computacional e aprendizado de máquina para guiar o projeto de sistemas de comunicação aumentativa e alternativa usando interação gestual personalizada [recurso eletrônico] Rúbia Eliza de Oliveira Schultz Ascari. – Curitiba, 2020.

Tese - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2020.

Orientador: Luciano Silva.
Coorientador: Roberto Pereira.

1. Interação homem-máquina. 2. Aprendizado do computador. 3. Visão Computacional. I. Universidade Federal do Paraná. II. Silva, Luciano. III. Pereira, Roberto. IV. Título.

CDD: 006.37

Bibliotecária: Vanusa Maciel CRB- 9/1928

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **RÚBIA ELIZA DE OLIVEIRA SCHULTZ ASCARI** intitulada: **Metodologia baseada em Visão Computacional e Aprendizado de Máquina para guiar o projeto de sistemas de Comunicação Aumentativa e Alternativa usando interação gestual personalizada**, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 19 de Junho de 2020.

Assinatura Eletrônica
13/07/2020 18:05:00.0

LUCIANO SILVA
Presidente da Banca Examinadora

Assinatura Eletrônica
07/07/2020 14:19:04.0

MARIA CECÍLIA CALANI BARANAUSKAS
Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica
06/07/2020 16:57:41.0

CLODIS BOSCARIOLI
Avaliador Externo (UNIVERSIDADE ESTADUAL DO OESTE DO
PARANÁ)

Assinatura Eletrônica
07/07/2020 15:56:57.0

NATASHA MALVEIRA COSTA VALENTIM
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

*À minha família, por todo o suporte
e confiança.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, que de forma singular me sustentou e me guiou para alcançar esta grande conquista.

Ao meu amado esposo Rafael, pela compreensão, incentivo, amor e companheirismo.

Aos meus filhos, Júlia e Davi, por compreenderem minhas ausências e me proporcionarem tantas alegrias.

Aos meus familiares, especialmente meus pais Rubin e Zélia e meus sogros Antonio e Salete, pelo carinho, incentivo e apoio constante.

Ao meu professor orientador Luciano Silva e meu professor coorientador Roberto Pereira, pela competência, colaboração, paciência e confiança.

À banca examinadora pelas excelentes contribuições que permitiram o aprimoramento desta tese.

Aos meus amigos e colegas do Dainf, especialmente Eliane De Bortoli Fávero, Soelaine Rodrigues Ascari e Robison Cris Brito. Embarcamos nessa jornada juntos, e foi muito bom poder compartilhar as incertezas, medos e conquistas com vocês.

Aos voluntários, professores, alunos e seus pais/representantes legais, e instituições coparticipantes que foram fundamentais para a construção desta tese: Universidade Tecnológica Federal do Paraná (UTFPR - Câmpus Pato Branco/PR), Secretaria Estadual de Educação do Estado do Paraná (SEED/PR), Núcleo Regional de Educação do município de Pato Branco, Colégio Estadual de Pato Branco (Premem), Centro Estadual de Educação Básica para Jovens e Adultos (CEEJA Pato Branco), Escola Carlos Almeida Educação Básica na Modalidade Educação Especial (APAE Pato Branco) e Colégio Estadual Castro Alves.

À UTFPR - Câmpus Pato Branco, pelo apoio e pela concessão do período de afastamento para que eu pudesse me dedicar exclusivamente aos estudos.

Ao Departamento Acadêmico de Informática da UTFPR - Câmpus Pato Branco e da UFPR por terem realizado um acordo de cooperação, que permitiu que professores da área de informática do sudoeste do Paraná pudessem se capacitar em um programa de Pós-Graduação de qualidade como o PPGInf.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) que apoiou direta ou indiretamente o andamento desta pesquisa, sendo instituições fundamentais no desenvolvimento tecnológico e científico do Brasil.

A todas as pessoas que de alguma forma contribuíram, me apoiaram e fizeram parte dessa jornada única.

RESUMO

A comunicação verbal é essencial para a socialização, construção e compartilhamento de conhecimento em uma sociedade. Pessoas que apresentam comprometimento de fala podem apresentar dificuldades de interação e aprendizagem, convivência familiar, entre outras situações que podem levar a problemas emocionais, sociais e cognitivos. Muitas vezes, essas pessoas também apresentam comprometimento motor associado, o que dificulta ainda mais sua capacidade de interagir com outras pessoas e com o ambiente em que estão inseridas. A Comunicação Aumentativa e Alternativa (CAA) é uma área de pesquisa que envolve o uso de modos não orais como complemento ou substituto da linguagem falada, aumentando a interação comunicativa de indivíduos sem oralidade. A CAA suportada por sistemas de Visão Computacional (VC) pode se beneficiar do reconhecimento dos movimentos funcionais restantes do usuário como uma abordagem alternativa de *design* de interação. Nesta tese, é apresentada uma nova metodologia, chamada MyPGI (*Methodology to yield Personalized Gestural Interaction*), construída para guiar o projeto de sistemas de CAA para pessoas com dificuldades motora e de fala, usando técnicas de VC e Aprendizado de Máquina para possibilitar interação gestual personalizada e não-invasiva. Muitos dos trabalhos existentes na literatura concentram-se em situações e características específicas, oferecendo pouca ou nenhuma flexibilidade para as pessoas com deficiência e seus diferentes contextos de uso, exigindo, portanto, que as pessoas se adaptem ao sistema em vez de adaptar o sistema às diferentes necessidades dos usuários. A metodologia MyPGI foi empregada no desenvolvimento de um sistema de CAA, chamado PGCA (*Personal Gesture Communication Assistant*), usando uma abordagem de baixo custo, em experimentos conduzidos com voluntários, incluindo alunos com dificuldades motora e de fala. Os resultados obtidos com os experimentos e também com entrevistas e avaliação de usabilidade conduzidas com profissionais da área de educação especial indicam a viabilidade da metodologia proposta e do sistema desenvolvido, além de agregar conhecimento sobre os desafios e necessidades específicas do público-alvo. A principal contribuição científica desta tese é, portanto, a definição da metodologia MyPGI, com o objetivo de promover maior autonomia para usuários de sistemas computacionais que possuem dificuldade motora e de fala e para seus cuidadores. A contribuição tecnológica desta tese se refere à produção de um sistema de CAA como uma forma de aplicação da metodologia proposta.

Palavras-chave: Visão Computacional, Interação Humano-Computador, Comunicação Aumentativa e Alternativa, Tecnologia Assistiva, Aprendizado de Máquina

ABSTRACT

Verbal communication is essential for socialization, building and sharing knowledge in society. People with speech impairment may experience difficulties in interaction and learning, family life, among other situations that can lead to emotional, social, and cognitive problems. Often, these people also have associated motor impairment, which further hinders their ability to interact with other people and the environment in which they are inserted. Augmentative and Alternative Communication (AAC) is a research area that involves the use of non-oral modes as a complement or substitute for spoken language, increasing the communicative interaction of individuals without orality. The AAC supported by Computer Vision (CV) systems can benefit from the recognition of the user's remaining functional movements as an alternative design approach to interaction. In this thesis, a new methodology, named MyPGI (Methodology to yield Personalized Gestural Interaction), is presented, built to guide the design of AAC systems for people with motor and speech difficulties, using CV techniques and Machine Learning to enable personalized and non-invasive gestural interaction. Many of the works in the literature focus on specific situations and characteristics, offering little or no flexibility for people with disabilities and their different contexts of use, thus requiring people to adapt to the system instead of adjusting the system to different user needs. The MyPGI methodology was used in the development of an AAC system, named PGCA (Personal Gesture Communication Assistant), employing a low-cost approach, in experiments conducted with volunteers, including students with motor and speech difficulties. The results obtained with the experiments and also with interviews and usability evaluation conducted with professionals from the Special Education area indicate the feasibility of the proposed methodology and the system developed, in addition to adding knowledge about the specific challenges and needs of the target audience. The main scientific contribution of this thesis is, therefore, the definition of the MyPGI methodology, to promote greater autonomy for users of computer systems that have motor and speech difficulties and for their caregivers. The technological contribution of this thesis refers to the production of an AAC system as a way of applying the proposed methodology.

Keywords: Computer Vision, Human-Computer Interaction, Augmentative and Alternative Communication, Assistive Technology, Machine learning

LISTA DE FIGURAS

1.1	Experimentos conduzidos com o objetivo de avaliar o sistema PGCA e a viabilidade da metodologia MyPGI. Foi utilizada a notação BPMN, dividindo em raias os responsáveis ou atores participantes de cada experimento. Os fluxos de trabalho indicam a ordem de execução dos experimentos.	27
2.1	Exemplos de estratégias de CAA de baixa tecnologia: a) Cartões de comunicação com símbolos gráficos representativos de mensagens, organizados por categorias de símbolos identificadas por molduras de cores diferentes (como expressões sociais, sujeitos, verbos); b) Prancha de comunicação temática com símbolos, fotos ou figuras, visualizando-se na imagem uma prancha com o tema "animais"; c) Prancha de comunicação alfabética, na qual o usuário aponta com o dedo indicador na letra "X".	30
2.2	Exemplos de estratégias de CAA de alta tecnologia, em que é possível a seleção de símbolos ou figuras para construção de frases: a) Interface web do sistema computacional de CAA chamado SCALA, onde pode-se visualizar a seleção de figura referente a expressão "eu quero", e possíveis complementos para a frase; b) Interface do sistema computacional de CAA chamado MeaVox, executado em dispositivo móvel, onde pode-se visualizar as seleções realizadas para construção da frase "eu quero beber suco de abacaxi".	30
2.3	Exemplo de um fluxo tradicionalmente empregado em sistemas para reconhecimento de gestos, usando vídeo. Como entrada a câmera captura uma imagem da mão do usuário que é processada para extrair a área de interesse (<i>pixels</i>) e segmentá-la em regiões (objetos). Cada objeto é representado por um vetor de características, com posição (x, y, z) e tamanho (t) que podem ser classificados para representar dados para a tomada de decisão no sistema. Neste exemplo, são identificadas seis regiões na imagem, das quais três representam dedos da mão de uma pessoa, indicando uma possível ação no sistema.	35
2.4	Principais subdivisões e aplicabilidades do Aprendizado de Máquina: aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem semi-supervisionada e aprendizagem por reforço.	38
3.1	Número de artigos publicados por ano, dentre os estudos selecionados no mapeamento sistemático.	51
3.2	<i>TagCloud</i> destacando a frequência de termos utilizados nas palavras-chaves dos estudos selecionados no mapeamento sistemático.	53
3.3	Partes do corpo humano utilizadas em pesquisas selecionadas no mapeamento sistemático que empregaram técnicas de Visão Computacional em recursos de Tecnologia Assistiva para pessoas com dificuldades motoras.	54

4.1	Estrutura da metodologia proposta, representada usando a notação BPMN. A metodologia é dividida em quatro raias que definem responsabilidades na execução das atividades e como os fluxos de trabalho cruzam funções ou transferem a atividade para outro responsável. As atividades relacionadas são agrupadas em etapas para facilitar sua descrição e seu entendimento..	68
4.2	Estrutura definida na metodologia MyPGI para a Etapa 1 (Calibração)..	73
4.3	Imagem representativa do processo de calibração realizado pelo sistema PGCA, com o objetivo de posicionar o usuário no centro de captura da câmera. Esse processo exige que a ponta do nariz do usuário passe em um momento pelo centro da imagem capturada, identificado por duas linhas retas que se cruzam. O círculo maior apresentado na figura corresponde à região da face do usuário e o círculo menor corresponde à localização do nariz do usuário.	74
4.4	Estrutura definida na metodologia MyPGI para a Etapa 2 (Criação de conjunto de dados personalizado)..	74
4.5	Exemplos de representação de movimento por meio de MHI (A) e OF-MHI (B) referente a um gesto executado com a cabeça.	75
4.6	Pontos usados para rastrear e calcular o fluxo óptico de um gesto de mão (movimentos para cima e para baixo) e de uma expressão facial (usuário levanta as pálpebras em uma expressão de surpresa). As imagens em tons de cinza correspondem à representação do movimento por meio de OF-MHI.	77
4.7	Estrutura definida na metodologia MyPGI para a Etapa 3 (Treinamento para reconhecimento de gestos)..	77
4.8	Exemplo de aplicação de Aumento de Dados. A partir de uma imagem representativa de gesto dinâmico (imagem original), outras oito variações são geradas por meio de operações de rotação e dimensionamento.. . . .	78
4.9	Esquema empregado pelo classificador com base em SVM para a etapa de treinamento do sistema PGCA. As amostras criadas de forma personalizada e representadas como imagens de histórico de movimento tem suas características descritas via descritor HOG, para então proceder a classificação dos gestos. . . .	79
4.10	Esquema empregado pelo classificador com base em CNN e Transferência de Aprendizado para a etapa de treinamento do sistema PGCA. A rede neural <i>TensorFlow Inception V3</i> é treinada originalmente com o conjunto de imagens coloridas <i>Imagenet</i> . O conhecimento adquirido nesta etapa de treinamento é mantido, e posteriormente apenas a última camada da rede é retreinada usando o conjunto de imagens (em tons de cinza) criado pelo usuário do sistema em conjunto com seu cuidador. Assim a rede torna-se apta a classificar os gestos personalizados.	80
4.11	Estrutura definida na metodologia MyPGI para a Etapa 4 (Avaliação do desempenho do sistema)..	80
4.12	Estrutura definida na metodologia MyPGI para a Etapa 5 (Interação gestual personalizada com sistema de CAA)..	81

4.13	Área do Cuidador onde conjuntos de dados são criados e o sistema é treinado para o reconhecimento de gestos. (A) Exemplo de captura de uma amostra de gesto. (B) Exemplo de mensagem apresentada ao usuário após análise da amostra por algoritmo de similaridade. (C) Mensagem informativa sobre o número de amostras capturadas por gesto. (D) Apresentação do desempenho do sistema após executar a etapa de treinamento e posterior avaliação do sistema.	83
4.14	Exemplo de uso da Área do Usuário para o reconhecimento de gestos. A borda amarela indica que o gesto está em execução, e a borda verde indica que o gesto foi concluído e reconhecido pelo sistema. O rótulo associado ao gesto reconhecido é apresentado em uma barra de progresso, a figura correspondente é apresentada ao usuário, o nome da classe é inserido na caixa de texto, e o som correspondente é reproduzido por um sintetizador de voz.	84
4.15	Tela de configurações da Área do Usuário, onde é possível customizar o funcionamento do sistema PGCA, relacionar gestos com a execução de gatilhos no sistema e com a simulação de uso de algumas teclas do teclado. Esta tela é oculta por padrão e exibida quando o usuário acessa a opção representada por um ícone de "engrenagem".. . . .	85
4.16	Exemplo de uso da Área do Usuário para a seleção de figuras em prancha de comunicação apresentada por padrão pelo sistema utilizando interação gestual. Ao selecionar uma opção no primeiro agrupamento apresentado pela prancha de comunicação, a navegação ocorre em um agrupamento menor, e então o usuário pode selecionar a figura referente à palavra que deseja escrever e o som correspondente é emitido.. . . .	86
4.17	<i>Layout</i> de pranchas alfabéticas para CAA: (A) <i>Layout</i> utilizado como base para criação de pranchas adaptadas; (B) <i>Layout</i> de prancha alfabética adaptada completa, contendo inclusive caracteres especiais e comandos diversos; (C) <i>Layout</i> de prancha alfabética adaptada simplificada, contendo letras, números e comandos básicos.	87
4.18	Área de Pranchas de Comunicação onde o cuidador pode gerar pranchas personalizadas usando imagens.. . . .	88
4.19	Área de Jogo, onde o usuário pode utilizar gestos para movimentar uma tartaruga com o objetivo de leva-la até o mar. (A) Tela inicial do jogo, a borda amarela indica que um movimento está sendo executado; (B) Exemplo de uso do jogo e alteração de estado da tartaruga, a borda verde indica que um gesto foi detectado e reconhecido.. . . .	89
4.20	Tela apresentada ao usuário após a conclusão do objetivo do jogo, que é levar a tartaruga até o mar. Uma imagem com balões e confetes é apresentada ao usuário, o estado da tartaruga é alterado, e um som de aplausos é reproduzido.	90
5.1	Exemplos de amostras de gestos executados pelos voluntários V1, V2 e V3 (pessoas sem deficiência motora e de fala), registradas durante a execução do Experimento 2 e representadas por meio de MHI convencional.. . . .	100
5.2	Exemplos de amostras de gestos executados pelos voluntários V4 e V5 (pessoas sem deficiência motora e de fala), registradas durante o Experimento 2 e representadas por meio de MHI convencional.	101

5.3	Amostras de gestos do <i>Keck Gesture Dataset</i> utilizado no Experimento 3, representados por MHI convencional e MHI com base em fluxo óptico (OF-MHI). As imagens em tons de cinza foram redimensionadas para dar ênfase às regiões em movimento.	104
5.4	Métodos (A) ABACADA e (B) Boquinhos empregados na alfabetização de crianças com necessidades especiais, citados por profissionais em entrevistas. . .	108
5.5	Exemplos de amostras de gestos executados pelos Alunos A e B (pessoas com dificuldades motora e de fala), registradas durante o Experimento 4 e representadas por meio de OF-MHI.	111
5.6	Pranchas de comunicação utilizadas pelos alunos com dificuldades motora e de fala durante o Experimento 4 para selecionar imagens e escrever em um editor de texto, e para indicar a letra inicial de uma palavra.	112
5.7	Exemplos de amostras de gestos executados pelos Alunos C e D (pessoas com dificuldades motora e de fala), registradas durante o Experimento 4 e representadas por meio de OF-MHI.	113
5.8	Prancha de comunicação usada para buscar um site do Ministério Público Brasileiro no navegador da Internet.	114
5.9	Exemplos de amostras de gestos executados pelo Aluno A (pessoa com dificuldade motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.	119
5.10	Exemplo de seleção de imagem em prancha de comunicação de figuras usando um gesto de cabeça. Esse teste de interação foi realizado por todos os alunos participantes do Experimento 5, cada um usando o seu conjunto de gestos personalizados.	120
5.11	Aluno A (pessoa com dificuldades motora e de fala) interagindo com a interface do sistema PGCA, jogando o jogo, como parte das tarefas executadas durante o Experimento 5.	121
5.12	Exemplos de amostras de gestos executados pelo Aluno B (pessoa com dificuldades motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.	122
5.13	Aluno B depois de interagir com a interface do sistema PGCA, e atingir o objetivo do jogo. Esta é uma das tarefas executadas durante o Experimento 5.	123
5.14	Exemplos de amostras de gestos executadas pela Aluna C (pessoa com dificuldades motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.	123
5.15	Aluna C interagindo com a interface do sistema PGCA, jogando o jogo, como parte das tarefas executadas durante o Experimento 5.	124
5.16	Pontuação obtida com a avaliação de usabilidade do sistema PGCA utilizando o questionário SUS: (A) Pontuação por participante (P1 a P5); (B) Pontuação média por questão (Q1 a Q10). A pontuação média do SUS de todos os participantes foi 81, que é aceitável.	127

6.1	Da personalização para a personificação: adequação da metodologia MyPGI levando em consideração a habilidade cognitiva e habilidade motora dos usuários, como uma possibilidade de trabalho futuro.	137
-----	---	-----

LISTA DE TABELAS

3.1	Exemplos de Sistemas de CAA considerados com objetivos semelhantes ou características desejáveis aos sistemas que podem ser desenvolvidos a partir da metodologia proposta..	42
3.2	Exemplos de trabalhos que empregaram reconhecimento de gestos, aplicando dispositivos diversos e técnicas de Visão Computacional para o desenvolvimento de soluções como recursos de Tecnologia Assistiva.	44
3.4	Protocolo de Estudo utilizado para condução do mapeamento sistemático, especificando o problema de pesquisa, seu objetivo, questão geral e questões de pesquisa..	47
3.5	Critérios de inclusão e exclusão empregados no primeiro filtro do mapeamento sistemático para seleção de estudos.	47
3.6	Critérios de seleção empregados no segundo filtro do mapeamento sistemático para seleção de estudos..	48
3.7	Expressões de busca empregadas e resultados obtidos em cada base de dados considerada no mapeamento sistemático: ACM, IEEE, Science Direct, e Springer. 33 estudos foram selecionados após a aplicação do primeiro e segundo filtros.	48
3.8	Formulário de extração de dados usado para padronizar o registro das informações obtidas a partir dos trabalhos resultantes do mapeamento sistemático.	49
3.9	Estudos selecionados categorizados por aplicação, região corporal utilizada para interação gestual e participação de pessoas com dificuldades motoras em alguma etapa do desenvolvimento da pesquisa.	52
3.10	Estudos selecionados no mapeamento sistemático, voltados para IHC com foco em viabilizar a comunicação, utilizando ou não o termo "Comunicação Aumentativa e Alternativa"..	53
3.11	Dispositivos e fonte de dados usados em estudos selecionados via mapeamento sistemático.	55
3.12	Métodos utilizados em trabalhos selecionados no mapeamento sistemático para realização das etapas de pré-processamento e segmentação de objetos, extração de características e representação.	55
3.13	Técnicas usadas para realizar a etapa de reconhecimento de gestos ou movimentos nos trabalhos selecionados no mapeamento sistemático..	56
4.1	Mudanças no estado da personagem tartaruga apresentadas durante a utilização do jogo desenvolvido..	89
5.1	Resumo das etapas de avaliação realizadas para avaliar o sistema PGCA e consequentemente a metodologia MyPGI.	94
5.2	Conjuntos de dados criados por voluntários no Experimento 2 — comparação de métodos de aprendizado de máquina: Precisão média, Cohen Kappa, Desvio padrão e Variância.	101

5.3	Uso do Keck Gesture Dataset no Experimento 3 — comparação de métodos de aprendizado de máquina: Precisão média, Cohen Kappa, Desvio padrão e Variância.	104
5.4	Questões utilizadas para guiar as entrevistas semi-estruturadas com especialistas de domínio.	106
5.5	Informações sobre os profissionais de educação especial (especialistas do domínio) entrevistados.	106
5.6	Características dos alunos com dificuldades motora e de fala participantes do Experimento 4.	110
5.7	Conjuntos de dados criados por alunos durante o Experimento 4 — comparação de desempenho de dois classificadores (SVM e CNN) e duas representações de movimento (MHI e OF-MHI): Precisão geral, Cohen Kappa, Desvio Padrão, Variância e Falsos Positivos.	115
5.8	Características dos alunos com dificuldades motora e de fala participantes do Experimento 5.	118
5.9	Tempo aproximado de execução dos classificadores para as etapas de treinamento e teste, usando os métodos Holdout e K-Fold no Experimento 5.	124
5.10	Conjuntos de dados criados por alunos durante a execução dos Experimentos 4 e 5 — desempenho de classificadores usando validação cruzada K-fold: Precisão, Kappa de Cohen, Desvio padrão, Variância e Falsos positivos.	125
5.11	Gestos reconhecidos pelo sistema durante a interação gestual personalizada do aluno A com o jogo e ações desencadeadas durante o Experimento 5.. . . .	125
5.12	Gestos reconhecidos pelo sistema durante a interação gestual personalizada do Aluno B com o jogo e ações desencadeadas durante o Experimento 5.	126
5.13	Gestos reconhecidos pelo sistema durante a interação gestual personalizada da Aluna C com o jogo e ações desencadeadas durante o Experimento 5.	126
5.14	Questionário SUS usado para avaliar a usabilidade do sistema PGCA após a execução de tarefas pelo público-alvo no Experimento 5. O questionário foi preenchido por profissionais da área de educação especial que acompanharam a condução do experimento.	127
5.15	Questionário usado para avaliar a metodologia MyPGI em um contexto diferente, por um pesquisador externo.	130

LISTA DE ACRÔNIMOS

AAM *Active Appearance Models*

ACM *Association for Computing Machinery*

AmCARA Ambiente e Comunicação Alternativo com Realidade Aumentada

ANN *Artificial Neural Network*

APAE Associação de Pais e Amigos do Excepcional

BCI *Brain-computer interfaces*

BOW *Bag of Words*

BPMN *Business Process Modeling Notation*

CAA Comunicação Aumentativa e Alternativa

CAAE Certificado de Apresentação para Apreciação Ética

CAVE *Cave Automatic Virtual Environment*

CNN *Convolutional Neural Network*

CRF *Conditional Random Fields*

DBN *Dynamic Bayesian Networks*

DCC *Design Centrado no Contexto*

DTW *Dynamic Time Warping*

EOG Eletrooculografia

EQM Erro Quadrático Médio

FPM *Frontier Point method*

GMM *Gaussian Mixture Model*

GRF Características Relacionais Geométricas

HMAGIC *Head Movement And Gaze Input Cascaded*

HMM *Hidden Markov Models*

HOG *Histogram of Oriented Gradients*

IHC *Interação Humano-Computador*

IHM *Interação Humano-Máquina*

IHR *Interação Humano-Robô*

INPI *Instituto Nacional da Propriedade Industrial*

KLT *Kanade-Lucas-Tomasi*

KNN *K-Nearest Neighbor*

KVAP *Kinect Virtual Art Program*

LED *Light Emitting Diode*

Libras *Língua Brasileira de Sinais*

MHI *Motion History Image*

MyPGI *Methodology to yield Personalized Gestural Interaction*

NWFE *Nonparametric Weighted Feature Extraction*

OF-MHI *Optical Flow-based Motion History Image*

OpenCV *Open Source Computer Vision*

OSC *Open Sound Control*

PECS *Picture Exchange Communication Systems*

PETRA *International Conference on PErvasive Technologies Related to Assistive Environments*

PGCA *Personal Gesture Communication Assistant*

RA *Realidade Aumentada*

RFID *Radio Frequency Identification*

ROC *Receiver Operating Characteristic*

SEED/PR *Secretaria Estadual de Educação do Estado do Paraná*

SIFT *Scale-Invariant Feature Transform*

SURF *Speeded Up Robust Features*

SUS *System Usability Scale*

SVM *Support Vector Machine*

TA *Tecnologia Assistiva*

TALE *Termo de Assentimento Livre e Esclarecido*

TAM *Technology Acceptance Model*

TCLE *Termo de Consentimento Livre e Esclarecido*

TDF *Transformada Discreta de Fourier*

TEA *Transtorno do Espectro Autista*

UFPR *Universidade Federal do Paraná*

UTFPR *Universidade Tecnológica Federal do Paraná*

VC *Visão Computacional*

VT *Vocabulary Tree*

LISTA DE SÍMBOLOS

α	taxa de atualização empregada para gerar OF-MHI
δ	parâmetro de decaimento empregado no MHI
ψ	função de atualização empregada para gerar MHI
τ	extensão temporal de um movimento (duração)

SUMÁRIO

1	INTRODUÇÃO	21
1.1	CONSIDERAÇÕES INICIAIS	21
1.2	DEFINIÇÃO DO PROBLEMA	22
1.3	MOTIVAÇÃO	24
1.4	OBJETIVOS	25
1.5	PASSOS METODOLÓGICOS	26
1.6	CONTRIBUIÇÕES	27
1.7	ORGANIZAÇÃO DA TESE	28
2	FUNDAMENTOS TEÓRICOS	29
2.1	COMUNICAÇÃO AUMENTATIVA E ALTERNATIVA	29
2.2	INTERAÇÃO HUMANO-COMPUTADOR	31
2.3	VISÃO COMPUTACIONAL	34
2.4	APRENDIZADO DE MÁQUINA	37
2.5	CONSIDERAÇÕES	39
3	ESTADO DA ARTE	41
3.1	ESTUDOS SOBRE CAA E VISÃO COMPUTACIONAL EM TECNOLOGIA ASSISTIVA	41
3.2	MAPEAMENTO SISTEMÁTICO - TENDÊNCIAS EM APLICAÇÕES DE VC PARA PESSOAS COM DIFICULDADES MOTORAS	46
3.2.1	Ameaças à validade do estudo	49
3.2.2	Resultados	50
3.2.3	Respostas às questões de pesquisa	62
3.2.4	Discussão	64
3.3	CONSIDERAÇÕES	65
4	METODOLOGIA MYPGI	66
4.1	PRESSUPOSTOS E VISÃO GERAL	66
4.1.1	Etapa 1 - Calibração	68
4.1.2	Etapa 2 - Criação de conjunto de dados personalizado	69
4.1.3	Etapa 3 - Treinamento para reconhecimento de gestos	69
4.1.4	Etapa 4 - Avaliação do desempenho do sistema	70
4.1.5	Etapa 5 - Interação gestual personalizada com sistema de CAA	71
4.2	CENÁRIOS DE APLICAÇÃO	71
4.3	IMPLEMENTAÇÃO - SISTEMA PGCA	72
4.3.1	Implementação da Etapa 1 - Calibração	73

4.3.2	Implementação da Etapa 2 - Criação de conjunto de dados personalizado.	74
4.3.3	Implementação da Etapa 3 - Treinamento para reconhecimento de gestos.	77
4.3.4	Implementação da Etapa 4 - Avaliação do desempenho do sistema	80
4.3.5	Implementação da Etapa 5 - Interação gestual personalizada com sistema de CAA	81
4.3.6	Considerações sobre as tecnologias empregadas	89
4.4	ETAPAS DE AVALIAÇÃO.	91
4.5	CONSIDERAÇÕES	92
5	RESULTADOS.	93
5.1	EXPERIMENTO 1: AVALIAÇÃO EXPLORATÓRIA DO SISTEMA POR PROFISSIONAIS QUE ATUAM COM IHC.	94
5.1.1	Resultado da avaliação	95
5.1.2	Discussão	98
5.2	EXPERIMENTO 2: CRIAÇÃO DE CONJUNTOS DE GESTOS POR USUÁ- RIOS SEM DIFICULDADES MOTORA E DE FALA	99
5.2.1	Conjuntos de dados criados.	99
5.2.2	Avaliação de desempenho dos classificadores	101
5.2.3	Discussão	102
5.3	EXPERIMENTO 3: AVALIAÇÃO DE TECNOLOGIAS USANDO <i>KECK GESTURE DATASET</i>	103
5.4	ENTREVISTAS COM ESPECIALISTAS DE DOMÍNIO.	105
5.4.1	Participantes.	106
5.4.2	Contribuições	106
5.5	EXPERIMENTO 4: USO DO SISTEMA PGCA POR ALUNOS COM DIFI- CULDADES MOTORA E DE FALA	109
5.5.1	Participantes.	109
5.5.2	Execução de tarefas por participante	110
5.5.3	Avaliação de desempenho do sistema no reconhecimento de gestos	115
5.5.4	Discussão	115
5.6	EXPERIMENTO 5: USO DO SISTEMA PGCA POR ALUNOS COM DIFI- CULDADES MOTORA E DE FALA INCLUINDO ABORDAGEM BASEADA EM JOGO.	117
5.6.1	Participantes.	118
5.6.2	Execução de tarefas por participante	118
5.6.3	Avaliação de desempenho do sistema no reconhecimento de gestos	124
5.6.4	Avaliação de usabilidade do sistema	126
5.6.5	Discussão	128
5.7	AVALIAÇÃO DA METODOLOGIA EM OUTRO CONTEXTO	129
5.7.1	Resultado da avaliação	130
5.8	CONSIDERAÇÕES	131

6	CONCLUSÃO	133
6.1	DISCUSSÃO	133
6.2	TRABALHOS FUTUROS	136
6.3	PUBLICAÇÕES GERADAS	138
	REFERÊNCIAS	139

1 INTRODUÇÃO

Este capítulo apresenta as considerações iniciais, definição do problema, motivação, os objetivos, passos metodológicos e contribuições desta tese. No final do capítulo encontra-se a organização da tese por meio de uma breve descrição dos demais capítulos.

1.1 CONSIDERAÇÕES INICIAIS

A linguagem oral é o meio mais usado para as pessoas se comunicarem, não obstante, a comunicação entre pessoas pode ser bem mais abrangente. O ser humano possui recursos verbais e não-verbais, que na interação interpessoal se misturam e se completam (SEESP, 2004). A comunicação entre pessoas é complementada por vários elementos comunicativos (como estados emocionais, gestos, expressões faciais) que permitem às pessoas se entenderem.

A comunicação é, portanto, essencial para a socialização (Turk, 2010). Quando há dificuldades em oralizar o que se pretende comunicar, surgem problemas de interação, dificultando a aprendizagem, a vida familiar, atividades sociais, práticas profissionais, entre outros, levando a problemas emocionais, sociais e cognitivos (Garbin et al., 2008). Assim, são necessárias alternativas para que os indivíduos com comprometimento de fala se comuniquem entre si e com o mundo ao seu redor. Nesse contexto, a Tecnologia Assistiva (TA) (Ravneberg e Söderström, 2017) tem potencial para contribuir com soluções interativas e de baixo custo, por meio do uso de recursos computacionais com a finalidade de proporcionar uma maior independência e autonomia às pessoas que possuem algum tipo de deficiência.

A Comunicação Aumentativa e Alternativa (CAA) visa complementar ou substituir a fala para compensar dificuldades de expressão usando sistemas de comunicação não-oral e estratégias de intervenção (Gevarter et al., 2013). A prática de CAA mediada por aplicações computacionais representa uma alternativa muito importante para seus usuários. As oportunidades de aprendizagem são numerosas e, considerando a realidade de pessoas já conectadas via *smartphones* e outros dispositivos móveis, criar meios para permitir o acesso de todos, em qualquer lugar e a qualquer momento, pode ser compensador. À medida que os recursos da CAA se tornam parte da vida diária das pessoas com o objetivo de trocas sociais eficientes, esses recursos contribuem para desenvolver sua cognição e linguagem, bem como na inclusão de pessoas com déficits cognitivos e comunicativos na escola e sociedade (Capovilla e Nunes, 2003).

O uso de computadores como ferramenta de comunicação é comum hoje em dia, porém o uso de sistemas de computador por pessoas com dificuldades motoras é particularmente desafiador (Ascari et al., 2020b). Muitas vezes, pessoas com deficiência de fala também apresentam comprometimento motor associado, o que torna ainda mais difícil a interação com o ambiente em que estão inseridas. Para pessoas que são parcialmente incapazes de mover ou controlar seus membros e não podem confiar na comunicação verbal em função de alguma deficiência, pode ser necessário o uso de tecnologias específicas, como as baseadas em reconhecimento de gestos ou informações obtidas via interface cerebral, para viabilizar a interação com sistemas computacionais de CAA (por exemplo, os sistemas desenvolvidos por Tu et al. (2007) e Gomez-Donoso et al. (2016)). Uma interface capaz de interpretar os movimentos voluntários limitados desses usuários pode permitir a sua comunicação com amigos, parentes e provedores de cuidados, ou enviar comandos para um sistema por meio da Interação Humano-Computador (IHC) (Montanini et al., 2015).

Permitir que usuários com deficiências motoras e de fala acessem computadores, utilizando-os como ferramenta educacional ou de treinamento, representa uma oportunidade importante para que exerçam suas habilidades comunicativas, influenciando afetivamente, emocionalmente e, conseqüentemente, contribuindo para sua formação e aprendizado (Ascari et al., 2018a). A área de conhecimento da TA, por meio de recursos, serviços, estratégias e práticas específicas, ajuda as pessoas com perdas funcionais advindas das deficiências ou como resultado do processo de envelhecimento a serem incluídas socialmente e se tornarem ou permanecerem independentes.

Existem muitos dispositivos de entrada e diferentes tecnologias que abrem novos paradigmas na IHC e podem beneficiar pessoas com deficiências. Sistemas com base em interação multimodal (ou seja, que tem a capacidade de se comunicar com um usuário por meio de diferentes modos de comunicação) proporcionam melhor acessibilidade a seus usuários, uma vez que uma interface multimodal pode ser utilizada por uma pessoa com deficiência por meio dos recursos de interação que consegue manipular, além de tornar o sistema mais flexível e com maior comodidade (Lalanne et al., 2009).

Hoje, praticamente todos os computadores comuns e mesmo os telefones celulares são equipados com uma câmera, portanto, as interfaces de Visão Computacional (VC) (Chen, 2015) podem ser facilmente disponibilizadas para todos nestes dispositivos. O acesso facilitado a dispositivos com câmera deu origem a uma nova geração de recursos de TA que não envolvem dispositivos eletromecânicos customizados e caros para acomodar necessidades especiais de acesso, mas são baseados em *software*, permitindo a redução de custos e a disponibilidade aprimorada de tecnologias, como previsto por Betke et al. (2002). Técnicas não-invasivas baseadas em VC (obtidas via câmera) permitem que métodos de interação não convencionais sejam considerados, incluindo o reconhecimento de movimentos das mãos, cabeça, rosto e outras partes do corpo para execução de ações em sistemas computacionais (Marnik, 2014). O reconhecimento de expressões faciais por meio de câmera também tem sido investigado, como na ferramenta desenvolvida por Pradi (2016) voltada ao reconhecimento e produção de expressões faciais por pessoas com Transtorno do Espectro Autista (TEA).

Por meio do reconhecimento de gestos, os seres humanos podem interagir diretamente com máquinas sem a necessidade de outros dispositivos, como mouse ou teclado, como no sistema proposto por Leite et al. (2017), que permite controlar aplicações CAVE (*Cave Automatic Virtual Environment*) baseadas no reconhecimento estático de gestos manuais. Este tipo de interação tem o potencial de considerar particularidades e limitações de cada usuário na realização de movimentos, sendo considerado "natural", e até mesmo intuitivo, pois os humanos aprendem a usar gestos desde a infância (Chen, 2013). Embora soluções em interação gestual tenham se popularizado, a sua aplicação como apoio à CAA ainda necessita de experimentos para avaliação de tecnologias e exemplos que demonstrem sua viabilidade técnica e que permitam o desenvolvimento de soluções computacionais de baixo custo para atender a diversidade das pessoas em suas capacidades e limitações físicas, cognitivas, sociais e econômicas (Ascari et al., 2020b).

1.2 DEFINIÇÃO DO PROBLEMA

Pessoas que possuem deficiência motora associada a alguma dificuldade de fala demandam meios alternativos para interagir com sistemas computacionais. Para essas pessoas, a interação gestual pode ser uma maneira de viabilizar a interação com sistemas de CAA quando a utilização de dispositivos tradicionais não é possível, oportunizando dessa forma todos os benefícios associados.

Muitas iniciativas já foram desenvolvidas empregando técnicas de VC para gerar recursos de TA como modos alternativos de interação para pessoas com dificuldade motora. No entanto, dentre os estudos existentes, percebe-se uma tendência em explorar o rastreamento de regiões corporais específicas (Ascari et al., 2019, 2020a). Ou seja, muitos sistemas específicos têm sido desenvolvidos, voltados exclusivamente ao rastreamento e detecção de gestos de mão, ou movimentos de cabeça, ou expressões faciais, entre outros. Essas soluções são em geral mais adaptadas, porém, pouco flexíveis, abrangendo um público alvo bastante específico. Observa-se também na literatura uma predominância no uso de conjuntos de gestos predefinidos (Ascari et al., 2019, 2020a), ou seja, o usuário precisa se adaptar aos gestos que o sistema é capaz de reconhecer, e não o contrário. Sistemas que consideram movimentações feitas com diferentes regiões corporais como um gesto válido para interação gestual podem estar mais próximos de atender as demandas reais das pessoas com dificuldades motoras. Como as habilidades motoras desses usuários podem ser muito particulares, a interação gestual para esse público-alvo deveria idealmente explorar os movimentos que cada usuário é capaz de executar, independente da região corporal utilizada.

Usuários com comprometimento motor podem apresentar posturas muito particulares e movimentos involuntários que às vezes podem ser incontroláveis, além de fadiga de curto prazo e diminuição (ou aumento) de capacidades motoras. Esses usuários podem se beneficiar do uso de soluções com base em interação gestual, porém, estão sujeitos a alterações esporádicas em suas capacidades motoras, o que pode implicar em modificações significativas na forma como os movimentos são executados. Nesse sentido, um sistema que possibilite ao usuário e seu cuidador personalizarem os gestos utilizados para interação, permitindo de forma facilitada a alteração de gestos previamente configurados, substituição ou inclusão de novos gestos, pode ser um importante passo para prover autonomia e poder de interação a essas pessoas.

Em se tratando de recursos de TA como um todo, algumas questões precisam ser melhor exploradas. Rivera e DeSouza (2014) destacam problemas enfrentados pela TA e que ainda são atuais:

p.14 "Em essência, o problema real dos recursos de Tecnologia Assistiva de hoje, como em muitas outras áreas da engenharia, é que o fardo de se adaptar a essas novas tecnologias ainda recai sobre o usuário. Somos constantemente forçados a aprender um novo conceito, uma nova linguagem ou um novo jargão para interagir com um novo *software* ou dispositivo, e essa imposição deve ser ainda menos aceitável quando aplicada a pessoas com deficiências. Além disso, à medida que a doença progride, o uso de qualquer interface específica pode tornar-se cada vez mais difícil devido à perda do controle motor de um músculo ou de um membro inteiro. Eventualmente, o usuário será forçado a aprender como usar uma nova interface, levando a tempo de inatividade, perda de independência e todos os efeitos psicológicos associados. É hora de mudar essa imagem: tempo para criar dispositivos de Tecnologia Assistiva que exijam menos intrusões e menos eletrodos; dispositivos que forneçam vários modos de interação ao mesmo tempo, para que o usuário possa alternar de um modo para outro em caso de fadiga de curto prazo e/ou perda de músculos específicos em longo prazo; e, finalmente, dispositivos que podem se adaptar perfeitamente à intenção do usuário — entender e antecipar essas intenções, em vez de obrigar o entendimento e a adaptação ao dispositivo."(Rivera e DeSouza, 2014) — traduzido pela autora.

Com o objetivo de gerar soluções computacionais que levem em consideração as características de pessoas com dificuldades motora e de fala, e buscando minimizar os problemas apontados por Rivera e DeSouza (2014), esta tese propõe uma metodologia voltada para guiar o projeto de sistemas de CAA para usuários com dificuldades motora e de fala, fazendo uso de técnicas de VC e Aprendizado de Máquina (Alpaydin, 2020) para viabilizar a interação gestual personalizada e não-invasiva. A interação personalizada é proposta com o objetivo de dar condições para que o usuário com o apoio de seu cuidador tenha autonomia para gerar (e atualizar sempre que necessário) um conjunto de gestos próprios que será utilizado para treinar um sistema computacional de CAA, com base em interação gestual. O usuário e seu cuidador poderão, portanto, criar uma linguagem de sinais personalizada, com propósitos de comunicação, levando em consideração as capacidades e limitações do usuário na realização de movimentos e também a experiência do cuidador na interpretação desses movimentos.

1.3 MOTIVAÇÃO

Muitas pessoas que sofrem com algum tipo de deficiência precisam de assistência para participar da vida social e econômica, e esse apoio geralmente vem de familiares, cuidadores e/ou soluções tecnológicas que facilitam a interação com o ambiente em que estão inseridos por meio da IHC (Kbar et al., 2017). Embora o desenvolvimento de recursos de TA e de sistemas computacionais interativos tenha avançado nos últimos anos, a acessibilidade na Computação ainda é um desafio, pois algumas modalidades de interação ainda são bastante restritivas para alguns usuários, como os usuários com deficiência física.

A deficiência física pode ser definida como "uma variedade de condições não sensoriais que afetam o indivíduo em termos de mobilidade, coordenação motora geral ou de fala, como decorrência de lesões neurológicas, neuromusculares e ortopédicas, ou ainda, de malformação" (Brasil, 1999). O aparelho locomotor compreende o sistema osteoarticular, o sistema muscular e o sistema nervoso. Doenças ou lesões que afetam quaisquer desses sistemas, isoladamente ou em conjunto, podem produzir quadros de limitações físicas de grau e gravidade variáveis, segundo os segmentos corporais afetados e o tipo de lesão ocorrida (Alves et al., 2006). Dentre diversas potencialidades, ferramentas computacionais podem apoiar, ou tornar possível, a comunicação de pessoas com seu meio social, especialmente aquelas com dificuldades e limitações originadas de alguma deficiência.

A Convenção sobre os Direitos das Pessoas com Deficiência, adotada pela Assembleia Geral das Nações Unidas, apresenta a acessibilidade às tecnologias de informação e comunicação como parte integrante dos direitos de acessibilidade, em paridade com a acessibilidade ao meio físico e ao transporte (Narasimhan, 2010). Essa noção de acessibilidade implica que tanto o *hardware* quanto o *software*, incluindo a interface humana, precisam ser acessíveis (Narasimhan, 2010). Uma possível perspectiva para o conceito de acessibilidade está fortemente relacionada à ideia de *design* universal (Steinfeld e Maisel, 2012), que se refere a tornar as coisas tão acessíveis quanto possível para um grupo de pessoas na sua maior extensão possível.

Acesso participativo e universal do cidadão brasileiro ao conhecimento constava dos Grandes Desafios da Pesquisa em Computação no Brasil 2006-2016 (Carvalho et al., 2006). A partir desse desafio, a acessibilidade e inclusão digital foram listadas também como parte dos Grandes Desafios de Pesquisa para a Interação Humano-Computador no Brasil (Baranauskas et al., 2015). O desenvolvimento de sistemas e interfaces para TA requer muito cuidado do ponto de vista da interação das pessoas com esses sistemas. O que funcionaria para uma população em geral não pode ser assumido para alguns grupos de usuários pretendidos (Munteanu et al., 2016). Selecionar um recurso de TA requer maximizar o fluxo de informações e minimizar o esforço

(físico, mental e emocional) para usá-lo (Abascal, 2008), sendo particularmente importante considerar as perspectivas dos usuários e seus cuidadores para se adotar (ou desenvolver) algo que funcione e faça sentido para as pessoas envolvidas no contexto de uso da solução.

Problemas de usabilidade e acessibilidade tornam-se críticos em TA — que visa justamente reduzir barreiras e promover a autonomia das pessoas em suas mais diferentes tarefas (Ascari et al., 2018b). Não há como um recurso de TA alcançar seu propósito se ele mesmo introduz barreiras que dificultam o seu uso. Problemas de usabilidade e acessibilidade no projeto de recursos de TA têm sido causados muitas vezes por negligência no processo de desenvolvimento, participação de profissionais sem noções básicas de qualidade em IHC, baixo orçamento para os projetos, erro de cronograma e o emprego de abordagens centradas no desenvolvimento do aparato técnico. Muitas vezes soluções são desenvolvidas negligenciando as habilidades e a experiência que se precisa promover para as partes interessadas, as quais deveriam ser centrais.

Com o objetivo de contribuir com a pesquisa voltada ao desenvolvimento de soluções para TA, nesta tese são identificadas duas perguntas de pesquisa a serem respondidas:

- É possível criar uma metodologia generalizável para o projeto de sistemas de CAA com base em interação gestual personalizada?
- Por intermédio de experimentos e utilizando-se da metodologia proposta, é possível auxiliar pessoas com comprometimento motor e de fala a se comunicarem por meio de TA baseada em interação gestual personalizada?

1.4 OBJETIVOS

O objetivo geral desta tese é desenvolver uma metodologia baseada em técnicas de VC e Aprendizado de Máquina para guiar o desenvolvimento de sistemas de CAA usando interação gestual personalizada. Com base em uma metodologia aplicável no desenvolvimento de soluções com diferentes requisitos de interação gestual, espera-se prover direcionamentos para desenvolvedores de *software* contribuírem na promoção de autonomia aos usuários de sistemas computacionais de CAA, aprimorando suas capacidades comunicativas.

Como objetivos específicos destacam-se:

- Desenvolver metodologia baseada em VC e Aprendizado de Máquina para guiar o projeto de sistemas de CAA usando interação gestual personalizada;
- Empregar os pressupostos da metodologia criada no desenvolvimento de um sistema computacional de CAA, usando abordagem de baixo custo;
- Aplicar técnicas viáveis para combinação de diferentes modalidades de entrada (como gestos, expressão facial) e saída (por exemplo síntese de voz, sugestão visual) associados à metodologia criada para viabilizar a IHC multimodal;
- Avaliar o uso da metodologia por meio do sistema desenvolvido, aplicado a diferentes grupos de usuário (com e sem dificuldades motora e de fala) por meio de experimentos e metodologias de avaliação de IHC;
- Verificar a viabilidade da interação gestual personalizada, criada pelo usuário e seu cuidador, e se essa forma de interação contribui para a CAA desses usuários.

1.5 PASSOS METODOLÓGICOS

Com o objetivo de verificar a viabilidade da metodologia proposta e também analisar os seus resultados, esta tese trilhou os seguintes passos metodológicos:

- Criação de um panorama da literatura sobre estudos focados no desenvolvimento ou descrição de sistemas computacionais de CAA e também no desenvolvimento de meios alternativos de interação para pessoas com deficiência motora empregando técnicas de VC;
- Levantamento de instituições coparticipantes, incluindo instituições de ensino do município de Pato Branco/PR frequentadas por alunos com deficiência motora e de fala;
- Submissão de projeto para avaliação do Comitê de Ética em Pesquisa com Seres Humanos do Setor de Ciências da Saúde da Universidade Federal do Paraná (UFPR);
- Entendimento dos aspectos elementares sobre a rotina de alunos com deficiência motora e de fala em seu ambiente escolar;
- Estruturação de uma metodologia baseada em VC e Aprendizado de Máquina para guiar o projeto de sistemas de CAA usando interação gestual personalizada;
- Com base na metodologia proposta, construção de um sistema computacional de CAA;
- Avaliação do artefato tecnológico desenvolvido, aplicado a diferentes grupos de usuário (com e sem dificuldades motora e de fala) por meio de experimentos e metodologias de avaliação de IHC;

O processo de construção do sistema desenvolvido a partir da metodologia proposta seguiu uma abordagem de *design* orientada a experimentos. Seguiu-se um procedimento iterativo a partir de *insights* experimentais obtidos em cada etapa de avaliação para aprimoramento ou extensão do sistema e da metodologia nas etapas subsequentes. A Figura 1.1 apresenta as etapas de avaliação realizadas, destacando os conjuntos de dados utilizados e o objetivo de cada etapa. Foi utilizada a notação BPMN (*Business Process Modeling Notation*), que é uma notação gráfica para modelagem de processos de negócios (Eroshkin et al., 2017), dividindo em raias os responsáveis ou atores participantes de cada experimento. Os fluxos de trabalho indicam a ordem de execução dos experimentos.

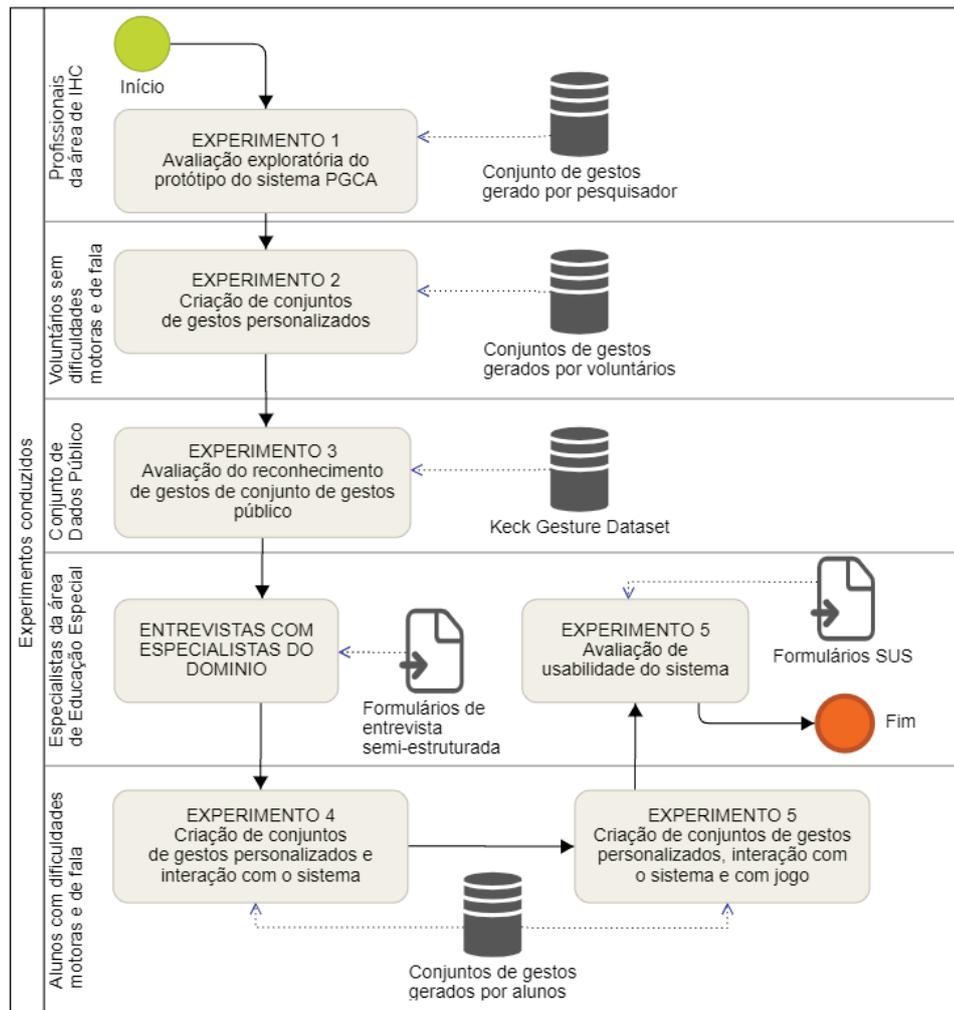


Figura 1.1: Experimentos conduzidos com o objetivo de avaliar o sistema PGCA e a viabilidade da metodologia MyPGI. Foi utilizada a notação BPMN, dividindo em raias os responsáveis ou atores participantes de cada experimento. Os fluxos de trabalho indicam a ordem de execução dos experimentos.

1.6 CONTRIBUIÇÕES

A principal contribuição científica desta tese é a definição de uma metodologia que visa guiar o projeto de sistemas de CAA usando interação gestual personalizada, promovendo maior autonomia para usuários de sistemas computacionais que possuem dificuldade motora e de fala e também para seus cuidadores. A efetividade dessa inovação foi avaliada por meio de experimentos com um sistema computacional de CAA desenvolvido especialmente com este propósito, possibilitando que as perguntas de pesquisa desta tese fossem respondidas. A contribuição tecnológica desta tese se refere à produção de um sistema de CAA como uma forma de aplicação da metodologia proposta.

Os esforços empregados deram origem aos seguintes resultados:

- Uma metodologia baseada em Visão Computacional e Aprendizado de Máquina, chamada MyPGI (*Methodology to yield Personalized Gestural Interaction*), construída para guiar o projeto de sistemas de CAA usando interação gestual personalizada^{1,2}.
- Um sistema computacional de CAA, chamado PGCA (*Personal Gesture Communication Assistant*), desenvolvido seguindo os pressupostos da metodologia MyPGI, incluindo uma abordagem baseada em jogo interativo por gestos³.
- Descrição do efeito do emprego da metodologia MyPGI a partir do uso do sistema PGCA por voluntários, incluindo alunos com dificuldades motora e de fala^{3,4}.
- Um registro de *software* de cotitularidade entre UFPR e UTFPR, certificado no INPI (BR512020000890-9).

1.7 ORGANIZAÇÃO DA TESE

Esta tese é composta por capítulos, organizados como descrito a seguir.

Capítulo 2 — Fundamentos Teóricos — abrange temas e conceitos relacionados a esta tese. A CAA e IHC são apresentadas. Por fim, são apresentadas VC e Aprendizado de Máquina juntamente com as principais técnicas utilizadas para reconhecimento de gestos.

Capítulo 3 — Estado da Arte — apresenta uma revisão dos trabalhos na literatura relacionados a esta tese. São apresentados os principais trabalhos focados em CAA com base em TA. E também, trabalhos que apresentam soluções para IHC baseada em VC e sua aplicação em recursos de TA, dando ênfase aos trabalhos voltados a pessoas que possuem dificuldades motora e de fala.

Capítulo 4 — Metodologia MyPGI — apresenta a metodologia MyPGI, desenvolvida para apoiar o projeto de sistemas de CAA usando interação gestual personalizada. Este capítulo descreve a estrutura da metodologia, pressupostos, possíveis cenários de aplicação e um exemplo de implementação da metodologia (sistema PGCA). Também é apresentado um breve resumo das etapas executadas para avaliação da metodologia.

Capítulo 5 — Resultados — apresenta os resultados obtidos a partir de diferentes experimentos conduzidos com o objetivo de avaliar a proposta desta tese. Também são apresentados resultados de entrevistas e avaliação de usabilidade conduzidas com profissionais da área de educação especial, e avaliação da metodologia em outro contexto.

Capítulo 6 — Conclusão — apresenta a discussão dos resultados obtidos, trabalhos futuros e publicações geradas a partir desta tese de doutorado, seguidas das referências bibliográficas.

¹Ascari, R. E. O. S., Pereira, R. e Silva, L. 2018. Towards a Methodology to Support Augmentative and Alternative Communication by means of Personalized Gestural Interaction. In Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018). ACM, 38, 1-12. Belém/PA, Brasil.

²Ascari, R. E. O. S., Pereira, R. e Silva, L. 2020. Computer vision-based methodology to improve interaction for people with motor and speech impairment (Aceito para publicação). ACM Transactions on Accessible Computing (ACM TACCESS). ACM, 1-33.

³Ascari, R. E. O. S., Silva, L. e Pereira, R. 2019. Personalized Interactive Gesture Recognition Assistive Technology. In Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems (IHC 2019). ACM, 38, 1-12. Vitória/ES, Brasil.

⁴Ascari, R. E. O. S., Silva, L. e Pereira, R. 2020. Personalized gestural interaction applied in a gesture interactive game-based approach for disabled people. In Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI 2020). ACM, 1-12. Cagliari, Itália.

2 FUNDAMENTOS TEÓRICOS

Este capítulo apresenta conceitos básicos ou fundamentais para o entendimento da abrangência desta tese. São apresentados conceitos relacionados a CAA, IHC, VC, e Aprendizado de Máquina com o objetivo de destacar as possibilidades de aplicação deles no contexto desta tese, assim como destacar tópicos específicos e atuais na pesquisa envolvendo essas áreas de conhecimento. Neste contexto, a IHC está inerentemente relacionada à TA, aplicada nesta tese com o objetivo de fornecer recursos para melhorar ou viabilizar a comunicação de pessoas com comprometimento motor e de fala. Para isso, técnicas de VC e Aprendizado de Máquina são empregadas, pois podem efetivamente permitir a interação personalizada e não-invasiva, viabilizando a comunicação por meio de sistemas computacionais de CAA.

2.1 COMUNICAÇÃO AUMENTATIVA E ALTERNATIVA

O termo Comunicação Aumentativa e Alternativa foi traduzido do inglês *Augmentative and Alternative Communication*. Além do termo resumido "Comunicação Alternativa", no Brasil também são utilizadas as terminologias "Comunicação Ampliada e Alternativa" e "Comunicação Suplementar e Alternativa" (Sartoretto e Bersch, 2014). CAA refere-se a todas as formas de comunicação que possam complementar ou substituir a fala. A CAA cobre necessidades de recepção, compreensão e expressão verbal, aumentando a interação comunicativa dos indivíduos sem oralidade (Von Tetzchner e Jensen, 1996).

Sistemas de Comunicação "Aumentativa" suplementam a linguagem oral quando não é possível a um indivíduo se comunicar eficazmente com outras pessoas e com o ambiente em que está inserido (como um meio de comunicação para complementar ou compensar deficiências de fala, mas sem substituí-la totalmente). Já os sistemas de Comunicação "Alternativa" substituem a linguagem oral de um indivíduo quando ela não pode ser entendida ou é perdida (como um meio de comunicação que difere da fala, substituindo-a). Ambos os tipos de sistema podem ser utilizados por pessoas com problemas de comunicação na interação com outras pessoas, expressando suas opiniões, sentimentos e tomando decisões pessoais para liderar e controlar suas vidas (Sanz et al., 2013).

Muitas pessoas que não falam usam tecnologias de CAA para aprimorar sua comunicação (por exemplo gestos não verbais e sons não léxicos, tais como o riso), bem como uma alternativa ao discurso oral (Alper, 2015). Os dispositivos CAA variam de baixa tecnologia, como cartões com fotos e placas de comunicação, a versões de alta tecnologia, como placas de comunicação eletrônicas, sintetizadores de voz computadorizada, ou *software* específicos. Interfaces tangíveis têm sido utilizadas com sucesso para potencializar a CAA, como no trabalho de Moreira e Baranauskas (2018) que investigaram o uso da tecnologia de RFID em cartões de comunicação para trabalhar a CAA via lúdico.

Para pessoas com deficiência de linguagem expressiva, esses dispositivos fornecem ferramentas que permitem a seleção de palavras, símbolos e imagens para comunicar seus pensamentos, intenções e conversas com outras pessoas por meio de voz digitalizada ou sintetizada (Alper, 2015).

As Figuras 2.1 e 2.2 apresentam, respectivamente, exemplos de estratégias de CAA de baixa e alta tecnologia. Na Figura 2.1 são apresentados cartões de comunicação com símbolos gráficos representativos de mensagens, pranchas de comunicação temáticas e alfabética, utilizadas pelos usuários para expressar seus desejos, vontades, assim como um recurso facilitador no

processo de interação e desenvolvimento da linguagem (Sameshima et al., 2009). A Figura 2.2 apresenta exemplos de interfaces de sistemas computacionais de CAA que permitem a comunicação expressiva para indivíduos com dificuldades de comunicação, por meio de composição de mensagens e leitura via sintetizador de voz.

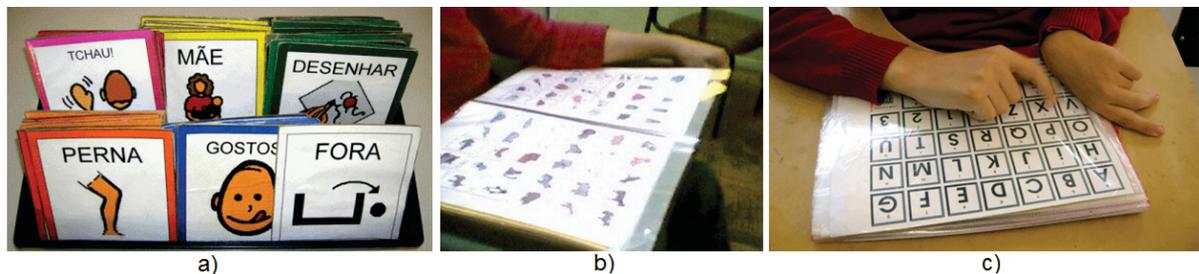


Figura 2.1: Exemplos de estratégias de CAA de baixa tecnologia: a) Cartões de comunicação com símbolos gráficos representativos de mensagens, organizados por categorias de símbolos identificadas por molduras de cores diferentes (como expressões sociais, sujeitos, verbos); b) Prancha de comunicação temática com símbolos, fotos ou figuras, visualizando-se na imagem uma prancha com o tema "animais"; c) Prancha de comunicação alfabética, na qual o usuário aponta com o dedo indicador na letra "X".

Fonte: Sartoretto e Bersch (2014).

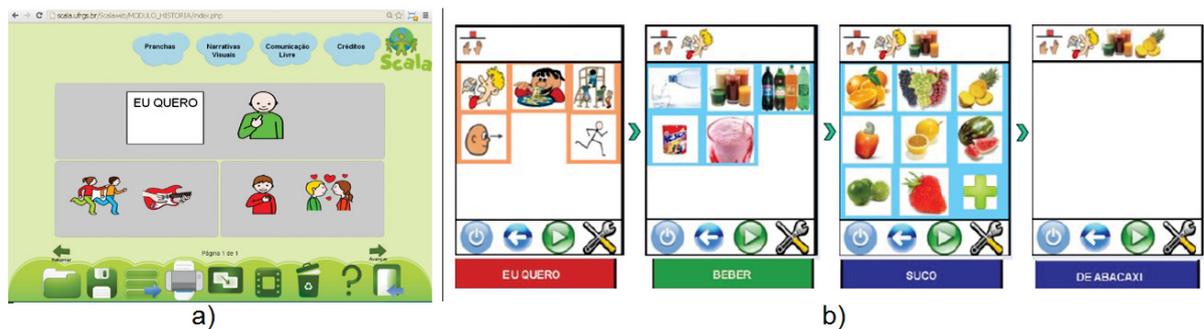


Figura 2.2: Exemplos de estratégias de CAA de alta tecnologia, em que é possível a seleção de símbolos ou figuras para construção de frases: a) Interface web do sistema computacional de CAA chamado SCALA, onde pode-se visualizar a seleção de figura referente a expressão "eu quero", e possíveis complementos para a frase; b) Interface do sistema computacional de CAA chamado MeaVox, executado em dispositivo móvel, onde pode-se visualizar as seleções realizadas para construção da frase "eu quero beber suco de abacaxi".

Fonte: Bez (2014), Caminha (2018).

No passado, pode ter havido críticas sobre a utilidade da CAA. Alguns autores argumentaram que a CAA afetaria a aprendizagem negativamente, uma vez que um indivíduo poderia preferir usá-la do que desenvolver capacidades de fala (Kangas e Lloyd, 1988). Atualmente, a CAA é vista como essencial para o desenvolvimento de indivíduos com dificuldades de comunicação, podendo inclusive auxiliar indivíduos com deficiência intelectual em seu processo de aprendizagem (Deliberato, 2005; Coe et al., 2015).

As tecnologias de acesso para comunicação começaram a partir de uma tarefa mais simples e mais lenta, como as placas E-tran (sistema de comunicação que permite a seleção de letras com movimentos dos olhos). Assim, com base no mesmo princípio de rastreamento ocular com pranchas físicas de olhar e nos princípios de reflexões oculares de Purkinje (formados por luz refletida da superfície frontal da córnea), nasceu a tecnologia de infravermelho e pesquisas de rastreamento ocular (Nerişanu et al., 2017). Essas tecnologias permitiram que os usuários com deficiência escrevessem palavras e frases no computador e conversassem por meio de voz sintetizada. Além disso, com o avanço das pesquisas na área de VC, dispositivos rastreadores de

movimentos estão se tornando mais acessíveis, eficientes e portáteis, encontrando sua aplicação em produtos do dia-a-dia, como *tablets* e *smartphones*, cadeira de rodas ou na própria cama do usuário. Para interagir efetivamente com esses dispositivos, critérios de qualidade em IHC devem ser considerados no desenvolvimento de sistemas interativos, de modo a contemplar os interesses de seus usuários e garantir a qualidade de uso, como descrito na seção a seguir.

2.2 INTERAÇÃO HUMANO-COMPUTADOR

A ACM (*Association for Computing Machinery*) define IHC como "uma disciplina preocupada com o *design*, avaliação e implementação de sistemas computacionais para uso humano e com o estudo dos principais fenômenos que os rodeiam" (Hewett et al., 1992). A qualidade em IHC pode ser vista como um conjunto de fatores importantes que influenciam a experiência do usuário e está relacionada a conceitos como usabilidade e acessibilidade.

Segundo a ISO 9241-11, usabilidade é a extensão na qual um produto pode ser usado por um grupo específico de usuários para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto de uso específico (ISO/IEC, 1998). Já a acessibilidade pode ser entendida como a flexibilidade proporcionada para o acesso à informação e à interação, de maneira que usuários com diferentes necessidades possam acessar e usar esses sistemas (Melo e Baranauskas, 2006). A adaptabilidade também é um fator importante a ser considerado, pois refere-se à capacidade do produto de software de ser adaptado para diferentes ambientes especificados, sem necessidade de aplicação de outras ações ou meios além daqueles fornecidos para essa finalidade pelo software considerado (ISO/IEC, 2001).

Uma possível perspectiva para o conceito de acessibilidade está fortemente relacionada à ideia de *design* universal (Story et al., 1998; Steinfeld e Maisel, 2012), que se refere a tornar as coisas tão acessíveis quanto possível para um grupo de pessoas o mais amplo possível, com pouco ou nenhum custo extra. O *design* universal traz benefícios para pessoas de diferentes idades e habilidades e a contribuição de IHC é que ela deve investigar, desenvolver e aproveitar essas novas áreas de possibilidade não apenas como tecnologias ou projetos, mas como um meio para melhorar a atividade, a experiência e a vida humana (Carroll, 2013).

O desenvolvimento de interação multimodal entre humanos e computadores busca fornecer meios alternativos e mais robustos de comunicação para com as aplicações e sistemas computacionais interativos (Vidakis et al., 2016). Está relacionado à integração de diferentes modalidades de entrada (como voz, gestos, olhar, emoções) e saída (por exemplo síntese de voz, sugestão visual), podendo potencialmente gerar interações mais intuitivas e acessíveis. Embora a interação multimodal nem sempre seja a melhor e preferível abordagem (Oviatt, 1999), seu potencial deve ser considerado no contexto assistivo.

A interação humana com o mundo é, inerentemente, multimodal (Quek et al., 2002). Assim, há um esforço crescente da comunidade científica para alavancar as habilidades de comunicação humana por meio da fala, gestos, toque, expressão facial e outras modalidades de comunicação com sistemas interativos (Turk, 2014). Considerando que os seres humanos interagem com o mundo principalmente por meio dos seus sentidos principais (visão, audição, tato, paladar e olfato), o objetivo da pesquisa nesta área é desenvolver tecnologias, métodos de interação e interfaces para eliminar limitações existentes, utilizando esses sentidos em conjunto (quando possível) para uma interação mais intuitiva.

A interação multimodal ganhou especial relevância com o surgimento de dispositivos de detecção/reconhecimento de movimentos corporais e gestuais de baixo custo associados a

consoles de videogame, como EyeToy1 (Playstation)¹, Wii Remote (Wii)² ou Kinect (Xbox)³ (Fernandes et al., 2015). Equipamentos independentes de console como o Leap Motion⁴ ou Parallax Si1143⁵ permitem a identificação de gestos dos dedos usando imagens obtidas via câmeras de infravermelho. O bracelete Myo⁶ identifica gestos mediante a detecção de atividade elétrica nos músculos do braço do usuário, empregando uma técnica conhecida como eletromiografia. Paralelamente à interação gestual de baixo custo, a realidade virtual e a realidade aumentada têm experimentado um ressurgimento, por meio de monitores de imersão de baixo custo e óculos de realidade aumentada. Com a ideia da computação de vestir, o conceito de joias inteligentes (pulseiras, braceletes, relógios) torna-se uma realidade que permite a seus usuários esquecerem a presença de seus dispositivos (Ju e Spasojevic, 2015).

É importante destacar que o crescente interesse na área de IHC, juntamente com todos esses dispositivos, amplia as possibilidades para criação de novas técnicas e o surgimento de aplicações de interação multimodal como soluções de TA. A pesquisa em IHC muitas vezes abrange a interação entre humanos e máquinas (IHM - Interação Humano-Máquina), e a interação entre humanos e vários sistemas robóticos (IHR - Interação Humano-Robô) (Myers, 1998). A IHM concentra-se na comunicação, cooperação e interação entre humanos e máquinas de maneira natural (Ma et al., 2019). Já o objetivo fundamental da IHR é desenvolver os princípios e algoritmos para sistemas robóticos que os tornam capazes de interação direta, segura e eficaz com seres humanos (Feil-Seifer e Mataric, 2009).

Assim como todo produto, processo ou *software* precisa passar por etapas de avaliação, recursos de TA também carecem de investigações nesse sentido, principalmente porque são destinados a um público que possui diferenças visuais, auditivas, mentais, físicas e/ou múltiplas (de Godoi e Valentim, 2019). Para que essas tecnologias se tornem mais acessíveis para o público pretendido, é importante considerar aspectos como acessibilidade, usabilidade e adaptabilidade, evitando o abandono do uso dos recursos de TA por parte dos usuários.

Existem diferentes técnicas de avaliação para apoiar avaliadores. Gediga et al. (2002) e Azarian e Siadat (2011) classificam as técnicas de avaliação em duas categorias: as técnicas de avaliação descritiva e as técnicas de avaliação preditiva, as quais devem estar presentes em todas as avaliações.

Técnicas de avaliação descritiva são usadas para descrever o *status* e os problemas reais do *software* de maneira objetiva, confiável e válida. Essas técnicas são baseadas no usuário e podem ser subdivididas em várias abordagens, como:

- Avaliação com base em comportamento: registram o comportamento do usuário enquanto utilizam um sistema. Esses procedimentos incluem técnicas de observação, protocolos de *Thinking aloud* (Nielsen, 1994) e o método de gravação de vídeo ou registro de *log* (Azarian e Siadat, 2011). O resultado desses métodos é um protocolo de entrevista cujas perguntas são focadas principalmente em pontos críticos, como interações. *Thinking aloud* permite verbalização simultânea de pensamentos durante a execução de uma tarefa, pois os usuários são solicitados a descrever suas considerações, problemas e alternativas enquanto realizam tarefas ou durante a exploração de um sistema por sua iniciativa. Na gravação de *log*, todas as observações e, em particular,

¹<https://www.playstation.com/en-us/games/eyetoy-play-2-with-camera-ps2/>

²<http://www.wii.com>

³<https://www.xbox.com/>

⁴<https://www.leapmotion.com>

⁵<https://www.parallax.com>

⁶<https://developerblog.myo.com/author/thalnic-labs/>

todas as interações com a interface do usuário dos sistemas são salvas em um arquivo. Uma gravação de vídeo pode ser preparada ao mesmo tempo.

- Avaliação com base em opiniões: visam obter as opiniões (subjetivas) do usuário. Exemplos são entrevistas, pesquisas e questionários, como o SUS (*System Usability Scale*) (Brooke, 1996), que visa revelar a opinião do usuário sobre a usabilidade do *software*. A aquisição de dados de questionários oferece a possibilidade de obter opinião sobre um determinado domínio do sistema. Resultados quantificados, como frequências, podem ser facilmente extraídos dos questionários. Questionários específicos ajudam a investigar áreas restritas dos sistemas de *software*, sendo que todos os tipos de questionário têm a sua utilidade. Nas entrevistas, o investigador precisa avaliar o *software* no *front end*, a fim de reconhecer possíveis pontos fracos e aprofundar o interrogatório. Se o número de testadores não for importante demais, as entrevistas continuam sendo uma boa técnica para elaborar a opinião dos usuários (Azarian e Siadat, 2011). Uma entrevista pode seguir diretrizes com entradas específicas e possíveis respostas (entrevista estruturada) ou de maneira não estruturada.
- Avaliação de usabilidade: Deriva de estudos clássicos de projeto experimental, sendo que o termo técnico "Teste de Usabilidade" é entendido como uma combinação de medidas baseadas em comportamento e opinião com alguma quantidade de controle experimental, geralmente escolhido por um especialista (Gediga et al., 2002).

Todas as técnicas de avaliação descritiva requerem algum tipo de protótipo e pelo menos um usuário (Nielsen, 1989). Além disso, para Gediga et al. (2002) os dados coletados por uma técnica descritiva precisam de alguma interpretação adicional por um ou mais especialistas para resultar em recomendações para o desenvolvimento futuro de *software*.

As técnicas de avaliação preditiva têm como objetivo prever aspectos da utilização não envolvendo utilizadores, permitindo fazer recomendações para o desenvolvimento futuro de *software* e a prevenção de erros de usabilidade. Essas técnicas são baseadas em especialistas - ou pelo menos em especialistas - como:

- Inspeções de especialistas e revisões heurísticas: o *software* é examinado por um especialista em usabilidade independente da equipe de desenvolvimento de *software*. Os especialistas anotam e avaliam alguns itens do *software* e geram uma lista de problemas como base para a otimização do *software*. As revisões heurísticas são uma variante desse método (Azarian e Siadat, 2011). A avaliação pode ser realizada em um estágio inicial de desenvolvimento com protótipos de papel. A avaliação heurística é um caso especial de uma inspeção especializada, e está focada nas seguintes heurísticas (Gediga et al., 2002): proporcionar um diálogo simples e natural; falar o idioma do usuário; minimizar a carga de memória do usuário; ser consistente; dar uma resposta; fornecer saídas claramente marcadas; fornecer atalhos; fornecer mensagens de erro informativas; evitar erros. Essas heurísticas devem ser consideradas como princípios gerais que orientam o avaliador durante as inspeções.
- Discussões em grupo: ajudam a resumir as ideias e comentários mantidos por membros individuais durante a avaliação da interface do usuário completa (Gediga et al., 2002). Cada participante age para estimular ideias e, por um processo de discussão, é estabelecida uma visão coletiva que é maior que as partes individuais (Azarian e Siadat, 2011). A discussão em grupo deve ser moderada por um especialista humano e é bastante flexível em sua orientação: pode ser focada em usuários, tarefas etc. Tende a ser mais eficiente se aplicada nos estágios iniciais do processo de desenvolvimento.

Técnicas de avaliação preditiva devem se basear em "dados". Mesmo que o especialista seja a força motriz desses métodos, os usuários também podem participar de alguns casos (Gediga et al., 2002). Em muitas técnicas de avaliação preditiva, esses "dados" são produzidos por especialistas que simulam usuários "reais". A objetividade e a confiabilidade dos critérios, que são a base das técnicas descritivas, são difíceis de aplicar nessa configuração. Como a validade deve ser o objetivo principal dos procedimentos de avaliação, há tentativas de provar a validade das técnicas de avaliação preditiva diretamente, por exemplo, comparando as taxas de "acertos" e "falsos alarmes" dos problemas detectados por uma técnica preditiva (Nielsen, 1994).

Para que um recurso de TA não seja abandonado e atenda a maior parte das necessidades dos usuários de forma clara e objetiva, é importante que a usabilidade, a acessibilidade e adaptabilidade sejam consideradas e trabalhadas em conjunto.

2.3 VISÃO COMPUTACIONAL

Visão Computacional refere-se à construção de descrições explícitas e significativas dos objetos em uma imagem (Ballard e Brown, 1982). É voltada ao estudo da obtenção e manipulação dos dados de uma imagem, incluindo quadros (*frames*) de vídeo, e do uso deles para diferentes propósitos.

Técnicas de VC são atraentes para implementar sistemas de entrada de gestos onde o objetivo é a captura de movimentos fluidos e livres das mãos ou do corpo do usuário (Wilson, 2006). Um gesto é a comunicação não-verbal feita com a mão, dedo, cabeça, rosto ou outra parte do corpo. Pessoas com deficiências motoras podem não conseguir produzir os mesmos gestos que outras pessoas, no entanto, elas ainda podem ser capazes de ativar outros músculos, mesmo com limitações de força ou duração dessa atividade. Sistemas computacionais de CAA com base em VC podem apoiar usuários com dificuldades motoras por meio do reconhecimento de movimentos funcionais restantes, especialmente partes do corpo humano que possam ser reconhecidas como um sinal visual (Gao e Duan, 2011).

Uma importante vantagem da VC é permitir a interação não-invasiva, ou seja, sem contato físico com o dispositivo. Praticamente, qualquer parte do corpo com mobilidade pode ser utilizada para realizar interação gestual, o que é especialmente importante para pessoas com comprometimentos físicos graves (Mauri et al., 2006). Conseqüentemente, essa forma de interação é extremamente flexível porque qualquer modificação detectada no vídeo é suscetível de ser interpretada pelo computador e utilizada para desencadear alguma ação. Também é possível combinar técnicas de VC com ferramentas distintas de entrada e saída a fim de explorar as potencialidades da interação multimodal em tempo real. Algumas das principais áreas de pesquisa da VC são o rastreamento ocular (movimento dos olhos), o reconhecimento de expressões faciais, o reconhecimento biométrico, o rastreamento de movimentos do corpo e o reconhecimento de gestos (linguagem gestual).

Na interação gestual, as imagens provenientes de uma ou mais câmeras são processadas para extrair características de movimentos que são interpretadas para classificação e reconhecimento por meio de técnicas de VC e Aprendizado de Máquina. O reconhecimento de gestos com base em VC é composto por etapas específicas que variam de acordo com o contexto da aplicação. A Figura 2.3 apresenta um diagrama de blocos de um fluxo padrão empregado em sistemas que usam como entrada imagens ou quadros de vídeo a partir de uma câmera. Às vezes, nem todos os blocos são incluídos em um sistema específico, mas a estrutura, no entanto, fornece uma diretriz relevante.

A etapa de aquisição de imagem, consiste basicamente na captura de uma imagem real e sua transformação em uma imagem digital, por intermédio de dispositivos que enviam

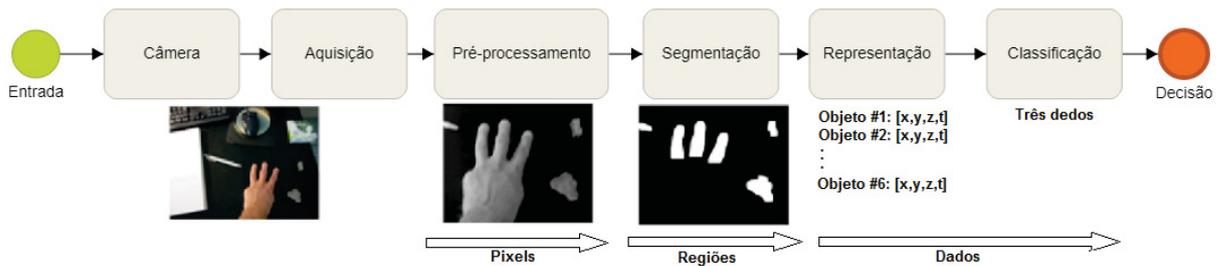


Figura 2.3: Exemplo de um fluxo tradicionalmente empregado em sistemas para reconhecimento de gestos, usando vídeo. Como entrada a câmera captura uma imagem da mão do usuário que é processada para extrair a área de interesse (*pixels*) e segmentá-la em regiões (objetos). Cada objeto é representado por um vetor de características, com posição (x, y, z) e tamanho (t) que podem ser classificados para representar dados para a tomada de decisão no sistema. Neste exemplo, são identificadas seis regiões na imagem, das quais três representam dedos da mão de uma pessoa, indicando uma possível ação no sistema.

Fonte: Moeslund (2012).

a informação da imagem adquirida a um conversor analógico digital (Baldner et al., 2017). Diferentes dispositivos de entrada podem ser utilizados para a aquisição de imagens, e a escolha do dispositivo pode interferir na qualidade da imagem obtida, permitindo o uso de mais ou menos informações nas etapas seguintes. São exemplos de dispositivos de aquisição utilizados em VC: *webcams*, câmeras de profundidade, câmeras térmicas, rastreadores de cabeça ou de olhar.

Em algumas situações, é necessário aplicar técnicas específicas para aprimorar a qualidade da imagem (corrigindo imperfeições, presença de *pixels* ruidosos, contraste e/ou brilho inadequado) e destacar regiões de interesse. Essa etapa é conhecida como pré-processamento e as técnicas empregadas envolvem métodos que operam no domínio espacial e métodos que operam no domínio da frequência (Gonzalez e Woods, 2002). Técnicas de processamento no domínio espacial baseiam-se em filtros que manipulam o plano da imagem, enquanto que as técnicas de processamento no domínio da frequência se baseiam em filtros que agem sobre o espectro da imagem (Gonzalez e Woods, 2002). É comum combinar vários métodos com base nestes dois domínios para realçar determinadas características de uma imagem, por exemplo para extrair regiões com cor de pele que representem dedos ou a mão de uma pessoa. Operações morfológicas são comumente empregadas nesta etapa, como a erosão (tende a remover objetos pequenos, possivelmente irrelevantes, por remoção de *pixels*) e dilatação (tende a fechar os buracos na imagem original, expandindo as regiões dos *pixels*, criando objetos maiores) que são aplicadas em uma imagem usando um conjunto de tamanho e forma conhecida, denominado elemento estruturante. Essas operações permitem construir funções importantes como abertura (erosão seguida de dilatação, eliminando conjuntos de *pixels* menores que o elemento estruturante) e fechamento (dilatação seguida de erosão, eliminando conjuntos de *pixels* maiores que o elemento estruturante), que podem suavizar as imagens de entrada de um sistema de VC.

Para algumas aplicações pode ser necessário segmentar objetos de interesse (veículos, pessoas, dedos da mão, olhos) a partir da imagem de entrada. Segmentar uma imagem significa, de modo simplificado, separá-la em partes que se diferenciam entre si. É usual denominar "objetos" da imagem os grupos de *pixels* de interesse, ou que fornecem alguma informação útil para a aplicação. Da mesma forma, a denominação "fundo" da imagem é utilizada para o grupo de *pixels* que podem ser desprezados ou que não têm utilidade para a aplicação. Ambos os grupos de *pixels* formam regiões na imagem sem que representem um objeto literalmente presente. A segmentação é considerada, dentre todas as etapas, a mais crítica, pois quaisquer erros ou distorções geradas nesta etapa refletem negativamente nas etapas seguintes e na eficiência do processo como um todo. Com base no posicionamento da câmera (estática ou em movimento), a

tarefa de segmentação pode ser feita de modos diferentes, como descrito por Ke et al. (2013), em uma revisão sobre reconhecimento de atividade humana baseada em vídeo.

Quando utilizada a câmera estática (a câmera é fixada em uma posição e ângulo específicos), um método de segmentação bastante utilizado é a subtração de fundo (a imagem de fundo sem qualquer objeto de primeiro plano é estabelecida pela primeira vez, e depois disso, a imagem atual na sequência de vídeo analisada é subtraída da imagem de fundo para obter os objetos em primeiro plano), que é simples e eficiente, mas sensível a mudanças de iluminação. Outros métodos com o mesmo propósito foram propostos, como o GMM (*Gaussian Mixture Model*) (Stauffer e Grimson, 1999), mistura de n distribuições gaussianas que apontam a mudança de estado de *pixels* correspondentes de um quadro a outro), que pode descrever um fundo mais complexo, mas implica em alto custo de processamento computacional, e o modelo estatístico (Horprasert et al., 1999), que é mais eficiente na construção de modelos de fundos complexos e pode ser usado para segmentar não apenas os objetos de primeiro plano, mas também as sombras (Ke et al., 2013). Esse último usa parâmetros estatísticos incluindo a mudança de intensidade e mudança de cromaticidade para cada *pixel*. Diferentemente dos métodos de segmentação citados anteriormente com base em pontos, é possível também segmentar o objeto de interesse rastreando regiões que são espacialmente coesas com movimento localmente suave (*smooth motion*), obtendo assim a informação global do objeto (região completa).

Já quando utilizada a câmera em movimento, é preciso considerar o movimento da câmera e a mudança de fundo. O método de segmentação mais comum nesse caso é a diferença temporal, ou seja, a diferença entre quadros de imagem consecutivos. Outra forma de segmentação para esta modalidade de câmera é o fluxo óptico (Horn e Schunck, 1981; Lucas e Kanade, 1981), que denota um deslocamento da mesma cena na sequência da imagem em um instante de tempo diferente. O fluxo óptico estima o movimento em nível de *pixel* entre duas imagens.

Após segmentada a região de interesse de uma imagem, pode-se executar a etapa de extração de características, resgatando dados sobre forma, silhueta, cores, movimentos ou outros considerados importantes, para posterior representação. De modo geral, as características podem ser categorizadas em quatro grupos, informações espaço-temporais, informações do domínio da frequência, descritores locais e modelagem corporal (Ke et al., 2013). Em uma sequência de vídeo, os recursos que capturam a relação de espaço e tempo são conhecidos como volumes espaço-temporais, formados pelo empilhamento temporário de quadros sobre uma sequência de vídeo como um cubóide 3D (tridimensional) de forma espaço-temporal. Diferentes métodos para extração de volumes espaço-temporais foram propostos, são exemplos: Blank et al. (2005), Ke et al. (2007) e Dollár et al. (2005). Além da informação espacial e temporal, a informação do domínio da frequência, ou seja, a variação de intensidade de uma imagem também pode ser aproveitada. É geralmente assumido que a intensidade do nível de cinza do objeto de primeiro plano é diferente daquela do objeto de segundo plano; portanto, a Transformada Discreta de Fourier (TDF) de uma imagem pode ser usada para obter informações da estrutura geométrica (forma) no domínio espacial. Geralmente, as características da TDF podem capturar a energia da forma dos quadros de imagem, mas também são sensíveis ao ruído e à oclusão.

Volumes espaço-temporais e a TDF são características globais que são extraídas considerando toda a imagem. No entanto, as características globais são sensíveis ao ruído, oclusão e variação do ponto de vista. Em vez de usar características globais, alguns métodos são propostos para considerar os fragmentos de imagens como características locais. Idealmente, as características locais são projetadas para serem mais robustas ao ruído e à oclusão e, possivelmente, para rotação e dimensionamento. São exemplos de descritores locais bastante utilizados: SIFT (Lowe, 1999), HOG (Dalal e Triggs, 2005), NWF (Lin et al., 2010), KLT (Tomasi e Kanade, 1991). Cada um deles apresenta vantagens e desvantagens que devem ser levadas em consideração

de acordo com o contexto da aplicação. Por exemplo, Jiang et al. (2017) empregaram descritores SIFT e BOW para reconhecimento de gestos de mão, Bougharriou et al. (2017) empregaram descritor HOG para detecção de veículos em um ambiente urbano.

Além de características globais e locais, outros métodos também são propostos para modelar direta ou indiretamente o corpo humano, para o qual as técnicas de estimativa de pose e rastreamento de partes do corpo podem ser aplicadas. Alguns métodos de modelagem humana foram propostos e aplicados com diferentes propósitos, incluindo *blobs* simples (Zhao et al., 2001), modelagem de corpos 2D (bidimensional) (Huo et al., 2009) e modelagem corporal 3D (Sedai et al., 2009). Além disso, as coordenadas da modelagem do corpo podem ser convertidas em características de menor dimensão ou mais discriminativas, como representação de coordenadas polares, características booleanas e Características Relacionais Geométricas (GRF) (Ke et al., 2013).

Depois de extraídas e representadas as características desejadas da imagem ou vídeo, o reconhecimento dessas características por meio de algoritmos de classificação são a próxima etapa a ser realizada. O objetivo do reconhecimento é realizar, de forma automática, a "identificação" dos objetos segmentados na imagem. Métodos como distância de Mahalanobis, Erro Quadrático Médio (EQM), distância euclidiana e *Dynamic Time Warping* (DTW) (Berndt e Clifford, 1994) podem ser usados para medir a similaridade entre duas sequências (um tipo de algoritmo de correspondência de modelos), possivelmente com diferentes comprimentos e diferentes taxas de ocorrência. Em alguns casos a aplicação desses métodos implica na necessidade de modelos extensivos para o conjunto de dados, resultando em alto processamento computacional. Por exemplo, somente para o reconhecimento de fala dos alfabetos de caracteres ingleses, para aplicar o DTW, pode ser necessário ter milhares de modelos com base em diferentes sotaques (Ke et al., 2013).

Para superar problemas dos algoritmos de medida de similaridade, métodos com base em modelos foram propostos. Basicamente, os algoritmos com base em modelos podem ser divididos em modelos gerativos e modelos discriminativos. Modelos gerativos, que simulam explicitamente o processo de geração das sequências de dados, como *Hidden Markov Model* (HMM) (Rabiner, 1986) e *Dynamic Bayesian Networks* (DBN) (Murphy e Russell, 2002), aprendem a distribuição de probabilidade conjunta $P(X, Y)$ sobre observação X e sequência Y , ou equivalentemente a probabilidade $P(X|Y)$. Por outro lado, modelos discriminativos aprendem a distribuição de probabilidade condicional (posterior) $P(Y|X)$ sobre um rótulo de classe Y de comprimento fixo não observado em um dado vetor de comprimento fixo observado X , como *Support Vector Machine* (SVM) (Cortes e Vapnik, 1995) e *Artificial Neural Network* (ANN) (McCulloch e Pitts, 1943). Outros algoritmos também utilizados para classificação são o filtro de Kalman (Welch e Bishop, 2006), árvore binária (Breiman, 2017), *Naive Bayes* (Duda et al., 1973), *K-Nearest Neighbor* (KNN) (Cover e Hart, 1967), *Conditional Random Fields* (CRF) (Lafferty et al., 2001), Adaboost (Viola e Jones, 2001) e *Vocabulary Tree* (VT) (Nister e Stewenius, 2006).

Após o reconhecimento de um objeto de interesse em uma imagem ou vídeo, ações podem ser executadas em uma interface, viabilizando assim a interação humano-computador baseada em um sistema de VC. Atraindo cada vez mais pessoas provenientes da academia e da indústria, no contexto da TA, os algoritmos de VC tem sido efetivamente explorados para atender às diferentes necessidades dos usuários (Leo et al., 2018).

2.4 APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina (*Machine Learning*) é o campo de pesquisa dedicado ao estudo formal de sistemas de aprendizado. Usado para ensinar às máquinas sobre como usar os

dados com mais eficiência (Khan et al., 2010), em termos gerais, o Aprendizado de Máquina refere-se a uma estrutura algorítmica que pode fornecer informações sobre os dados, facilitando a inferência e fornecendo uma configuração provisória para determinar os relacionamentos funcionais (Triantafyllidis e Tsanas, 2019). É uma área de pesquisa que avança rapidamente e tem sua popularidade ampliada em diversos domínios, como reconhecimento de fala e de escrita, Visão Computacional, diagnóstico médico, sistemas de busca, entre outros. Isso é em parte explicado pela amplitude das aplicações que abrange.

Um processo de Aprendizado de Máquina consiste em duas fases: a fase de aprendizado, na qual o sistema analisa os dados e gera regras, encontrando algumas semelhanças entre os dados, e a fase de validação, na qual as regras geradas devem ser verificadas, computando alguma função de avaliação de desempenho em um novo conjunto de dados (Jain e Kacprzyk, 2013). Algoritmos de Aprendizado de Máquina podem empregar diferentes abordagens de aprendizagem (Khan et al., 2010), sendo importante destacar a aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem semi-supervisionada e aprendizagem por reforço (Rebala et al., 2019). Algumas das principais subdivisões e aplicabilidades do Aprendizado de Máquina são apresentadas na Figura 2.4, criada a partir de informações disponíveis em Jha (2017), Hunt (2017), e Litslink (2019).

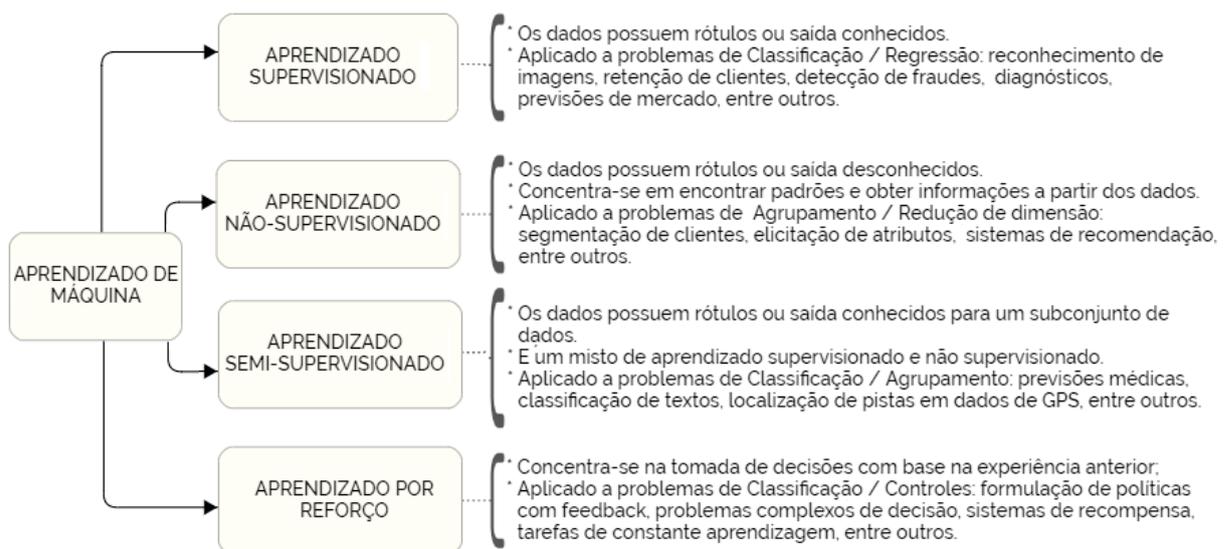


Figura 2.4: Principais subdivisões e aplicabilidades do Aprendizado de Máquina: aprendizagem supervisionada, aprendizagem não supervisionada, aprendizagem semi-supervisionada e aprendizagem por reforço.

Fonte: Adaptado de Jha (2017), Hunt (2017) e Litslink (2019).

Os algoritmos de aprendizado supervisionado utilizam a detecção de padrões para estabelecer previsões e podem ser subdivididos em algoritmos de classificação e algoritmos de regressão (Rebala et al., 2019). Os algoritmos de classificação (comumente empregados em soluções de VC) tem como objetivo classificar algo (itens ou amostras) em um conjunto distinto de classes ou categorias, de acordo com as características observadas pelo supervisor. Os algoritmos de regressão funcionam com a compreensão de relação da máquina, quanto às variáveis para prever valores de uma variável contínua.

Na aprendizagem não supervisionada não há rótulos, o algoritmo automaticamente agrupa os dados em algum formato para melhor entendimento da estrutura, ou faz a simplificação dos dados (identifica padrões para rotular os dados). Dada uma grande quantidade de dados, a máquina pode identificar tendências de similaridade, ou seja, o sistema aprende por conta própria (Jain e Kacprzyk, 2013).

No aprendizado semi-supervisionado, a máquina recebe um grande conjunto de dados, no qual apenas alguns pontos de dados são rotulados. O algoritmo usa técnicas de *clustering* (agrupamento) para identificar grupos dentro do conjunto de dados fornecido e usa os poucos pontos de dados rotulados dentro de cada grupo para fornecer rótulos para outros pontos de dados no mesmo *cluster* (grupo) (Rebala et al., 2019).

Na aprendizagem por reforço, o algoritmo toma uma decisão com base no conjunto de dados que é disposto, e sinais de recompensa em resposta aos resultados observados das tomadas de decisão (tentativa e erro). Com base nisso, o algoritmo ajusta sua estratégia para receber a recompensa mais alta. O processo está, portanto, relacionado ao uso da experiência adquirida ao interagir com o mundo e ao *feedback* avaliativo para melhorar a capacidade de um sistema tomar decisões comportamentais (Littman, 2015).

Recentemente a pesquisa na área de Aprendizado de Máquina foi impulsionada pelo impacto do paradigma de aprendizado profundo (*Deep Learning*), cuja arquitetura consiste em redes neurais artificiais composta de neurônios, funções de ativação e pesos que aprendem por conta própria usando algoritmos de aprendizado (Shah e Philip, 2019; LeCun et al., 2015). Ravì et al. (2016) apresentam diferentes abordagens de aprendizado profundo comumente usadas no campo da informática em saúde. O aprendizado profundo descobre uma estrutura complexa em grandes conjuntos de dados usando o algoritmo de retropropagação para indicar como uma máquina deve alterar seus parâmetros internos que são usados para calcular a representação em cada camada da representação na camada anterior (LeCun et al., 2015).

Após tornar-se popular, o aprendizado profundo passou a ser empregado em aplicações inovadoras de diversas áreas, como entendimento de imagem e vídeo, reconhecimento de fala, imagens médicas e veículos autônomos (Leo et al., 2018). A disponibilidade de poder computacional e grande quantidade de dados tornou essas grandes estruturas muito eficazes no aprendizado de recursos ocultos, juntamente com padrões de dados (Rebala et al., 2019).

Redes neurais convolucionais profundas (*Convolutional Neural Networks* — CNN) e seu sucesso em problemas de reconhecimento de padrões em imagem trouxeram avanços significativos para diversas áreas, inclusive para o desenvolvimento de recursos de TA (Leo et al., 2018). Contudo, existe um desafio relacionado a como ajustar CNN para um pequeno conjunto de dados, mantendo o desempenho semelhante ao do conjunto de dados em larga escala. A solução comum para o uso de CNN em pequenos conjuntos de dados é a Transferência de Aprendizado, que elimina a camada classificadora de uma CNN pré-treinada e a ajusta no conjunto de dados de destino (Han et al., 2018). A Transferência de Aprendizado é a melhoria do aprendizado em uma nova tarefa através da transferência de conhecimento de uma tarefa relacionada que já foi aprendida (Torrey e Shavlik, 2010). O objetivo é melhorar o aprendizado na tarefa de destino, aproveitando o conhecimento da tarefa de origem, o que amplia as possibilidades de aplicação da CNN, beneficiando também o desenvolvimento de recursos de TA.

2.5 CONSIDERAÇÕES

Considerando que a abordagem apresentada nesta tese visa direcionar o desenvolvimento de soluções de CAA que possam promover a autonomia de usuários com dificuldade motora e de fala por meio de interação gestual, está vinculado ao objetivo deste capítulo destacar o que foi considerado mais relevante dentre os conceitos sobre as diferentes áreas de conhecimento relacionadas. Conceitos sobre CAA foram apresentados, a fim de familiarizar o leitor com os termos utilizados, propósitos e princípios dessa área. A IHC e seu vínculo com o contexto de TA também foram apresentados, juntamente com alguns exemplos de técnicas de avaliação que podem ser úteis para avaliação da proposta. Por fim, a VC com possíveis aplicações e exemplos

de métodos utilizados para reconhecimento de objetos ou ações foram apresentados, seguida do Aprendizado de Máquina, destacando diferentes abordagens de aprendizagem. Essas áreas de pesquisa foram cruciais para o desenvolvimento da metodologia proposta e têm sido amplamente investigadas. Trabalhos relacionados desenvolvidos para auxiliar pessoas com dificuldades motora e de fala são apresentados no próximo capítulo.

3 ESTADO DA ARTE

Este capítulo apresenta os trabalhos mais relevantes relacionados ao tema desta pesquisa. Estudos voltados ao desenvolvimento ou descrição de sistemas computacionais de CAA são apresentados, com o objetivo de exemplificar as possibilidades atuais dos recursos de TA voltados para CAA. Também é exibido um panorama da literatura, incluindo um mapeamento sistemático contemplando trabalhos desenvolvidos para apoiar ou melhorar de alguma forma a qualidade de vida de pessoas com deficiência motora empregando técnicas de VC e dispositivos diversos.

3.1 ESTUDOS SOBRE CAA E VISÃO COMPUTACIONAL EM TECNOLOGIA ASSISTIVA

O campo da CAA abrange a pesquisa e desenvolvimento em educação, sistemas e práticas que possibilitam a participação de diversas áreas, e por isso, requer uma abordagem multidisciplinar (Pavliša et al., 2012). A partir de um mapeamento sistemático da literatura (Ascari et al., 2018a) conduzido com o objetivo de identificar oportunidades e desafios sobre a pesquisa aplicada em dispositivos móveis para CAA, percebeu-se a necessidade de promover o uso e adoção efetiva da TA.

Na literatura, diferentes iniciativas podem ser encontradas com o propósito de tornar sistemas de CAA efetivamente utilizáveis. Alguns estudos têm seu foco em métodos de *design* participativo, tais como os trabalhos de Borges et al. (De Faria Borges et al., 2011; de Faria Borges et al., 2012; Borges et al., 2013). Esses estudos estão voltados para a inclusão de *stakeholders* (partes interessadas no projeto) com deficiências em práticas de *design* participativo para conceber recursos de TA customizados. Como esses autores sugerem, atividades participativas são promissoras para projetar soluções para a prática de CAA.

Considerando a relevância das abordagens de CAA e o frequente uso de símbolos para promover a comunicação, alguns trabalhos exploraram as possibilidades relacionadas a integração da VC e Realidade Aumentada (RA) nas intervenções de comunicação. São exemplos: o sistema STAR (da Silva et al., 2014), que reforça os símbolos de comunicação para crianças com TEA sobrepondo objetos virtuais e animações sobre a tela do ambiente real; o ambiente AmCARA (Garbin et al., 2008; Garbin e Dainese, 2009) que permite interação de crianças com paralisia cerebral com *software* de comunicação assistiva utilizando recursos de RA por meio de teclado, placas de controles e marcadores; o projeto PictogramRoom (Spain e Spain, 2017) que envolve uma sala de RA para ensinar como interpretar os pictogramas (símbolos que representam um objeto ou conceito por meio de desenhos figurativos) que permitem a comunicação com pessoas afetadas por TEA.

Existem atualmente mais de uma centena de aplicações de CAA disponíveis no mercado, a um custo consideravelmente mais baixo quando comparado com sistemas de comunicação tradicionais. A popularidade relativa dessas tecnologias levou a uma proliferação de *software* concorrentes. Alguns sistemas descritos ou citados em publicações científicas, considerados com características desejáveis a sistemas de CAA, são apresentados na Tabela 3.1. Os trabalhos listados nesta tabela são bastante efetivos para o que se propõem, mas em geral demandam interação por meio de dispositivos tradicionais, como mouse, teclado, ou via toque.

Tabela 3.1: Exemplos de Sistemas de CAA considerados com objetivos semelhantes ou características desejáveis aos sistemas que podem ser desenvolvidos a partir da metodologia proposta.

Dynavox e TouchSpeak (Al Mahmud, 2012); Proloquo2Go, iComm e TapToTalk (Drigas e Kokkalia, 2016): Comunicação por meio de construção de palavras e frases e reprodução de áudio. Uso de símbolos (ou ícones). TouchSpeak é voltado para pessoas que sofrem com afasia (perda da capacidade de produzir ou compreender a linguagem falada), iComm para crianças com dificuldades de aprendizagem e Proloquo2Go, TapToTalk e Dynavox para crianças e adultos com dificuldades de comunicação.
Lingraphica (Al Mahmud, 2012): Comunicação explorando cenas familiares para praticar palavras e construir mensagens de comunicação funcionais. Uso de símbolos (ou ícones) e metáfora de storyboards (pranchas com recursos semelhantes a histórias em quadrinhos). Voltado para pessoas que sofrem com afasia.
Lets Talk (Torii et al., 2012); Grace (Drigas e Kokkalia, 2016): Comunicação por meio de construção de sentenças de imagens. Uso de símbolos (ou ícones). Voltado para crianças com TEA.
MyVoice (Jellinek e Abrahams, 2012): Comunicação por meio de construção de frases usando palavras e figuras. Possui interface móvel e web, e vocabulário de localização consciente que pode detectar que um usuário entrou em um restaurante ou biblioteca e apresentá-los com palavras relacionadas a esse local. Voltado para pessoas com déficits de memória e comunicação.
MyChoicePad (Jellinek e Abrahams, 2012): Comunicação por meio de construção de sentenças de imagens. Uso de sinais e símbolos do Makaton (programa de linguagem que usa sinais e símbolos para ajudar na comunicação). Voltado para adultos e crianças com dificuldades de aprendizagem ou de comunicação.
HeyTell e Tango (Smith, 2013): Auxílio no entendimento do contexto emocional de uma mensagem. Uso de gravação e reprodução de voz para retransmissão ao parceiro de comunicação. Voltado para alunos com dificuldades intelectuais.
Vox4all (Quintela et al., 2013): Comunicação por meio de placas de comunicação interativas. Uso opcional de sistema pictográfico <i>Widgit Symbols for Literacy</i> . Voltado para pessoas com dificuldades de comunicação.
SCALA (Passerino e Bez, 2013): Comunicação por meio de construção de palavras e frases, construção de pranchas e histórias, centrado no ambiente educacional e familiar do usuário, denominado de Design Centrado no Contexto (DCC). Voltado para pessoas com TEA.
iCommunicate (Drigas e Kokkalia, 2016): Comunicação por meio de storyboards. Uso de símbolos (ou ícones), fotos e metáfora de storyboards. Voltado para crianças com TEA.
iConverse (Drigas e Kokkalia, 2016): Comunicação por meio de construção de sentenças de imagens. Uso de PECS (<i>Picture Exchange Communication System</i>). Voltado para crianças com dificuldades de comunicação.
AutismExpress (Drigas e Kokkalia, 2016): Estímulo a comunicação e melhoria de habilidades sociais com base no entendimento e expressão de emoções. Uso de imagens representativas de emoções comuns na forma de desenhos animados. Voltado para crianças com TEA.
Stories2Learn (Drigas e Kokkalia, 2016): Comunicação por meio de construção de histórias sociais personalizadas. Uso de fotografias, texto e mensagens de áudio. Voltado para crianças com TEA.

Para as pessoas que são total ou parcialmente incapazes de mover ou controlar seus membros e não podem confiar na comunicação verbal em função de alguma deficiência ou dificuldade, é importante obter uma interface capaz de interpretar seus movimentos voluntários limitados para permitir a interação com sistemas computacionais de CAA. Técnicas não invasivas com base em VC permitem que diferentes métodos de interação sejam considerados para esses usuários. Uma história de estudos nos últimos 30 anos relatada em Bhuiyan e Picking (2011) sugere que as interfaces controladas por gestos agora oferecem oportunidades realistas e acessíveis, o que pode ser apropriado para pessoas com deficiência ou idosos.

Provavelmente, o exemplo mais antigo usando vídeo para reconhecer os movimentos das mãos como um modo de interação é o trabalho de Krueger et al. (1985). Jacob (1991) foi um dos primeiros a introduzir técnicas de interação baseadas em aparência em aplicativos em tempo real voltados para pessoas com deficiências. O autor discutiu alguns fatores e considerações técnicas que surgem na tentativa de usar os movimentos dos olhos como interface de entrada de dados para sistemas computacionais. Desde então vários trabalhos foram e continuam sendo desenvolvidos com o objetivo de melhorar a qualidade de vida das pessoas com deficiências físicas, apoiados em recursos de TA e modalidades de interação baseadas no reconhecimento de movimentos corporais.

A Tabela 3.2 apresenta um agrupamento de trabalhos voltados para o reconhecimento de gestos, aplicando dispositivos diversos e em sua maioria técnicas de VC para recursos de TA. Esses trabalhos estão categorizados de acordo com o tipo de dispositivo empregado e a parte do corpo humano utilizada para rastreamento na solução em questão. É apresentada também a informação do público-alvo de cada estudo. O termo "pessoas com algum tipo de deficiência física" foi empregado como uma tradução do termo em inglês "*disabled people*". Quanto às partes do corpo humano utilizadas nas pesquisas, os estudos identificados com o termo "Várias" referem-se ao uso de mais de uma parte do corpo como sinal visual ou rastreamento de movimentos corporais de um modo geral. No caso específico do estudo desenvolvido por Kane et al. (2012), a VC é aplicada para identificação de contexto e localização para fins de CAA, não está, portanto, relacionado à identificação ou reconhecimento de alguma parte do corpo do usuário.

Como pode-se visualizar na Tabela 3.2, dispositivos simples como *webcams* foram utilizados por um grande número de trabalhos (43% dos trabalhos apresentados), pois representam uma alternativa viável para detecção e rastreamento de movimentos, especialmente no que diz respeito ao baixo custo associado. Essa vantagem também é válida para o uso de câmeras de dispositivos móveis, a cada dia mais acessíveis. Trabalhos mais recentes têm investigado também o uso de dados de profundidade (7 trabalhos desde 2014 foram apresentados na Tabela 3.2), por meio de dispositivos como o Kinect, BumbleBee depth sensor, Monocular infrared depth camera e Image range sensor.

As pesquisas que se dedicaram à detecção/rastreamento de olhos receberam muita atenção da comunidade acadêmica (37% dos trabalhos apresentados na Tabela 3.2), possivelmente porque, em algumas situações, os movimentos oculares são os únicos movimentos que algumas pessoas com deficiência podem controlar voluntariamente. Dentre os dispositivos *Eye-tracker* (rastreadores oculares comerciais) empregados nos estudos relacionados, pode-se citar rastreadores da Tobii Technology, da EyeTech Digital Systems, EyeTribe, e RED da SensoMotoric Instruments.

O rastreamento de regiões específicas do corpo humano como forma de interação para um público-alvo específico, tende a gerar soluções mais adaptadas aos usuários de interesse. Contudo, a acessibilidade dessas soluções é prejudicada, pois usuários com capacidades ou limitações motoras diferentes do público especificado podem ficar impedidos de utilizá-las. Soluções mais flexíveis tendem a explorar diversas partes do corpo humano como forma de viabilizar a interação humano-computador por meio de reconhecimento de gestos e são consideradas mais próximas ao objetivo da proposta apresentada nesta tese.

Tabela 3.2: Exemplos de trabalhos que empregaram reconhecimento de gestos, aplicando dispositivos diversos e técnicas de Visão Computacional para o desenvolvimento de soluções como recursos de Tecnologia Assistiva.

Partes do corpo X Dispositivos empregados	Câmara de dispositivos móveis	Câmara de profundidade	Câmara simples ou webcam	Câmara térmica	Eye-tracker	Outros / Não informado	Mais de um dispositivo	Total
Boca / Língua								
Cabeça / Face	Montanini et al. (2015) ^f	Bian et al. (2016) ^f	Leung e Chau (2010) ⁿ ; Niu et al. (2018) ^d Toyama (1998) ^a ; Chauhan e Morris (2001) ^a ; Mauri et al. (2006) ⁿ ; Tu et al. (2007) ^f ; Varrona et al. (2008) ^a ; Yee et al. (2009) ^a ; Antunes et al. (2016) ^a ; Xavier et al. (2017) ^k ; Henzen e Nohama (2017) ^f	Memarian et al. (2009a) ^d ; Memarian et al. (2009b) ^a ;		Fu e Huang (2007) ^a ; Azmi et al. (2009) ^a	Alves et al. (2016) ^j Simpson et al. (2008) ^k ;	5 14
Nariz / Narinas			Manresa-Yee et al. (2010) ⁿ			El-Afifi et al. (2004) ^d		2
Mãos	Vidhya et al. (2016) ^g	Gomez-Donoso e Bakheet (2017) ^a	Vafadar e Behrad (2008) ^a ; Pirani e Kolte (2010) ^m ; Federici e Scherer (2012) ^h ; Dehankar et al. (2017a) ^a ; Dehankar et al. (2017b) ^a			Patel e Roy (1998) ^f ; Morrison e McKenna (2002a) ^d ; Morrison e McKenna (2002b) ^a		11
Olhos	Mohammed e Shereen (2014) ^a ; Zhang et al. (2017) ^j	Cazzato et al. (2014) ^a	Como et al. (2002) ^j ; Koceljko et al. (2009) ^a ; Missimer e Betke (2010) ^a ; Parmar et al. (2012) ^a ; Lemahieu e Wyns (2011) ^a ; Marnik (2014) ^c ; Zhang et al. (2016) ^a ; Liu et al. (2016) ^a ; Naqvi et al. (2017) ^a	Grane et al. (2011) ^p	Jacob (1991) ^a ; Hornof e Caverder (2005) ⁱ ; MacKenzie e Ashtiani (2009) ⁱ ; Ashtiani e MacKenzie (2010) ⁱ ; Biswas e Langdon (2011) ⁱ ; Eidam et al. (2016) ^p ; Blignaut (2017) ^q ; Lupu et al. (2017) ^e ; Hemmingsson et al. (2018) ^l	Grauman et al. (2003) ⁱ ; Adjouadi et al. (2004) ⁱ ; Fejtová et al. (2009) ^k	Moon et al. (2003) ⁱ ; Zander et al. (2010) ⁱ ; Biswas e Langdon (2015) ^a ; García et al. (2017) ⁱ ; Stawicki et al. (2017) ^j	30
Pés		Yoda et al. (2017) ^a						1
Várias		Martins et al. (2015a) ^a ; Negin et al. (2018) ^o	Betke et al. (2002) ^a ; Betke (2008) ^a ; Perini et al. (2006) ^f ; González et al. (2010) ⁿ ; Tuisku et al. (2013) ^a ; Epstein et al. (2014) ^k ; Kurauchi et al. (2015) ^a ; Cristina e Camilleri (2016) ^a			Roy et al. (1994b) ⁿ ; Roy et al. (1994a) ⁿ ; Hansen et al. (2004) ^a	Biswas e Langdon (2013) ^a ; Rozado et al. (2017) ^a	15
Nenhuma	Kane et al. (2012) ^b							1
Total	5	7	34	3	9	12	9	79

^a Público-alvo — Pessoas com:

^a algum tipo de deficiência física ; ^b afasia ; ^c atrofia muscular espinhal ; ^d deficiência de destreza ; ^e deficiência neuro-motora ; ^f deficiências motoras e de fala ; ^g dificuldades de audição e fala ; ^h dificuldades de fala ; ⁱ dificuldades motoras severas ; ^j Esclerose Lateral Amiotrófica ; ^k incapacidades motoras dos membros superiores ; ^l lesão alta da medula espinhal ; ^m lesão cerebral adquirida ; ⁿ paralisia cerebral ; ^o patologias corticais (Alzheimer) ; ^p Síndrome de Bloqueio Total ; ^q estágio avançado de esclerose múltipla ; ^r tetraplegia.

Os trabalhos de Roy et al. (1994a,b), apesar de não empregarem câmeras, mas sensores diretamente em contato com o usuário (rastreador magnético e eletromiografia), mostraram que muitos indivíduos não falantes são capazes de fazer gestos que são repetíveis e podem ser mapeados para palavras ou conceitos. Devido à natureza da deficiência física, esses gestos podem não seguir qualquer forma padronizada nem ser facilmente reconhecidos como representações icônicas. Os autores empregaram ANN de propagação reversa para reconhecimento da trajetória de braço e mostraram que pessoas com paralisia cerebral são capazes de realizar ações ou gestos com seus braços que são reconhecíveis por observadores familiares; eles descobriram que, incentivando a livre expressão, o número de gestos diferentes possíveis de serem realizados pelos indivíduos é muito maior do que se pensava e esse é um aspecto a ser explorado pela metodologia proposta nesta tese. O conhecimento dos observadores de uma pessoa com comprometimento motor, ou seja, dos cuidadores, sobre a capacidade do indivíduo na realização de movimentos, é fundamental para possibilitar a criação de uma linguagem gestual personalizada, que de acordo com a proposta desta tese, poderá ser utilizada para interação do usuário com um computador, e consequentemente, com um sistema computacional de CAA.

Betke et al. (2002) desenvolveram o CameraMouse, um sistema capaz de rastrear o movimento de diversas partes do corpo especificadas pelo usuário (como nariz, olhos, queixo, pé), por meio de um algoritmo com base em correspondência de modelos espaço-temporais. O movimento da parte do corpo rastreada é então mapeado no movimento do ponteiro do mouse na tela, sendo útil para interação baseada na seleção de itens dinâmicos ou estáticos apontando sem clicar. Esse trabalho foi bastante representativo para a área de TA, e inspirou outros trabalhos como Betke (2008), Kwan et al. (2011), Epstein et al. (2014), Marnik (2014), Antunes et al. (2016), Bian et al. (2016), que propuseram melhorias no sistema original ou aplicações para serem utilizadas em conjunto. Sistemas similares também foram desenvolvidos por Perini et al. (2006) e Fu e Huang (2007).

Em alguns trabalhos, o olhar é utilizado em conjunto com movimentos realizados por outras partes do corpo ou outros dispositivos, gerando modalidades de interação multimodal. Rozado et al. (2017) investigaram a interação com uma interface de usuário por meio de olhar como apontador para seleção de alvo e gestos faciais para execução de comandos de ação específicos do alvo. Com propósito semelhante, Biswas e Langdon (2013) desenvolveram interação com base em rastreamento ocular e técnicas de interação de varredura de chave única, combinando um sistema básico de movimentação de ponteiro na tela usando o olhar, um sistema de varredura emulador de mouse sensível a piscar de olhos e um sistema de varredura emulador de mouse sensível a um comutador físico. Esses autores utilizaram em conjunto o Kinect para reconhecimento de gestos e o Tobii X120 *Eye-Tracker* para rastreamento ocular.

Visando identificar maneiras de trabalhar com detecção e rastreamento de diferentes partes do corpo humano, sem demandar que o usuário especifique a área a ser rastreada, trabalhos voltados para o reconhecimento de ações ou comportamento humano também foram investigados.

Huang et al. (2011) empregaram um método de reconhecimento de ação humana usando o HOG (*Histogram of Oriented Gradients*) da imagem MHI (*Motion History Image*) e demonstraram resultados promissores que poderiam ser empregados para interação humano-computador. Vafadar e Behrad (2008) empregaram imagens MHI para reconhecimento de gestos de mão como forma de interação para pessoas com deficiências motoras e obtiveram resultados melhores em comparação a duas outras abordagens existentes. Esses autores utilizaram informações do histograma de orientação de movimento e HMM para classificar os gestos. Zhang et al. (2016) empregaram também a imagem MHI em uma proposta que combinou o uso de recursos obtidos via HOG e SURF para reconhecimento de ações. Como as ações humanas incluem informações de tempo e espaço, a imagem MHI reflete não apenas a posição de ação

espacial, mas também a ordem de movimento, podendo ser usada para descrever a tendência do movimento e descrição da informação temporal da ação (Zhang et al., 2016). Esses trabalhos serviram de inspiração para gerar um sistema de CAA como prova de conceito de implementação da metodologia proposta, descrito no Capítulo 4.

Diante de tantas possibilidades permitidas pelo uso da VC, viu-se oportuno realizar uma pesquisa das últimas tendências de pesquisa sobre IHC assistiva baseada em VC para auxiliar pessoas com deficiência motora. Soluções desenvolvidas exclusivamente para pessoas com comprometimento motor podem beneficiar também pessoas que possuem comprometimentos de fala associados, uma vez que a IHC possibilita o uso de modos alternativos de interação com sistemas computacionais, incluindo sistemas de CAA, tornando possível a comunicação. Assim, a seção seguinte apresenta uma breve descrição das principais técnicas utilizadas em trabalhos recentes que empregaram a VC para recursos de TA voltados para pessoas com dificuldades motoras.

3.2 MAPEAMENTO SISTEMÁTICO - TENDÊNCIAS EM APLICAÇÕES DE VC PARA PESSOAS COM DIFICULDADES MOTORAS

Um mapeamento sistemático permite identificar, analisar e interpretar toda a pesquisa disponível pertinente a questões de pesquisa específicas, áreas de pesquisa ou fenômeno de interesse (Kitchenham, 2004). Em uma revisão sistemática, o processo de busca é conduzido de acordo com uma sequência bem definida de etapas, seguindo um protocolo de estudo previamente planejado (Munzlinger et al., 2012). Os mapeamentos sistemáticos são um tipo particular de revisão sistemática com um escopo mais amplo, projetado para cobrir e fornecer uma visão geral de uma área de pesquisa, classificando e contando as contribuições de acordo com as categorias predefinidas (Petersen et al., 2015). Um mapeamento sistemático estuda a literatura para identificar quais tópicos estão sendo abordados e como, onde a pesquisa foi publicada e por quem, quais são as práticas e ferramentas mais comuns, as lacunas e oportunidades e assim por diante. Enquanto as revisões sistemáticas da literatura se concentram na coleta e síntese de evidências, considerando a força das evidências, o mapeamento sistemático da literatura é usado principalmente para estruturar uma área de pesquisa (Petersen et al., 2015).

As possibilidades de usar a VC como suporte à CAA são diversas, interessantes e desafiadoras em termos de dispositivos, interface, interação, questões sociais e econômicas. Assim, viu-se oportuno realizar um mapeamento sistemático da literatura para investigar as recentes iniciativas de pesquisa no campo da VC aplicada à IHC para auxiliar a interação e comunicação de pessoas com deficiência motora. Nesse mapeamento foram considerados artigos científicos indexados pelas bases de dados ACM¹, IEEE², Science Direct³ e Springer⁴, publicados entre janeiro de 2009 e dezembro de 2019. O intervalo de 10 anos foi utilizado com o objetivo de mapear as pesquisas mais recentes sobre o tema investigado.

O mapeamento sistemático foi desenvolvido de acordo com as diretrizes de Munzlinger et al. (2012) e Petersen et al. (2015). O primeiro passo foi planejar e formalizar o protocolo do estudo, especificando o problema de pesquisa, seu objetivo, questão geral e questões de pesquisa. A Tabela 3.4 apresenta detalhes sobre o protocolo de estudo empregado.

Para o protocolo do estudo, os critérios de seleção foram definidos e aplicados como um primeiro filtro para os estudos recuperados. Os critérios de seleção foram divididos em critérios

¹<https://dl.acm.org/>

²<https://ieeexplore.ieee.org>

³<https://www.sciencedirect.com/>

⁴<https://link.springer.com/>

de inclusão e exclusão e aplicados para classificar os estudos de acordo com seus metadados (título, resumo e palavras-chave). Os estudos que atenderam a pelo menos um dos critérios de inclusão foram incluídos e os estudos que atenderam a pelo menos um dos critérios de exclusão foram excluídos.

Tabela 3.4: Protocolo de Estudo utilizado para condução do mapeamento sistemático, especificando o problema de pesquisa, seu objetivo, questão geral e questões de pesquisa.

Problema de pesquisa: Investigar as iniciativas de pesquisa que empregaram técnicas de Visão Computacional (VC) em recursos de Tecnologia Assistiva voltados para viabilizar a interação e comunicação de pessoas com dificuldades motoras.

Objetivo: Mapear as pesquisas desenvolvidas recentemente na área de Tecnologia Assistiva para pessoas com dificuldades motoras empregando técnicas de VC.

Questão geral: Quais são as recentes soluções teóricas ou práticas desenvolvidas empregando VC para apoiar a comunicação e interação de pessoas com dificuldades motoras?

Questões de pesquisa:

- 1) Como a interação gestual baseada em VC tem sido explorada para auxiliar a comunicação e interação de pessoas com dificuldades motoras?
- 2) Quais tipos de câmera ou dispositivos têm sido utilizados?
- 3) Quais regiões corporais têm sido utilizadas para rastreamento?
- 4) Quais técnicas têm sido empregadas para reconhecimento de gestos?
- 5) As soluções permitem a personalização dos gestos utilizados?
- 6) Existem metodologias vinculadas ao desenvolvimento das soluções?
- 7) As soluções voltadas para viabilizar comunicação usam o termo Comunicação Aumentativa e Alternativa?
- 8) Pessoas com dificuldades motoras participam da construção ou avaliação das soluções propostas?

A Tabela 3.5 apresenta os critérios de inclusão e exclusão empregados para filtragem de estudos no primeiro filtro.

Tabela 3.5: Critérios de inclusão e exclusão empregados no primeiro filtro do mapeamento sistemático para seleção de estudos.

Primeiro Filtro	
Critérios de Inclusão	Critérios de Exclusão
CI1: O estudo define ou apresenta instrumentos para auxiliar a comunicação ou interação de pessoas com dificuldades motoras usando VC.	CE1: O estudo não apresenta instrumentos, investigações, comparações ou avaliações de soluções voltadas para auxiliar a comunicação ou interação de pessoas com dificuldades motoras usando VC.
CI2: O estudo investiga, compara ou avalia soluções para auxiliar a comunicação ou interação de pessoas com dificuldades motoras usando VC.	CE2: O estudo refere-se a um artigo resumido.
CI3: O estudo apresenta instrumentos voltados para Comunicação Aumentativa e Alternativa usando VC.	CE3: O estudo não foi publicado em inglês.
CI4: O estudo apresenta instrumentos voltados para comunicação não verbal usando VC.	CE4: O estudo não está disponível para acesso integral.

Os critérios de seleção para o segundo filtro (Tabela 3.6) foram definidos e aplicados na leitura completa dos estudos resultantes do primeiro filtro. Foram mantidos apenas os estudos que tratavam de temas relacionados a pelo menos um dos critérios de seleção. As buscas foram realizadas no mês de dezembro de 2019, retornando 221 estudos: 167 estudos foram excluídos pelo primeiro filtro e 21 foram excluídos pelo segundo filtro, resultando em um conjunto de 33 estudos.

Tabela 3.6: Critérios de seleção empregados no segundo filtro do mapeamento sistemático para seleção de estudos.

Segundo Filtro - Critérios de seleção
A: Ambientes inteligentes / Ambiente de moradia assistida.
B: Interação Humano-Computador (IHC) / soluções computacionais.
C: Interação gestual aplicada em atividades de lazer.
D: Interação Humano-Máquina (IHM).
E: Interação Humano-Robô (IHR).
F: Levantamento bibliográfico.

As expressões de busca para cada base de dados selecionada foram definidas, calibradas e adaptadas de acordo com as funcionalidades disponíveis. A Tabela 3.7 apresenta uma visão geral do processo de seleção.

Tabela 3.7: Expressões de busca empregadas e resultados obtidos em cada base de dados considerada no mapeamento sistemático: ACM, IEEE, Science Direct, e Springer. 33 estudos foram selecionados após a aplicação do primeiro e segundo filtros.

Base de dados e Expressão de busca	Número de estudos	Primeiro Filtro	Segundo Filtro
ACM: [[Publication Title: "assistive"] OR [Publication Title: "accessible"] OR [Publication Title: "motor disability"] OR [Publication Title: "motor impairment"] OR [Publication Title: "disabled"] OR [Publication Title: "cerebral palsy"]] AND [[Abstract: "computer vision"] OR [Abstract: "camera"] OR [Abstract: "gesture"]] AND [[Abstract: "communication"] OR [Abstract: "interaction"]]	53	29	15
IEEE: (((("Document Title":"assistive") OR ("Document Title":"accessible") OR ("Document Title":"motor disability") OR ("Document Title":"motor impairment") OR ("Document Title":"disabled") OR ("Document Title":"cerebral palsy")) AND (("Abstract":"computer vision") OR ("Abstract":"camera") OR ("Abstract":"gesture")) AND (("Abstract":"communication") OR ("Abstract":"interaction"))))	27	17	13
Science Direct: Title, abstract, keywords: ("computer vision" OR "camera" OR "gesture") AND ("communication" OR "interaction") Title: "assistive" OR "accessible" OR "motor disability" OR "motor impairment" OR "disabled" OR "cerebral palsy"	39	6	4
Springer: with at least one of the words: "computer vision" "camera" "gesture" where the title contains: "assistive" "accessible" "motor disability" "motor impairment" "disabled" "cerebral palsy"	2	2	1
Total de estudos selecionados	221	54	33

A Tabela 3.8 mostra o formulário de extração de dados usado para padronizar os dados extraídos das publicações lidas, com o objetivo de reduzir o viés dos resultados e a informalidade do processo. Esse formulário de extração de dados foi inspirado na taxonomia de gestos apresentada por Escalera et al. (2017) sobre os vários componentes envolvidos na

realização de pesquisas em reconhecimento de ações/gestos, mantendo apenas as informações de maior interesse para o domínio de pesquisa desta tese.

Tabela 3.8: Formulário de extração de dados usado para padronizar o registro das informações obtidas a partir dos trabalhos resultantes do mapeamento sistemático.

FORMULÁRIO DE EXTRAÇÃO DE DADOS	
Aspectos do significado do gesto	Informação espacial (onde ocorre); Informação direcional (o caminho seguido); Informação simbólica (o sinal que faz); Ação (o tipo de ação que está sendo executada); Pose do corpo (localização dos membros e ângulos articulares); Forma de mão (a postura da mão); Informação afetiva (sua qualidade emocional).
Partes do corpo utilizadas	Mãos, dedos das mãos, pés, braços, corpo, cabeça, rosto, boca, língua, olhos, outros.
Atores envolvidos no reconhecimento de gestos	Interação Humano Máquina/Robô/Computador; Gestos humanos isolados; Gestos Humano-Humano; Interação Humano-Objeto.
Gestos estáticos/dinâmicos	Estáticos (apenas considerando características de um quadro de entrada ou qualquer outro dispositivo de aquisição descrevendo a configuração espacial dos membros do corpo). Dinâmicos (a trajetória e pose de membros do corpo fornecem a mais alta informação discriminativa).
Fonte de Dados	Visual (RGB, <i>Depth</i> , Térmico). Dispositivos empregados.
Avaliação, métricas	Spotting (distingue gestos significativos de movimentos não intencionais); Classificação (sequências contendo gestos ou gestos segmentados); avaliação com o público-alvo.
Reconhecimento	Modelo com base em aparência; Pré-processamento; Extração de características; Segmentação do início-fim do gesto; Reconhecimento: classificação final do rótulo do gesto/ação.
Aplicações	Interação Humano-Computador, Interação Humano-Máquina, Interação Humano-Robô, Interação gestual aplicada em atividades de lazer, Ambientes inteligentes, Levantamento bibliográfico.

A extração foi realizada rastreando as informações no formulário de extração para as declarações de cada artigo e verificando sua exatidão. As categorias foram criadas dinamicamente à medida que os dados foram extraídos para refletir o conjunto de dados resultante do próprio processo de extração. Um esquema dinâmico foi adotado em vez de um esquema predefinido para refletir o conteúdo extraído. Por um lado, uma categorização predefinida pode representar uma categorização mais estruturada; por outro lado, impede a identificação de outras categorias relevantes para representar o conjunto selecionado. Como a natureza deste estudo é um mapeamento aberto e abrangente, optou-se por essa forma mais flexível de classificação, gerada a partir da forma de extração de dados utilizada.

3.2.1 Ameaças à validade do estudo

Esta seção discute as ameaças à validade que podem ter afetado os resultados desse mapeamento sistemático. O protocolo de revisão foi validado para garantir que a pesquisa fosse a mais correta, completa e objetiva possível. No entanto, foram identificadas possíveis limitações em dois momentos do processo de revisão: na seleção da publicação e na extração dos dados.

Diferentes ameaças à validade podem ser apontadas neste estudo. Uma das ameaças corresponde à falta de algum estudo relevante na área. De fato, não é possível garantir que todos os trabalhos relacionados publicados estejam incluídos neste mapeamento. A busca de publicações foi realizada apenas em um conjunto limitado de bases de dados, supondo que esses mecanismos de pesquisa tendem a conter a maioria dos estudos relevantes. Os termos usados nas expressões de busca podem ter muitos sinônimos e a própria expressão de busca restringe os resultados possíveis. É possível que documentos relevantes não tenham sido incluídos (como

alguns citados anteriormente neste mesmo capítulo) devido à divergência no uso de termos e expressões, como observado no caso do termo "*nonverbal communication*". Como outra ameaça, é possível que algum tipo de imprecisão ou classificação incorreta tenha ocorrido na extração de dados realizada neste mapeamento sistemático, principalmente porque a extração de dados foi feita individualmente por um pesquisador. A coerência do esquema de classificação também pode introduzir viés na análise dos dados, e outros pesquisadores podem propor diferentes esquemas de classificação. Para reduzir essas ameaças, a extração e classificação dos dados foram conduzidas pelo autor desta tese e validadas por um segundo pesquisador.

Petersen et al. (2015) revisaram os esquemas de classificação de validade existentes e discutiram sua aplicabilidade à engenharia de *software*. Com base no estudo desses autores, os seguintes tipos de validade devem ser levados em consideração para minimizar as ameaças à validade do estudo: validade descritiva, validade teórica, generalização, e validade interpretativa.

Validade descritiva é a extensão em que as observações são descritas com precisão e objetividade. Para reduzir essa ameaça, um formulário de extração de dados foi projetado para suportar a gravação de dados. O formulário direcionou o processo de extração de dados de forma objetiva e permite que as informações extraídas possam sempre ser revisitadas. Portanto, essa ameaça é considerada sob controle.

A validade teórica é determinada pela capacidade de capturar o que se pretende capturar. Para reduzir essa ameaça, o conjunto de questões de pesquisa foi avaliado pelo autor e, posteriormente, por outro pesquisador. O primeiro e o segundo filtros foram aplicados e os artigos selecionados foram lidos na íntegra, possibilitando a extração de respostas a essas questões.

Generalização refere-se ao quanto é possível fazer uma pesquisa e generalizar os resultados provenientes do processo de pesquisa proposto. Petersen et al. (2015) apresentaram uma distinção entre generalização externa (generalização entre grupos ou organizações) e generalização interna (generalização dentro de um grupo). Para evitar a ameaça de generalização externa, foi utilizado um protocolo próprio e teoricamente defendido (Petersen et al., 2015), e para evitar generalização interna foram definidas questões de pesquisa e critérios de inclusão/exclusão próprios que permitem a expansão ou reprodução da pesquisa em um período diferente, garantindo a generalização do estudo.

A validade interpretativa é alcançada quando as conclusões tiradas são razoáveis, a partir dos dados e, portanto, são mapeados para a validade da conclusão. Uma ameaça na interpretação dos dados é o viés do pesquisador, que é minimizado por meio do processo de revisão realizado por outro pesquisador.

A repetibilidade requer relatórios detalhados do processo de pesquisa. Nesta seção o processo de mapeamento sistemático seguido é relatado e também as ações tomadas para reduzir possíveis ameaças à validade.

3.2.2 Resultados

A partir dos dados extraídos dos trabalhos selecionados no mapeamento sistemático, foi possível identificar alguns métodos e abordagens empregadas, bem como algumas características dos estudos que aplicaram a VC como recurso de TA para pessoas com dificuldades motoras. As informações consideradas relevantes são apresentadas na sequência e podem servir de referência para o desenvolvimento de sistemas de CAA com base na metodologia proposta nesta tese de doutorado.

A Figura 3.1 mostra o número de artigos publicados em cada ano, dentre os estudos selecionados no mapeamento sistemático. O maior número de trabalhos (oito) foi publicado no ano de 2015. Desses, cinco estudos foram publicados em conferências distintas, e três

estudos (Graham-Knight e Tzanetakis (2015), Kurauchi et al. (2015) e Saleh e Berns (2015)) foram publicados em uma mesma conferência: *8th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (PETRA). Outros cinco estudos selecionados no mapeamento sistemático (Missimer e Betke (2010), Paquette et al. (2011), McMurrrough et al. (2012), Feng et al. (2014) e Ghanem et al. (2017)), também foram publicados na conferência PETRA em outros anos. Essa conferência é considerada altamente interdisciplinar, concentrando-se em abordagens computacionais e de engenharia para melhorar a qualidade de vida e aprimorar o desempenho humano em uma ampla variedade de configurações, no local de trabalho, em casa, em espaços públicos, ambientes urbanos e outros.



Figura 3.1: Número de artigos publicados por ano, dentre os estudos selecionados no mapeamento sistemático.

A realização de movimentos por pessoas com deficiência é muito particular, e diferentes movimentos remanescentes podem ser explorados para interação. Pesquisas relacionadas à detecção e rastreamento de uma ou várias partes do corpo humano foram identificadas no mapeamento, e a ocorrência verificada é apresentada na Tabela 3.9. Nessa tabela os estudos selecionados são categorizados de acordo com sua aplicação e identificados quanto à participação do público-alvo (pessoas com dificuldade motora) no processo de construção ou avaliação da solução em questão.

Tabela 3.9: Estudos selecionados categorizados por aplicação, região corporal utilizada para interação gestual e participação de pessoas com dificuldades motoras em alguma etapa do desenvolvimento da pesquisa.

Aplicação	Região corporal / interação gestual	Estudo selecionado	Usuários com dificuldade motora
IHC	Cabeça	Paquette et al. (2011)	X
		Zhang et al. (2013)	
	Cabeça e Olhos	Kumar et al. (2010)	
		Zhao et al. (2014)	
		Kurauchi et al. (2015)	
		Saleh e Berns (2015)	
	Expressões corporais	Feng et al. (2014)	
		Rosales et al. (2017)	X
		Ascari et al. (2019)	X
	Expressões Faciais	Xu et al. (2010)	
	Mãos	Chattoraj et al. (2017)	
		Kakkoth e Gharge (2018)	
		Krishnamurthi et al. (2019)	
	Olhos	Missimer e Betke (2010)	
		McMurrough et al. (2012)	
		Parmar et al. (2012)	
		Karamchandani et al. (2015)	
		Sambrekar e Ramdasi (2015)	
		Drawdy e Yanik (2015)	
		Cristina e Camilleri (2016)	X
Utaminingrum et al. (2016)			
Pal et al. (2017b)		X	
Nakazawa et al. (2018)			
Não se aplica	Sharma et al. (2018)	X	
IHM	Expressões Faciais	Fine e Tsotsos (2009)	
	Mãos	Lamb e Madhe (2016)	
		Gao et al. (2017)	
Olhos	Mohammad e Anas (2015)		
IHR	Cabeça	Zhang et al. (2013)	
	Cabeça e Olhos	Zhao et al. (2014)	
		Saleh e Berns (2015)	
Olhos	Drawdy e Yanik (2015)		
Lazer	Cabeça, Mãos e Pés	Diment e Hobbs (2014)	X
	Mãos	Graham-Knight e Tzanetakis (2015)	X
Levantamento Bibliográfico	Não se aplica	Martins et al. (2015b)	-
		Ghanem et al. (2017)	-
		Tavares et al. (2018)	-

As palavras-chave indicadas em cada um dos estudos selecionados foram utilizadas para criar uma imagem representativa dos termos com maior ocorrência na forma de uma *TagCloud* apresentada na Figura 3.2. *TagClouds* são exibições visuais de um conjunto de palavras selecionadas por alguma justificativa, nas quais atributos do texto como tamanho ou cor são usados para representar propriedades relevantes, por exemplo a frequência dos termos associados (Schrammel et al., 2009). Pode-se visualizar na Figura 3.2 que entre os termos mais representativos estão "*interaction*", "*assistive*", "*detection*", "*recognition*", "*gesture*", "*human-computer*", "*technology*" que são bastante relacionados aos objetivos do mapeamento sistemático realizado. O termo "*nonverbal communication*" foi encontrado em alguns artigos e pode ser uma opção interessante a ser incluída nas palavras-chaves de expressões de busca em novos mapeamentos que venham a ser realizados.

Feng et al., 2014; Kurauchi et al., 2015; Pal et al., 2017b) dos trabalhos selecionados indicaram ter utilizado o sistema CameraMouse em suas pesquisas, seja para propor melhorias ou para comparação de resultados. Apenas 24% (8 de 33) do total de trabalhos selecionados incluíram pessoas com dificuldades motoras no processo de construção ou avaliação da solução em questão.

A Figura 3.3 apresenta as regiões corporais empregadas nos estudos selecionados (foi possível identificar essa informação em 29 de 33 estudos) para viabilizar interação gestual com soluções computacionais, máquinas ou robôs. Como já observado na Tabela 3.2 apresentada no início desse capítulo, o uso dos olhos para interação continua sendo uma alternativa bastante explorada (37% — 11 de 29), seja de forma única, ou associada com outras regiões corporais. Estudos voltados para a detecção e reconhecimento de gestos de mão também representam uma parcela significativa dos trabalhos selecionados (20% — 6 de 29). Algumas dificuldades ocorrem ao utilizar métodos de VC com imagens de câmeras de baixo custo para detecção e rastreamento do olhar, devido à baixa resolução e frequência de captura desses dispositivos. Possivelmente em função disso, muitos estudos voltados ao desenvolvimento de soluções de baixo custo, exploram o rastreamento e detecção de outras regiões corporais.

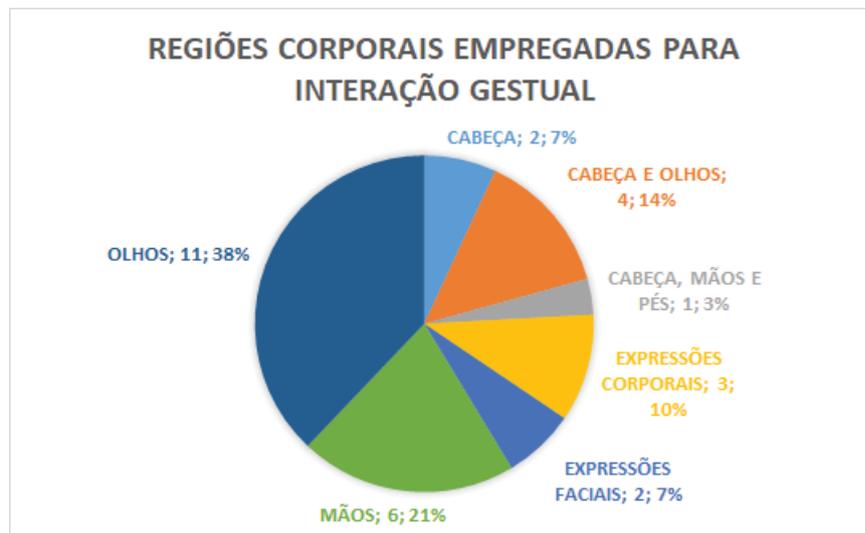


Figura 3.3: Partes do corpo humano utilizadas em pesquisas selecionadas no mapeamento sistemático que empregaram técnicas de Visão Computacional em recursos de Tecnologia Assistiva para pessoas com dificuldades motoras.

A Tabela 3.11 apresenta informações relacionadas a dispositivos e fonte de dados visuais utilizados (dados RGB e de profundidade — *Depth*). As publicações relacionadas ao símbolo “-” não continham informações sobre o dispositivo utilizado ou a natureza do trabalho não se aplica a essa identificação. É possível perceber uma diversidade grande de dispositivos empregados, e dentre esses, câmera simples ou *webcam* foi a opção utilizada pela maioria dos trabalhos (cerca de 48% — 16 de 33), sendo 6 desses trabalhos voltados para detecção e rastreamento de olhar.

Os principais objetivos ou aplicações dos trabalhos selecionados não são os mesmos e como resultado os métodos de VC utilizados, principalmente nas etapas de pré-processamento, segmentação e extração de características são muito variados, dificultando uma apresentação padronizada. Em função disso, a Tabela 3.12 destaca os principais métodos utilizados para realização das etapas de pré-processamento e segmentação de objetos, extração de características e representação, citados nos trabalhos selecionados para gerar as informações necessárias às etapas seguintes no processo de reconhecimento.

Estudos de Fine e Tsotsos (2009), Gao e Duan (2011), Karamchandani et al. (2015), Cristina e Camilleri (2016), Rosales et al. (2017), Kakkoth e Gharge (2018), Ascari et al. (2019)

Tabela 3.11: Dispositivos e fonte de dados usados em estudos selecionados via mapeamento sistemático.

Dispositivo	Fonte de dados		Estudo
	RGB	Depth	
Câmera infravermelha	X		Zhao et al. (2014)
Câmera Stereo Wii mote e Câmera Playstation Eye	X		McMurrough et al. (2012)
Câmera USB com led in- fravermelho	X		Nakazawa et al. (2018)
Digital câmera	X		Chattoraj et al. (2017)
IP câmera	X		Rosales et al. (2017)
Microsoft Kinect	X	X	Diment e Hobbs (2014), Graham-Knight e Tzanetakis (2015), Saleh e Berns (2015), Gao et al. (2017), Krishnamurthi et al. (2019)
Óculos de visão	X		Zhang et al. (2013)
Tobii eye tracker	X		Drawdy e Yanik (2015), Pal et al. (2017b)
Webcam	X		Fine e Tsotsos (2009), Kumar et al. (2010), Missimer e Betke (2010), Xu et al. (2010), Paquette et al. (2011), Parmar et al. (2012), Feng et al. (2014), Karamchandani et al. (2015), Kurauchi et al. (2015), Mohammad e Anas (2015), Sambrekar e Ramdasi (2015), Cristina e Camilleri (2016), Lamb e Madhe (2016), Utaminingrum et al. (2016), Kakkoth e Gharge (2018), Ascari et al. (2019)
Não se aplica	-	-	Martins et al. (2015b), Ghanem et al. (2017), Sharma et al. (2018), Tavares et al. (2018)

Tabela 3.12: Métodos utilizados em trabalhos selecionados no mapeamento sistemático para realização das etapas de pré-processamento e segmentação de objetos, extração de características e representação.

Pré-processamento e segmentação de objetos
Subtração de fundo, Morfologia, Remoção de ruídos, Threshold, Etiquetagem de regiões, Detecção de bordas, Suavização de imagem, conversão de espaço de cores.
Extração de características e representação
AAM, HOG, SIFT, BTE, coordenadas de Landmarks, extração de textura, extração de formas, extração de contornos, extração de trajetória, informação de postura, cor de pele, centróide de movimento, Fourier, Haar-like, Gaussian function, Padrão Direção-Magnitude, Transformada Hough, Transformada Wavelet.

citam o uso de funções disponíveis na biblioteca OpenCV (*Open Source Computer Vision*). Para cálculo de fluxo óptico os estudos de Kumar et al. (2010), Missimer e Betke (2010) e Ascari et al. (2019) citam o uso do fluxo óptico de Lucas-Kanade.

As técnicas utilizadas para realizar a etapa de reconhecimento de gestos, expressões ou movimentos nos trabalhos selecionados são apresentadas na Tabela 3.13. A grande variedade de métodos e abordagens indica a necessidade de se avaliar o contexto de cada aplicação para escolher a melhor maneira de implementar essa etapa.

Dentre as linguagens de programação citadas nos trabalhos que mencionaram essa informação, pode-se destacar: C++ (Fine e Tsotsos, 2009; Diment e Hobbs, 2014; Ascari et al., 2019); C# (Graham-Knight e Tzanetakis, 2015); C#.Net (Pal et al., 2017b); Python (Rosales et al., 2017; Ascari et al., 2019); e Matlab (Karamchandani et al., 2015; Mohammad e Anas, 2015; Sambrekar e Ramdasi, 2015; Lamb e Madhe, 2016; Chattoraj et al., 2017).

Tabela 3.13: Técnicas usadas para realizar a etapa de reconhecimento de gestos ou movimentos nos trabalhos selecionados no mapeamento sistemático.

Técnica para reconhecimento	Estudo
Adaboost	Kumar et al. (2010), Xu et al. (2010), Zhao et al. (2014), Sambrekar e Ramdasi (2015), Rosales et al. (2017)
ANN	Drawdy e Yanik (2015)
Bayes	Cristina e Camilleri (2016)
CNN	Ascari et al. (2019) .
Haar Cascade	Saleh e Berns (2015), Utaminingrum et al. (2016), Kakkoth e Gharge (2018)
KNN	Fine e Tsotsos (2009), Ghanem et al. (2017)
, Ascari et al. (2019)rshortsvm	Parmar et al. (2012), Zhang et al. (2013), Saleh e Berns (2015), Ghanem et al. (2017), Ascari et al. (2019)
VT	Zhang et al. (2013)
Template matching	Missimer e Betke (2010), Karamchandani et al. (2015), Cristina e Camilleri (2016), Ghanem et al. (2017)

Uma breve descrição das contribuições de cada trabalho selecionado no mapeamento sistemático é apresentada nas subseções a seguir.

3.2.2.1 *Estudos referentes a levantamentos bibliográficos*

Publicações referentes a levantamentos bibliográficos são uma maneira importante de compartilhar conhecimento entre pesquisadores interessados em contribuir para o crescimento de um determinado campo. Dentre os trabalhos selecionados, três referem-se a levantamentos bibliográficos em que o emprego de VC para pessoas com dificuldade motora é citado.

Martins et al. (2015b) apresentaram um estudo sobre possíveis soluções tecnológicas para aprimorar o processo de comunicação para surdos em plataformas de e-learning por meio da tradução da linguagem de sinais. Dentre as opções tecnológicas apresentadas para o reconhecimento, tradução e apresentação de sinais está a abordagem baseada em VC. Segundo esses autores, a captura de todo o corpo por uma câmera, pode ser uma solução para a tradução correta da linguagem gestual. Contudo, apesar dos dispositivos existentes poderem capturar facilmente gestos e expressões, enfrentam alguns problemas no reconhecimento de gestos, como: quantidade muito grande de gestos e similaridade entre eles; línguas de sinais diferentes devido à cultura, à vida social individual e à maneira como os gestos foram ensinados; e, a sequência de gestos para expressar uma sentença pode ser difícil de calcular, porque é difícil detectar onde o gesto começa e termina e onde começa o próximo. Assim, os autores acreditam ainda que existem alguns desafios críticos para resolver e ainda falta uma integração efetiva dessas tecnologias em plataformas de e-learning. Além disso, ainda não há soluções imediatas para resolver a comunicação síncrona em tempo real entre surdos e não surdos.

Ghanem et al. (2017) apresentaram um levantamento bibliográfico das técnicas mais recentes em sistemas de reconhecimento de linguagem de sinais baseados em dispositivos móveis. Os autores classificam as soluções existentes em sensores e visão e tem como foco principal algoritmos de detecção e classificação de sinais. A abordagem baseada em visão refere-se ao uso da câmera do telefone para capturar a imagem ou o vídeo da mão executando sinais. O artigo mostra uma comparação entre os métodos e técnicas empregados em diversos trabalhos baseados em VC, voltados principalmente para o reconhecimento de sinais estáticos.

Tavares et al. (2018) realizaram uma revisão bibliográfica sistemática com o objetivo de elucidar recursos de TA que permitam a inclusão digital de pessoas com paralisia cerebral e levantar métodos para sua avaliação, considerando possibilidades motoras, realização de tarefas e satisfação dos usuários. Dentre os trabalhos selecionados, pode-se destacar os estudos que empregaram VC como Manresa-Yee et al. (2014) que investigaram o uso de movimentos de cabeça para substituição do mouse. Esses autores citam soluções existentes e apresentam uma série de fatores de *design* relevantes para seu uso, propondo uma lista de parâmetros de concepção e avaliação para futuros *designers* usarem. Outro trabalho resultante da revisão é de Pauly e Sankar (2016) que investigaram métodos não intrusivos para avaliar dispositivos de detecção de piscar de olhos como forma alternativa de interação, tendo o descritor HOG combinado ao classificador SVM, mostrando melhor desempenho em comparação com outros métodos.

Os estudos citados dão uma amostra da pesquisa relacionada ao emprego da VC em abordagens inclusivas para pessoas com dificuldade motora. Esses estudos fornecem uma visão geral útil sobre o domínio pesquisado de forma condensada, e também representam fontes de referências adicionais.

3.2.2.2 *Interação gestual aplicada em atividades de lazer*

O *Kinect Virtual Art Program* (KVAP) usa a tecnologia de reconhecimento de gestos do Microsoft Kinect para permitir a criação de arte rastreando diferentes regiões corporais do usuário. Diferentes efeitos, como formas e brilhos, são ativados por diferentes gestos e velocidades dos membros. Diment e Hobbs (2014) apresentaram um estudo piloto realizado com cinco crianças com deficiência motora grave para determinar o nível de envolvimento físico que essas crianças poderiam atingir ao usar o KVAP em cinco sessões. Os resultados indicaram que os participantes gostaram do uso do KVAP e se envolveram cada vez mais com ele ao longo das sessões. O KVAP incentivou a atividade física e permitiu que as crianças criassem suas próprias obras de arte, uma atividade que antes era inacessível para elas usando abordagens tradicionais.

Graham-Knight e Tzanetakis (2015) apresentaram uma nova abordagem para pessoas com deficiência física tocarem música por meio de uma interface musical adaptável usando o Microsoft Kinect. Um sistema foi desenvolvido, em que os dados posicionais do Kinect são enviados para a linguagem de programação visual Max/MSP por meio do protocolo *Open Sound Control* (OSC) para análise e reprodução. Um teste foi realizado com o objetivo de ensinar usuários (com e sem deficiência física) a tocar um instrumento sem toque, o que não é intuitivo, e também medir a latência geral do tempo de reação e do sistema. O teste envolvia tocar um som de bateria, clicando com o mouse em uma mensagem de 'estrondo', levando o usuário a dedilhar um acorde de guitarra por meio de um gesto executado com a mão. O estudo fornece uma linha de base para futuras melhorias, medindo a latência do sistema Kinect atual, bem como o tempo de treinamento e a repetibilidade. Segundo os autores, a quantidade de latência foi considerada muito alta para a apresentação de um concerto, mas o sistema se mostrou agradável para um participante em uma situação de musicoterapia.

Com base nos estudos apresentados nesta seção é possível observar que a interpretação da interação corporal comumente explorada em jogos de entretenimento, também tem sua aplicação em outras atividades de lazer, como no domínio da arte e da música. As soluções apresentadas ampliam as possibilidades de interação para pessoas com dificuldades motora, incentivando o aprendizado de novas habilidades.

3.2.2.3 Interação Humano-Computador

Foram considerados trabalhos voltados para IHC os estudos que apresentaram soluções computacionais ou dispositivos que viabilizassem a interação de pessoas com dificuldades motoras com algum sistema computacional.

Kumar et al. (2010) criaram um sistema óptico de controle de mouse baseado no reconhecimento de movimentos de cabeça e piscar de olhos com o objetivo de gerar tecnologia de IHC assistencial para pessoas com dificuldades motoras.

Missimer e Betke (2010) propuseram um algoritmo que permite ao usuário interagir com o computador usando piscar de olhos para simular o clique em um mouse tradicional. O algoritmo é capaz de localizar automaticamente os olhos do usuário e aprender a aparência dos olhos abertos e fechados do usuário. O sistema apresentado amplia a funcionalidade dos sistemas de computador binário baseados em câmera de Grauman et al. (2003) e Chau e Betke (2005) fornecendo um método mais intuitivo para controlar o mouse. Ao interpretar o movimento de três regiões faciais, incluindo os dois olhos, o sistema permite que o usuário controle o ponteiro do mouse em um nível semelhante ao do mouse tradicional.

Xu et al. (2010) desenvolveram um sistema para pessoas com dificuldades de movimento dos membros superiores (ou amputação). O sistema emprega detecção e rastreamento de movimentos faciais para controlar o movimento do cursor na tela e acionar os eventos apropriados do mouse. Será aplicado em um sistema de reabilitação para pessoas com deficiência com amputação de membro superior.

Paquette et al. (2011) desenvolveram o Menu Controller, uma ferramenta que pode coletar as entradas de menu de aplicativos existentes e apresentá-las ao usuário de uma forma mais acessível e utilizável, para indivíduos com controle muscular severamente limitado. O sistema foi testado com um sistema de substituição de mouse baseado em câmera — CameraMouse (Betke et al., 2002). A maneira pela qual o Menu Controller renderiza novamente as entradas de menu de um aplicativo é baseada na interface do usuário de uma nova versão do CameraCanvas (Kwan e Betke, 2011), um programa de edição de imagens para pessoas com deficiências motoras graves, projetadas para uso com o CameraMouse.

McMurrough et al. (2012) apresentaram uma solução de baixo custo para rastreamento em tempo real da posição da cabeça de um usuário humano em relação a uma fonte de exibição de vídeo para estimativa do olhar em um ambiente assistencial. A solução utiliza um fone de ouvido vestível equipado com sensores encontrados em dispositivos de videogame. Um par de sensores de imagem Nintendo Wiimote é usado para criar uma câmera estéreo para rastreamento de posição 6DOF do fone de ouvido, enquanto uma câmera monocular Playstation Eye modificada é usada para rastrear a posição da pupila. O *hardware* de rastreamento resultante é capaz de medir a posição 3D de quatro LED infravermelhos montados em locais conhecidos na exibição de vídeo usando a triangulação dos dados da câmera estéreo. Também é descrita a integração desta abordagem de rastreamento posição de cabeça com uma solução de rastreamento de pupila baseada em VC, a fim de calcular o ponto de vista do usuário.

Parmar et al. (2012) apresentaram um sistema projetado para usuários com deficiência motora que são capazes de piscar voluntariamente. O algoritmo proposto permite detectar piscadas de olhos voluntárias, a duração do piscar e interpretar sequências de piscadas em tempo real para controlar uma interface humano-computador não invasiva. A técnica descrita usa câmeras que permitem rastrear características do nariz, as sobrancelhas e a posição da cabeça nas coordenadas 2D e 3D.

Feng et al. (2014) avaliaram o desempenho do método de seleção de cruzamento reverso do alvo, para uso com um sistema de substituição de mouse baseado em câmera para pessoas com dificuldades motoras — CameraMouse (Betke et al., 2002). O método avaliado apresentou

desempenho satisfatório. Tanto a análise estatística quanto as medidas subjetivas mostraram que o cruzamento reverso foi mais eficiente do que o clique regular no tempo de permanência.

Karamchandani et al. (2015) desenvolveram um sistema de rastreamento ocular simples, portátil, não invasivo e de baixo custo, destinado ao uso por crianças com deficiências graves. O sistema é capaz de diferenciar entre diferentes pontos de vista, permitindo ao usuário controlar uma interface em uma grade 4x4. Dois jogos baseados em grade foram desenvolvidos para testar o desempenho do sistema de rastreamento ocular como uma prova de conceito.

Kurauchi et al. (2015) apresentaram a técnica *Head Movement And Gaze Input Cascaded* (HMAGIC) que combina interação baseada no movimento da cabeça e no olhar. HMAGIC foi incorporado ao sistema CameraMouse (Betke et al., 2002) e posiciona o ponteiro do mouse para o ponto de olhar do usuário na tela e, em seguida, permite ao usuário controlar o ponteiro através do movimento da cabeça para manipulação fina. O HMAGIC combina as vantagens da velocidade dos movimentos dos olhos com a precisão do posicionamento baseado no movimento da cabeça em uma única interface de substituição de mouse.

Sambrekar e Ramdasi (2015) apresentaram um sistema desenvolvido para pessoas com deficiência de membros interagirem com computadores usando a direção do olhar. O sistema apresenta um teclado de 4 teclas que pode ser pressionado por meio da estimativa da direção do olhar do usuário.

Cristina e Camilleri (2016) apresentaram um método para estimativa tridimensional do olhar sob movimento livre da cabeça usando uma única câmera integrada de um notebook, como uma ferramenta de CAA para auxiliar indivíduos com deficiências motoras, como paralisia cerebral, afetadas por movimentos involuntários da cabeça e do rosto. A validade do método foi investigada em um conjunto de dados publicamente disponível e em dados da vida real capturados por meio da colaboração voluntária de um grupo de indivíduos normais e uma pessoa que sofre de paralisia cerebral.

Utamingrum et al. (2016) apresentaram uma nova estrutura para detectar o movimento dos olhos e sinalizar um comando de navegação. Os autores focaram na detecção de esclera, em vez de pupila e íris, comumente utilizadas para rastrear o movimento dos olhos. Esse estudo faz parte de uma pesquisa contínua e os autores apontam como trabalhos futuros que essa solução será incorporada em um mini assento elétrico, com o objetivo de fornecer mais acessibilidade para as pessoas com deficiência controlar os meios de subsistência.

Chattoraj et al. (2017) propuseram um método para reconhecer os diferentes gestos de mão usados pelas pessoas surdas para se comunicar usando escala invariável em recursos. Segundo esses autores, por meio do método proposto a linguagem de sinais americana é reconhecida através de imagens de sinais que ajudarão na comunicação com pessoas que não possuem o conhecimento dela.

Pal et al. (2017b) desenvolveram um aplicativo de CAA integrado ao rastreador ocular Tobii EyeX e ao sistema CameraMouse (Betke et al., 2002) para simplificar a comunicação para pessoas com deficiência. O sistema desenvolvido permite ao paciente expressar suas necessidades diárias, enviar mensagens pela Internet e a números de contato, conversar por meio de sintetizador de voz, gerar um sinal de alarme, além de utilizar outros módulos básicos (calculadora e lembrete) fornecidos.

Rosales et al. (2017) (trabalho já citado no início deste capítulo) apresentaram um protótipo desenvolvido para a interação entre crianças e computador que permitiu responder afirmativamente à seguinte pergunta de pesquisa: "Podemos melhorar a comunicação com crianças que tem paralisia cerebral utilizando um protótipo de VC para detecção de padrões corporais?" Por meio das observações feitas, foi possível distinguir certas expressões que um paciente utiliza para tentar se comunicar e que têm maior probabilidade de serem reconhecidas

pelo uso da VC. A interface desenvolvida detecta e rastreia movimentos da mão, pé ou cabeça do usuário, e foi utilizada para detectar padrões corporais referentes a dor de cabeça, felicidade, fome, medo e recreação.

Kakkoth e Gharge (2018) apresentaram um sistema de reconhecimento de gestos de mão em tempo real baseado em descritor visual. Foram reconhecidos principalmente dez gestos com base na detecção da ponta dos dedos, que podem ser utilizados por pessoas com deficiência para transmitir esses gestos na forma de texto e som.

Nakazawa et al. (2018) desenvolveram um dispositivo vestível de comutação do tipo óculos com foco no movimento do globo ocular. As imagens ao redor do globo ocular são obtidas por uma câmera USB equipada com LED infravermelhos e as pupilas são extraídas pelo procedimento de transformação de Hough. O sistema de reconhecimento de direção do olhar pode distinguir doze níveis de "movimento do globo ocular", obtendo o centróide constituído por imagens de reflexão da superfície da córnea direita e imagem da pupila. Os autores desenvolveram também uma ferramenta de comunicação para pacientes com incursão neurológica usando o movimento do globo ocular como forma de interação.

Como os aplicativos baseados em gestos ganham espaço no domínio da educação especial, Sharma et al. (2018) estudaram seu potencial no contexto indiano. Uma das principais contribuições desse estudo é a disponibilização de diretrizes para projetar e desenvolver aplicativos baseados em gestos para indivíduos com deficiências de desenvolvimento. Os autores combinaram resultados e experiências de estudos com usuários utilizando aplicativos próprios (Kirana, Balloons e HOPE) para apresentar catorze diretrizes de *design* para sistemas baseados em gestos na expectativa de que educadores e cuidadores de todo o mundo transmitam habilidades sociais, motoras e de vida a indivíduos com deficiências de desenvolvimento. O aplicativo Kirana emprega gestos socialmente apropriados para ensinar habilidades para a vida, como comprar itens do dia-a-dia de uma mercearia indiana local. Ballons promove habilidades de atenção conjunta por meio de interação colaborativa. HOPE melhora a coordenação motora e as habilidades sociais e cognitivas dos usuários, com níveis crescentes de dificuldade.

O estudo publicado por Ascari et al. (2019) refere-se a um dos resultados desta tese de doutorado em que é apresentado o sistema PGCA e detalhes de um experimento conduzido com o público-alvo. O sistema PGCA é um sistema inteligente (baseado em VC e Aprendizado de Máquina) que permite a criação de interação gestual personalizada como recurso de TA para fins de comunicação. Esse sistema leva em consideração as habilidades e limitações motoras de seus usuários e o conhecimento dos cuidadores em reconhecer os gestos realizados pelos usuários. Os resultados de entrevistas com profissionais da área de educação especial e de um experimento com o público-alvo sugerem que o uso de gestos personalizados é uma prática comum para a CAA. Além disso, os autores destacaram sua percepção de que a criação de conjuntos de dados personalizados pode ser desafiadora, principalmente devido ao nível de entendimento dos participantes, à semelhança entre gestos e variações na execução dos mesmos gestos.

Krishnamurthi et al. (2019) desenvolveram um dispositivo ponteiro alternativo (*Frontier Point method* — FPM) para manipular uma apresentação do PowerPointTM, por meio de gestos, principalmente por pessoas com restrições nos movimentos de mãos ou braços. Os autores desenvolveram um sistema baseado no reconhecimento dinâmico de gestos de mãos, para execução de comandos para interação com slides de uma apresentação.

Soluções com base em VC desenvolvidas para permitir que pessoas com deficiência motora interajam com o computador são maioria dentre os estudos selecionados no mapeamento sistemático. Em geral os estudos citados exploram a capacidade dos usuários fazerem gestos distintos e reconhecíveis para desenvolver recursos de TA empregando abordagens diversas. Uma

preocupação comum a esses trabalhos é promover um modo de interação mais natural e intuitivo para pessoas com deficiência, contribuindo potencialmente para melhorar sua qualidade de vida.

3.2.2.4 *Interação Humano-Máquina*

Esta subseção apresenta estudos voltados a viabilizar a interação gestual de pessoas com dificuldades motoras e as máquinas, dispositivos ou equipamentos disponíveis no ambiente em que elas estão inseridas.

Fine e Tsotsos (2009) investigaram a viabilidade de um sistema capaz de obter *feedback* visual do usuário, por meio de expressões faciais, para uso em uma cadeira de rodas automática.

Mohammad e Anas (2015) apresentaram o modelo de *design* de uma cadeira de rodas elétrica para pessoas com tetraplegia e/ou quadriplegia, guiada por meio do rastreamento de movimentos da retina do usuário. O movimento ocular é detectado por uma webcam, e capturado por um sinal chamado eletrooculografia (EOG). Os sinais EOG são processados para que possam guiar o microcontrolador em interface serial, que, por sua vez, controla os movimentos da cadeira de rodas.

Lamb e Madhe (2016) apresentaram um estudo desenvolvido para controlar a posição de uma cama automaticamente usando o reconhecimento de gestos de mãos. Quatro movimentos podem ser reconhecidos de acordo com o gesto de mão apresentado, direcionando a cama para cima, para baixo, para a direita e para a esquerda.

Gao et al. (2017) projetaram um sistema de controle para cadeira de rodas robótica para pessoas com deficiência e idosos, utilizando interação por gestos de mão. Um novo método de detecção foi apresentado, unindo informações da cor da pele e de profundidade, para obter informações de velocidade de acordo com a posição da palma da mão em tempo real.

Como observado nos estudos citados, a interação gestual tem sido investigada como uma forma de obter uma interação humano-máquina eficaz e natural, acessível para usuários com capacidade de movimento muito limitada. Um dos estudos permitiu ao usuário controlar o posicionamento de uma cama usando gestos. Três dos quatro estudos citados investigaram o emprego de gestos para interagir com cadeiras de rodas elétricas, uma vez que o público-alvo não teria condições de controlar cadeiras de rodas elétricas tradicionais. Essas iniciativas contribuem para ampliar a independência das pessoas com deficiências motora na vida diária, usando sinais corporais que o usuário é capaz de executar para interagir com máquinas.

3.2.2.5 *Interação Humano-robô*

O rastreamento de movimentos de cabeça foi investigado por Zhang et al. (2013) em um protótipo de sistema de robô assistencial baseado em visão egocêntrica. Usando um par de óculos com uma câmera prospectiva, o usuário é envolvido ativamente no circuito de controle do robô em tarefas de navegação. A aplicação pretendida deste sistema de robô é ajudar uma pessoa com deficiência grave a agarrar um objeto alvo utilizando movimentos de cabeça.

Zhao et al. (2014) apresentaram uma plataforma interativa "Integração Homem-Robô" com multimodalidades, como postura da cabeça, olhar, voz e outros gestos corporais naturais. Os autores propõem um método interativo de julgamento de intenção, combinando estimativa de pose do modelo 3D da cabeça e a direção do olhar detectada pelo método de reflexão pupila-córnea. O objetivo do trabalho é aumentar o entendimento mútuo da situação de colaboração entre humanos e robôs de serviço.

Drawdy e Yanik (2015) apresentaram uma pesquisa voltada para robótica assistiva por meio de IHR utilizando a estimativa de ponto de interesse de um usuário (com base em rastreamento de olhar), para auxiliar na emissão de comandos para um robô, provendo um melhor

planejamento do caminho a ser percorrido. Segundo esses autores, para pessoas com deficiências, prevalece a necessidade de um paradigma de interação intuitiva entre humanos e robôs.

Saleh e Berns (2015) apresentaram um modelo biologicamente inspirado que realiza a comunicação verbal e não verbal entre humanos e robôs. A comunicação não verbal usada inclui movimentos de cabeça e comportamento dos olhos. Nesse modelo, o robô e seu parceiro de interação enviam e recebem informações ao mesmo tempo. O robô envia informações via fala e recebe o *feedback* do ser humano através de sugestões não verbais.

Os sistemas robóticos assistenciais podem ser muito valiosos para fornecer ajuda nas atividades da vida diária de pessoas com deficiência motora. Os resultados obtidos nos estudos citados indicam que dispositivos auxiliares controlados por gestos podem ser um meio viável para comunicação não verbal entre humanos e robôs. Os benefícios da IHR para pessoas com deficiência motora são importantes tanto do ponto de vista humanitário quanto de economia de custos.

3.2.3 Respostas às questões de pesquisa

Com base nos estudos mapeados, foram respondidas as questões de pesquisa definidas para o mapeamento sistemático apresentado. A questão principal do mapeamento teve como objetivo identificar as recentes soluções teóricas ou práticas desenvolvidas empregando VC para apoiar a comunicação e interação de pessoas com dificuldades motoras. O mapeamento apresentou instrumentos que são usados como recursos de CAA para viabilizar a comunicação ou interação de pessoas com dificuldades motoras, seja por meio de interação com o ambiente em que estão inseridos, com aplicativos de CAA ou para outros fins, conforme apresentado nas Tabelas 3.9 e 3.10.

Para cada questão apresentada na Tabela 3.4, foi elaborada uma resposta com base nas informações extraídas dos estudos mapeados. Com relação à pergunta "1) Como a interação gestual baseada em VC tem sido explorada para auxiliar a comunicação e interação de pessoas com dificuldades motoras?" Atualmente, existem muitas opções para interação gestual, oferecendo diferentes graus de acessibilidade para diversos propósitos. Entre as iniciativas identificadas pode-se destacar o uso de interação gestual para execução de atividades de lazer (música, arte), controle de equipamentos como cadeira de rodas ou cama, interação com robôs assistivos e interação com sistemas computacionais de um modo geral. Todas essas aplicações estão vinculadas de alguma forma com o objetivo de oferecer mais qualidade de vida para as pessoas com dificuldades motoras empregando técnicas de VC, seja dando-lhes algum nível de independência, facilitando o acompanhamento dos cuidadores e a comunicação, ou viabilizando a execução de atividades que de outra forma não seriam possíveis.

Considerando a questão "2) Quais tipos de câmera ou dispositivos têm sido utilizados?" Dentre os dispositivos identificados pode-se observar que câmeras de baixo custo e imagens RGB foram mais utilizadas como fonte de dados do que dados de profundidade, como apresentado na Tabela 3.11. Foi possível observar que há uma tendência em gerar alternativas mais acessíveis em termos de custo, inclusive substituindo dispositivos de rastreamento ocular, frequentemente baseados em câmeras infravermelhas com custo elevado.

Quanto à questão "3) Quais regiões corporais têm sido utilizadas para rastreamento?" A Tabela 3.9 e a Figura 3.3 apresentam informações extraídas dos estudos selecionados referentes a regiões corporais utilizadas como meio alternativo de interação para pessoas com dificuldades motoras. Dentre 29 estudos em que a identificação de regiões corporais utilizadas para rastreamento foi possível, 37% (11 de 29) investigaram o uso dos olhos como modalidade de interação, seja por meio de detecção de piscadas, estimativa da direção do olhar, detecção de pupila, íris ou esclera. Além desses, cerca de 13% dos trabalhos (4 de 29) investigaram o uso do

olhar associado a movimentos de cabeça. O rastreamento de movimentos de cabeça também é empregado em vários trabalhos, de forma isolada (2 trabalhos de 29), ou em conjunto com outras regiões corporais como olhos, mãos e pés. Foi possível verificar a grande variedade de possibilidades que já estão sendo utilizadas, explorando regiões corporais diversas, como mostra a Figura 3.3.

Sobre a questão "4) Quais técnicas têm sido empregadas para reconhecimento de gestos?" Foram identificados vários métodos para processamento de imagens, reconhecimento de padrões, algoritmos diversos e linguagens de programação empregadas conforme apresentado nas Tabelas 3.12 e 3.13. Quanto às linguagens de programação empregadas, percebe-se uma tendência na utilização de linguagens bem estabelecidas, que disponibilizam funções específicas para VC, como a biblioteca OpenCV. Muitas das técnicas apresentadas na Tabela 3.12 são empregadas por meio de funções disponíveis nessa biblioteca.

Com relação à questão "5) As soluções permitem a personalização dos gestos utilizados?" Com exceção do estudo vinculado a esta tese de doutorado (Ascari et al., 2019), não foram identificados outros estudos que descrevessem como seria possível aos usuários ou seus cuidadores realizarem atualizações nos conjuntos de gestos criados e utilizados para interação ou comunicação por meio de VC. É possível que esse resultado seja um viés decorrente da não inclusão de termos específicos como "*tailoring*", "*customization*", ou "*personalization*" na expressão de busca utilizada.

Quanto à questão "6) Existem metodologias vinculadas ao desenvolvimento das soluções?" O protótipo citado no estudo de Rosales et al. (2017) foi desenvolvido usando a metodologia de desenvolvimento de *software* Iconix (Rosenberg e Scott, 1999), que ajudou na análise dos requisitos, na análise preliminar, no *design* e implementação. O estudo de Ascari et al. (2019) vinculado a esta tese de doutorado citou o uso da metodologia proposta neste documento para orientar o desenvolvimento de soluções computacionais baseadas no reconhecimento de gestos como uma forma alternativa de interação ou comunicação especificamente para pessoas com deficiência motora e de fala. Com objetivo semelhante, mas voltado para um público mais abrangente, pode-se destacar o estudo de Sharma et al. (2018) que apresentam catorze diretrizes de *design* para projetar e desenvolver aplicativos que empregam gestos para indivíduos com deficiências de desenvolvimento. Não foram identificados estudos que apresentassem uma metodologia que possa ser comparada à apresentada nesta tese de doutorado. Esse resultado sugere uma lacuna na literatura e uma demanda por estudos para esse fim, ou mesmo que a nomenclatura "metodologia" não tenha sido utilizada para definir estudos relacionados à definição de formas de conduzir o desenvolvimento de sistemas de CAA para pessoas com dificuldades motoras e de fala, usando interação gestual.

Sobre a questão "7) As soluções voltadas para viabilizar comunicação usam o termo Comunicação Aumentativa e Alternativa?" Percebeu-se que o termo CAA não foi empregado por todos os trabalhos que apresentaram soluções voltadas para viabilizar a comunicação de pessoas com dificuldades motoras. Apenas 4 trabalhos de 7 usaram o termo CAA. Em alguns estudos o termo "*nonverbal communication*" foi empregado como referência ao uso de gestos para viabilizar a comunicação.

E por fim, quanto a questão "8) Pessoas com dificuldades motoras participaram da construção ou avaliação das soluções propostas?" Foi possível identificar que apenas 8 trabalhos de 33 envolveram pessoas com dificuldade motora no processo de construção ou avaliação das soluções apresentadas. Desses estudos, 5 incluíram representantes do público-alvo em etapas de avaliação das soluções desenvolvidas: Paquette et al. (2011), Diment e Hobbs (2014), Cristina e Camilleri (2016), Pal et al. (2017b) e Ascari et al. (2019). Já nos estudos de Graham-Knight e Tzanetakis (2015), Rosales et al. (2017) e Sharma et al. (2018) foi adotada uma abordagem

colaborativa envolvendo diferentes partes interessadas no processo de construção e avaliação das soluções.

3.2.4 Discussão

Os trabalhos apresentados neste capítulo representam iniciativas voltadas para diferentes contextos assistivos e para pessoas com diferentes níveis de deficiência física, mas de um modo geral, apresentam soluções tecnológicas para minimizar problemas enfrentados por pessoas que sofrem com algum tipo de comprometimento motor ou de fala. Os trabalhos acima relacionados de modos diversos destacam a viabilidade de empregar técnicas de VC para apoiar a interação de usuários com computadores.

Trabalhos como o de Hemmingsson et al. (2018) que permitiu a uma criança que não fala, com lesão medular alta, navegar na Internet, ler e-books, digitar documentos, entre outras ações, evidenciam as potencialidades da TA atual. Percebe-se que a deficiência não é mais tão limitante quanto foi no passado, e o foco dos trabalhos não consiste mais em verificar se é possível interagir com um computador ou dispositivo, mas sim, em tornar a interação melhor ou efetiva. Soluções diversas foram identificadas, e o próximo passo no campo de conhecimento da TA pode estar relacionado a diminuir o abandono das soluções existentes tornando-as acessíveis, mais flexíveis, fáceis de usar e disponíveis a um baixo custo.

Uma tendência no desenvolvimento de soluções de baixo custo, tem sido o emprego de câmeras simples ou *webcam* para aquisição de dados. Contudo, câmeras de profundidade também estão tornando-se mais acessíveis e já representam, portanto, uma alternativa viável. Outros dispositivos como *Eye-Tracker*, *Head-Tracker*, BCI, entre outros permanecem sendo utilizados, de forma isolada ou em conjunto. A metodologia proposta neste documento visa gerar uma estrutura acessível, e para isso prevê o uso de mais de uma modalidade de entrada e saída a fim de prover a interação multimodal.

A aprendizagem e reconhecimento de padrões em imagens é uma tarefa complexa que pode ser suportada por métodos de extração de características, como verificado em muitos dos trabalhos apresentados. Como identificado no levantamento bibliográfico, diferentes descritores de características foram utilizados. No entanto, a escolha das características é uma tarefa complexa, que requer muito conhecimento sobre o domínio do problema. Atualmente, as redes de aprendizado profundo vêm sendo bastante utilizadas para processar dados complexos e, em particular, as CNN vêm sendo empregadas eficientemente em várias tarefas de reconhecimento de padrões, inclusive no reconhecimento de gestos, como observado no trabalho de Ascari et al. (2019), vinculado a esta tese de doutorado. Nos trabalhos que utilizam CNN, a preocupação em escolher o melhor descritor de características não existe, uma vez que a CNN não requer etapas comuns de pré-processamento e extração de características, ou seja, os dados são utilizados de forma bruta, e dessa forma, evita-se que a precisão de um dado classificador seja determinada pela escolha de descritores que melhor representam o problema em questão.

Dentre os classificadores utilizados, foi possível observar que opções tradicionais como Adaboost, ANN, KNN e SVM permanecem populares e eficientes para o que se propõem. No entanto, novas alternativas têm sido exploradas pela comunidade científica, como transferência de aprendizado e CNN, e apresentado resultados promissores em aplicações de classificação (Han et al., 2018).

Considerando o processo de reconhecimento de padrões ou de gestos como um todo, desde a aquisição até a interação com interface computacional, foi possível observar que apenas o trabalho vinculado a esta tese de doutorado (Ascari et al., 2019) apresenta a possibilidade de o próprio usuário definir o conjunto de gestos a ser utilizado de forma flexível. Dentre os trabalhos apresentados como tendências recentes, muitos trabalharam com conjuntos de

gestos predefinidos, e o rastreamento de uma região corporal específica foi empregado pela maioria dos trabalhos apresentados neste capítulo. Os estudos em geral têm sido conduzidos com pacientes em ensaios clínicos, comprovando apenas a viabilidade técnica para transformar os sinais biológicos em comandos, mas falhando em fornecer percepções sobre a operação do dispositivo diariamente. No caso de trabalhos voltados ao reconhecimento de gestos, após sair das condições de laboratório, caso ocorra a necessidade de substituição de algum gesto pré-configurado, ou inclusão de novos, não foram observados outros trabalhos descrevendo como seria possível ao usuário e seu cuidador realizar esse tipo de personalização. A adoção com sucesso de um dispositivo depende de vários fatores, entre eles a capacidade do cuidador em aprender e personalizar a nova ferramenta (Kintsch e DePaula, 2002).

Outro ponto observado, foi que, os estudos em geral não apresentam detalhes sobre as dificuldades encontradas para criação do conjunto de dados, possivelmente por utilizarem conjuntos de dados existentes, ou criados com gestos executados por pessoas sem deficiência. Esse é um ponto crítico e de certa forma compreensível, pois em geral os projetistas de sistemas de CAA enfrentam dificuldades para coleta de dados devido às seguintes razões (Bhattacharya et al., 2008): é muito difícil obter número suficiente de usuários com deficiência para avaliar projetos alternativos; a coleta de dados de uso suficientemente grande para análise também é problemática, pois as incapacidades físicas impedem que os usuários trabalhem continuamente por um longo período de tempo, e em alguns casos são necessários vários meses para coletar dados de uso para avaliação.

Dentre os trabalhos descritos, percebeu-se que o uso do termo Comunicação Aumentativa e Alternativa não é padrão, visto que vários trabalhos foram desenvolvidos com foco na interação humano-computador como forma de facilitar a comunicação de pessoas com algum tipo de deficiência, mas não são descritos como uma contribuição específica para o campo da CAA.

Também, com base no levantamento bibliográfico, não foram encontrados outros estudos (com exceção de Ascari et al. (2019)) que apresentassem uma metodologia para apoiar a CAA por meio de reconhecimento de gestos, o que sugere que a proposta desta tese de doutorado é válida e pode suprir uma lacuna existente na literatura.

3.3 CONSIDERAÇÕES

Este capítulo apresentou uma revisão dos trabalhos na literatura relacionados a esta proposta. Foram apresentados os principais trabalhos voltados para CAA com base em TA. Também foram apresentados trabalhos que investigaram o uso ou o desenvolvimento de soluções que empregaram técnicas de VC em um contexto de TA, dando ênfase aos trabalhos voltados a pessoas que possuem dificuldades motoras. Foram selecionados via mapeamento sistemático 33 trabalhos publicados desde o ano de 2009, com o objetivo de identificar as principais técnicas e abordagens utilizadas para reconhecimento de gestos ou expressões descritas em publicações recentes. E por fim, uma visão panorâmica do estudo é discutida, estabelecendo relações entre os diversos resultados apresentados na literatura e suas implicações gerais para o problema de pesquisa abordado nesta tese.

4 METODOLOGIA MYPGI

Com base nos estudos mapeados em levantamento bibliográfico, observou-se que há uma tendência das soluções computacionais para CAA serem empregadas e focadas em situações específicas, oferecendo pouca flexibilidade e adaptabilidade que são características essenciais para a acessibilidade. Assim, a interação multimodal em interfaces adaptativas, informada por teorias e boas práticas da IHC, parece ser um caminho necessário e promissor para que as aplicações de CAA proporcionem aos seus usuários independência e competência em suas funções comunicativas. Com relação à adaptabilidade como fator de promoção de acessibilidade, o uso de sistemas de CAA por pessoas que têm, além de dificuldades de comunicação, comprometimento motor, é um desafio que demanda diferentes abordagens. Independentemente da origem do problema motor, é comum que esses usuários apresentem posturas muito particulares e movimentos involuntários que às vezes podem ser incontroláveis, impossibilitando o uso de muitas interfaces tradicionais.

Com o propósito de explorar alternativas que sejam adaptáveis e de baixo custo, é apresentada neste documento uma proposta de metodologia baseada em VC e Aprendizado de Máquina para guiar o projeto de sistemas de CAA usando interação gestual personalizada. Neste capítulo é descrita a metodologia proposta, destacando sua estrutura, possíveis cenários de aplicação, o sistema desenvolvido como prova de conceito a partir de seus pressupostos e a forma de avaliação empregada.

4.1 PRESSUPOSTOS E VISÃO GERAL

Esta pesquisa apresenta a proposta de uma metodologia que visa direcionar o projeto de sistemas computacionais de CAA utilizados principalmente por pessoas que possuem comprometimento motor e de fala. A metodologia, chamada *Methodology to yield Personalized Gestural Interaction* (MyPGI), baseia-se em boas práticas de IHC e em um processo de reconhecimento de gestos generalizável, criado e configurado pelo usuário e seu cuidador, corroborando com as observações de Rivera e DeSouza (2014), apresentadas no Capítulo 1. O termo "metodologia" empregado nesta tese refere-se ao meio ou método de fazer algo, como descrito por Mouton (1996). O objetivo da metodologia é ajudar no entendimento, nos termos mais amplos possíveis, não dos produtos da investigação científica, mas do próprio processo (Cohen et al., 2000).

O problema de pesquisa desta tese foi examinado a partir de uma perspectiva de resolução de problemas em IHC descrita por Oulasvirta e Hornbæk (2016). Esses autores elaboraram o conceito de capacidade de resolução de problemas de Larry Laudan (Laudan, 1978) como um critério universal para determinar o progresso de soluções (resultados): em vez de perguntar se a pesquisa é 'válida' ou segue a abordagem 'certa', incitam os pesquisadores a perguntarem-se como suas soluções avançam e qual sua capacidade de resolver problemas importantes no uso de computadores por humanos. Nessa perspectiva, os principais problemas de pesquisa em IHC classificam-se em três categorias principais: empíricos, conceituais e construtivos.

Considerando os objetivos desta tese, pode-se tratar o problema desta pesquisa como sendo de natureza mista, com características de natureza empírica e construtiva. É empírica, no sentido de testar e descrever o efeito de uma metodologia projetada para apoiar a CAA com base em interação gestual personalizada. E é construtiva, no sentido em que agrega informação na compreensão do uso de um sistema computacional de CAA por pessoas com dificuldades motora e de fala.

Em função do público-alvo desta tese apresentar muitas particularidades com relação à capacidade de realizar movimentos, a metodologia foi proposta com base nos resultados obtidos a partir da revisão de literatura e com o objetivo de atender os seguintes requisitos:

- a metodologia deve estar apoiada por técnicas de VC baseadas em abordagens não-invasivas;
- o método empregado para detecção e rastreamento de gestos deverá considerar qualquer movimentação realizada em frente à câmera, sejam movimentos de cabeça, braços, mãos ou outras partes do corpo humano;
- o método empregado para criação do conjunto de dados e treinamento do sistema deve ser generalizável, permitindo personalização de gestos e seus significados pelos próprios cuidadores e usuários;
- a interface dos sistemas de CAA desenvolvidos a partir da metodologia deve atender aos princípios de qualidade em IHC, principalmente quanto a usabilidade e acessibilidade;
- a metodologia deve permitir a interação multimodal e adaptação da interface de forma a possibilitar a interação gestual personalizada;
- a metodologia deve ser avaliada por meio de implementação de um sistema e por meio de experimentos conduzidos principalmente com representantes do público-alvo.

A metodologia proposta foi concebida para viabilizar o reconhecimento de padrões de movimentos usados para a interação gestual, capturados por meio de uma câmera de baixo custo. Essa mesma metodologia pode ser instanciada usando diferentes tipos de câmera ou dispositivos de entrada complementares (como BCI ou sensores de dispositivos móveis) a fim de viabilizar a interação multimodal com sistema de CAA. Embora a proposta esteja focada em um contexto específico, ou seja, usuários com dificuldades motora e de fala, é desejável que essa metodologia possa ser instanciada em diferentes contextos, e que as soluções projetadas com base nela atendam à maior diversidade possível de pessoas.

A Figura 4.1 representa a metodologia elaborada, usando a notação BPMN. As atividades relacionadas são agrupadas em etapas para facilitar sua descrição e seu entendimento. A estrutura da metodologia pode ser entendida em um nível macro, entretanto, ela depende de uma série de manipulações e processos específicos executados em um nível micro descritos a seguir por etapas. Detalhes sobre a implementação de cada uma dessas etapas são apresentados na Seção 4.3.

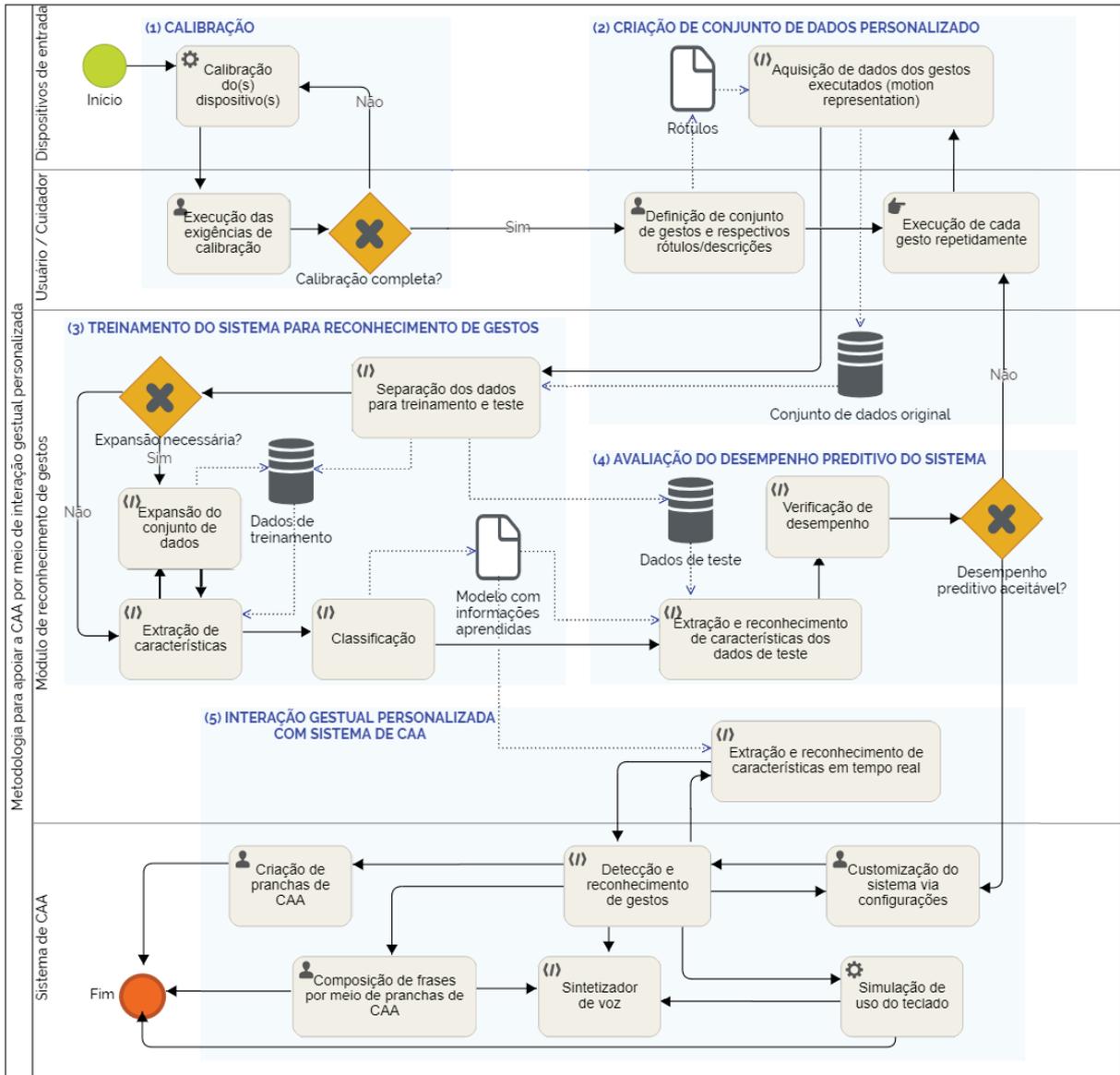


Figura 4.1: Estrutura da metodologia proposta, representada usando a notação BPMN. A metodologia é dividida em quatro raias que definem responsabilidades na execução das atividades e como os fluxos de trabalho cruzam funções ou transferem a atividade para outro responsável. As atividades relacionadas são agrupadas em etapas para facilitar sua descrição e seu entendimento.

4.1.1 Etapa 1 - Calibração

A etapa de calibração é uma condição necessária para que os resultados obtidos a partir de um dispositivo de entrada sejam confiáveis. A metodologia propõe que esse processo seja realizado sempre que um dispositivo de entrada for iniciado. Cada dispositivo utilizado pode apresentar requisitos específicos que devem ser atendidos nessa etapa inicial. A escolha do(s) dispositivo(s) empregado(s) deve levar em consideração as características dos gestos utilizados como forma de comunicação pelo público pretendido, e a estratégia a ser empregada pelo sistema para o treinamento e reconhecimento dos gestos.

A tarefa de calibrar uma câmera, por exemplo, consiste em encontrar as relações geométricas e os parâmetros envolvidos entre *pixels* e posições do mundo real (Everts et al., 2007). A calibração de uma câmera pode ter como objetivo exigir um determinado posicionamento do

usuário, como distância mínima pré-estabelecida ou centralização em relação ao dispositivo de entrada. As exigências dessa etapa devem ser atendidas pelo usuário que fará a execução dos gestos, tendo o auxílio de seu cuidador, caso seja necessário.

4.1.2 Etapa 2 - Criação de conjunto de dados personalizado

Para criação de um conjunto de dados, é necessário identificar o usuário (por exemplo informando seu nome) que fará a execução dos gestos, e definir um rótulo para cada gesto que será realizado para captura pelo dispositivo de entrada. Esse rótulo será utilizado para identificar o gesto executado e posteriormente para interação com o sistema ou composição de frases. Assim, os rótulos devem ser definidos pelo usuário e seu cuidador, de forma a representar palavras significativas para comunicação ou interação. Cada rótulo pode ser associado a uma imagem, que será apresentada no sistema de CAA após o reconhecimento do gesto em questão.

O cuidador é responsável por informar o rótulo de cada gesto, e o usuário é responsável por executar o gesto correspondente de forma que o dispositivo de entrada possa capturá-lo (gestos devem ter início e fim bem definidos). O processo de identificação de gestos pode ser feito via detecção de início de movimentação, ou a partir de algum outro gatilho utilizado pelo dispositivo de entrada. Gestos dinâmicos e estáticos podem exigir diferentes estratégias de implementação. A captura e representação final de um gesto pode ocorrer quando o sistema detecta diminuição significativa de movimentação, ou quando identifica que o gesto foi executado por completo. Assim, ao considerar que o gesto foi finalizado, os dados considerados relevantes para representação desse gesto (e seu respectivo rótulo) são gravados como uma amostra no conjunto de dados original. Esse processo de captura de amostras deve ser feito repetidamente, a fim de formar um conjunto de dados com diversas amostras para cada gesto.

4.1.3 Etapa 3 - Treinamento para reconhecimento de gestos

Após concluir a etapa de criação do conjunto de dados personalizado para um determinado usuário, pode-se executar a etapa de treinamento, na qual um módulo específico para reconhecimento de gestos permitirá que o sistema aprenda por meio de algoritmos de Aprendizado de Máquina a identificar as características dos gestos que compõem o conjunto de dados original, para então usar o conhecimento adquirido na classificação de novos gestos.

Por medida de segurança, pode-se inicialmente criar um *backup* das amostras originais, e então proceder com a divisão do conjunto de dados original (que contém todas as amostras cujos rótulos são conhecidos) em conjunto de dados de treinamento e conjunto de dados de teste. Há diversos métodos de amostragem que podem ser utilizados para execução dessa etapa, por exemplo Bootstrap (Efron, 1992), Holdout (Kohavi, 1995), validação cruzada K-fold (Geisser, 1975), e *Leave-one-out* (Nguyen e Poo, 2017). Cada estratégia procede a separação dos dados para treinamento e teste de maneira distinta, mas em todas um modelo de classificação é induzido a partir do conjunto de dados de treinamento e tem o seu desempenho avaliado utilizando um conjunto de dados de teste.

Para avaliação do desempenho do sistema, um modelo de classificação é criado a partir dos dados separados para treinamento, e as informações aprendidas nesta etapa são utilizadas para reconhecer os dados separados para teste. Para uso em tempo real, pode-se utilizar o modelo criado para avaliação de desempenho, ou considerar a criação de outro modelo, utilizando todas as amostras disponíveis no conjunto de dados. A partir do treinamento do modelo de classificação que será utilizado, cada novo gesto executado pelo usuário durante a interação com o sistema representará uma nova amostra de teste para esse modelo.

Dependendo do dispositivo de entrada usado, e da tecnologia empregada, é possível gerar novos exemplos a partir dos originais, por meio de um processo chamado "Aumento de Dados" (*Data Augmentation*), proposto originalmente por Tanner e Wong (1987). Esse processo consiste em criar amostras adicionais, a partir dos dados existentes, sendo útil em um cenário de dados de treinamento insuficientes, com o objetivo de melhorar o desempenho do classificador empregado. Portanto, se considerado necessário e viável, o conjunto de dados de treinamento pode ser expandido com a inclusão de novas amostras, geradas a partir de variações das amostras originais.

Tendo os conjuntos de dados de treinamento e teste definidos, pode-se realizar a etapa de extração de características, que consiste em analisar os dados de entrada a fim de extrair e derivar informações úteis para o processo de reconhecimento. Em uma abordagem tradicional, a extração de características pode ser feita em dois níveis: global ou localmente. Representações globais representam uma ação como um todo. As representações locais descrevem um videoclipe, por exemplo, como uma coleção de fragmentos. Os fragmentos são amostrados de forma escassa ou densa e, em seguida, são descritos usando descritores locais. Em uma abordagem mais recente, como no caso de aprendizagem profunda, a extração de características é realizada internamente, de forma automática, pois o algoritmo é exposto à uma grande quantidade de dados e de forma dinâmica determina quais características são relevantes. Independente da forma escolhida para extração de características, esse processo deve ser executado de igual forma para os dados do conjunto de treinamento, do conjunto de testes e também para os dados gerados durante a interação com o sistema de CAA em tempo-real (gestos executados para interação e uso do sistema).

Depois de selecionar as características desejadas de cada amostra, a detecção e reconhecimento dessas características por meio de algoritmos de Aprendizado de Máquina, chamados de classificadores, são a próxima etapa a ser realizada. Os classificadores identificam o conjunto de categorias que os subconjuntos de dados suportam e usam essas informações para atribuir uma classe a uma amostra não identificada ou desconhecida (Das, 2016). Diferentes algoritmos de classificação geralmente requerem conjuntos diferentes de representações de características.

Após a finalização da etapa de classificação, as informações aprendidas pelo sistema devem ser armazenadas para então serem utilizadas no reconhecimento de gestos em etapas seguintes.

4.1.4 Etapa 4 - Avaliação do desempenho do sistema

Na etapa de avaliação, o módulo de reconhecimento de gestos avalia o classificador gerado quanto ao seu desempenho preditivo, levando em consideração principalmente a proporção de classificações corretas para o total de gestos disponíveis para classificação.

A estimativa do desempenho do sistema é realizada utilizando-se um conjunto de gestos cujos rótulos são conhecidos, ou seja, o conjunto de dados de teste. Em um classificador ideal o número de acertos é igual ao número de amostras utilizadas para teste, ou seja, o classificador não comete erros. Desse modo, por meio de análise comparativa, pode-se calcular o percentual de gestos corretamente classificados para o conjunto de gestos em questão. Se a precisão do modelo for estimada a partir dos mesmos exemplos utilizados para sua construção, provavelmente, uma estimativa otimista será obtida. Em função disso, a metodologia utiliza exemplos que não fizeram parte do conjunto de treinamento, ou seja, dados do conjunto de testes, para estimar o desempenho do classificador.

Existem diferentes métricas de qualidade e precisão para avaliar o desempenho de um classificador, e muitas delas baseiam-se na matriz de confusão, uma tabela que contém informações sobre os valores reais e os valores previstos para um classificador. A partir desses

dados, pode-se calcular precisão, sensibilidade ou taxa de falsos positivos, taxa de classificação incorreta, especificidade, e curva ROC (*Receiver Operating Characteristic*) (Das, 2016).

Se o teste de desempenho utilizado produzir resultados considerados aceitáveis para o contexto em questão, de acordo com a métrica desejada, o classificador poderá ser utilizado no reconhecimento de novos gestos, ou seja, gestos cujas classes (rótulos) são desconhecidas. Caso contrário, se o sistema não estiver reconhecendo de forma satisfatória os gestos utilizados para compor o conjunto de teste, o próprio sistema poderá sugerir ao usuário a realização de uma nova coleta de dados. Nesse caso é importante informar o usuário sobre os gestos que estão apresentando menor taxa de acerto, destacando gestos que possam estar sendo executados de forma muito semelhante, e que estejam sendo confundidos pelo classificador. Se for optado por ajustar as amostras do conjunto de dados original, realizando uma nova coleta de dados, é importante que o sistema restaure o *backup* feito na Etapa 1, antes de se iniciar a nova captura ou realizar a exclusão de amostras para um determinado gesto.

4.1.5 Etapa 5 - Interação gestual personalizada com sistema de CAA

A interação gestual personalizada com um sistema de CAA é a etapa final prevista na metodologia, na qual os gestos executados pelo usuário, juntamente com as demais entradas consideradas pela solução desenvolvida, serão reconhecidos e utilizados com propósito de comunicação e interação com a interface de um sistema computacional.

Sempre que forem identificados início e fim de movimentação, o sistema deve considerar que um gesto foi executado, e o módulo de reconhecimento de gestos deve acionar o classificador já treinado para tentar classificar o gesto em questão. Ao reconhecer um gesto, o sistema poderá usá-lo como um gatilho para execução de ações diversas na interface do sistema.

Entre as ações previstas para um sistema de CAA implementado a partir da metodologia proposta, destacam-se: (a) customização; (b) criação de pranchas de CAA personalizadas; (c) simulação de uso do teclado, permitindo que o usuário navegue entre os controles da interface, altere configurações ou interaja com outros aplicativos; d) composição de frases utilizando os rótulos dos gestos cadastrados ou realizando seleções nas pranchas de comunicação disponíveis.

Como uma forma de melhorar a acessibilidade do sistema de CAA para usuários não alfabetizados, um sintetizador de voz pode ser utilizado para reproduzir ao usuário o som correspondente às opções selecionadas nas pranchas de comunicação ou à descrição dos rótulos/ações reconhecidas. Mensagens e instruções de uso do sistema também podem ser reproduzidas com som, ou caso o sistema possa ser utilizado por pessoas com dificuldades auditivas, deve ser considerada a reprodução de mensagens em linguagem de sinais.

Para que a interação gestual ocorra de forma efetiva, critérios de qualidade em IHC, principalmente relacionados a acessibilidade e usabilidade, devem ser considerados no desenvolvimento do sistema de CAA, de modo a contemplar os interesses dos usuários e garantir a qualidade de uso.

4.2 CENÁRIOS DE APLICAÇÃO

Pode-se apresentar como exemplo, pelo menos três cenários distintos em que a metodologia proposta, poderia ser aplicada:

- Uma criança com comprometimento da fala, utiliza um sistema de CAA com base em reconhecimento de gestos em seu dispositivo móvel particular ou computador pessoal, em seu ambiente escolar, para poder interagir com seus colegas e professores, e conseqüentemente, ampliar ou facilitar seu aprendizado.

- Um adulto com dificuldades motora e de fala, que se locomove por meio de uma cadeira de rodas, na qual poderia ser acoplada uma *webcam* ou um dispositivo móvel como celular ou *Tablet*, ou ainda, um óculos inteligente. A câmera desse dispositivo permitiria que ele interagisse com o sistema para poder realizar suas requisições diárias, ou mesmo formar narrativas pessoais.
- Um idoso enfermo, com movimentos e fala comprometidos, poderia realizar movimentos em frente a uma *webcam* adaptada em seu quarto para realizar requisições ou indicar quando precisa de auxílio de seu cuidador.

Em todas essas situações, espera-se que um sistema de CAA instanciado pela metodologia MyPGI seja utilizado, de forma a oferecer uma solução simples, flexível e personalizada para cada usuário explorando e aliando as potencialidades das áreas de VC, IHC, Aprendizado de Máquina e TA. Os resultados apresentados nesta tese não contemplam testes em todos os cenários citados, mas espera-se que a autonomia proporcionada às pessoas com dificuldades motora e de fala por meio do emprego dessa metodologia no desenvolvimento de sistemas de CAA represente um passo importante à reintegração dessas pessoas na sociedade ou para melhoria da sua qualidade de vida.

4.3 IMPLEMENTAÇÃO - SISTEMA PGCA

Com o objetivo de colocar em prática todas as etapas previstas na metodologia proposta, um sistema computacional de CAA foi desenvolvido como prova de conceito. O sistema desenvolvido foi chamado de Assistente de Comunicação Pessoal por Gestos (*Personal Gesture Communication Assistant* — PGCA) e segue os pressupostos da metodologia MyPGI, considerando que gestos e seus significados são criados e configurados pelos usuários e seus cuidadores. PGCA é um sistema inteligente (baseado em VC e Aprendizado de Máquina) desenvolvido para apoiar a CAA como um recurso de TA para interação gestual personalizada, explorando alternativas adaptáveis e de baixo custo. A primeira versão do sistema foi avaliada por profissionais que atuam com IHC e os resultados, descritos em Ascari et al. (2018b), direcionaram a implementação de melhorias no sistema para uma nova versão apresentada e avaliada por voluntários sem dificuldade motora e de fala, e também por representantes do público-alvo.

Para desenvolvimento do sistema foi adotado um modelo de processo da área de desenvolvimento de *software* prescrito por Pressman e Maxim (2016), chamado de prototipação evolucionária. Esse modelo prevê um refinamento sucessivo do protótipo em várias etapas, auxiliando na identificação dos requisitos, até que o objetivo final seja atingido. O objetivo da prototipação evolucionária é fornecer aos usuários finais um sistema funcionando. O desenvolvimento começa com aqueles requisitos que são melhor compreendidos. O protótipo é mantido, e é continuamente alterado e expandido, evoluindo para o produto final. Em todas as iterações, os requisitos e o *design* do sistema são refinados com base no *feedback* dos usuários ou do proprietário do produto. Para o desenvolvimento do sistema apresentado nesta tese, cinco experimentos foram conduzidos com usuários distintos, que forneceram contribuições significativas para a adição, alteração ou aprimoramento de requisitos, conforme descrito no Capítulo 5.

O sistema PGCA emprega uma abordagem de baixo custo, utilizando uma câmera simples (*webcam*) e computador pessoal para realizar a captura e rastreamento dos gestos executados. O sistema captura movimentos da câmera e os representa na forma de imagens: (i) imagem de histórico de movimento convencional (MHI); ou (ii) imagem de histórico de movimento com base em fluxo óptico (OF-MHI). Dois classificadores podem reconhecer esses

movimentos (gestos): (i) classificador desenvolvido com base em SVM; e (ii) classificador desenvolvido com base em CNN e Transferência de Aprendizado. Para a implementação do sistema, utilizou-se principalmente a linguagem de programação C++ e a plataforma QT. O classificador com base em CNN foi implementado usando linguagem de programação Python embutida no aplicativo C++. Todas as ações realizadas no sistema por meio de interação gestual são armazenadas em um arquivo texto (*log*) com o objetivo de registrar eventos relevantes na interação com a interface. A seguir são apresentados detalhes de como cada etapa prevista na metodologia MyPGI foi implementada no sistema PGCA.

4.3.1 Implementação da Etapa 1 - Calibração

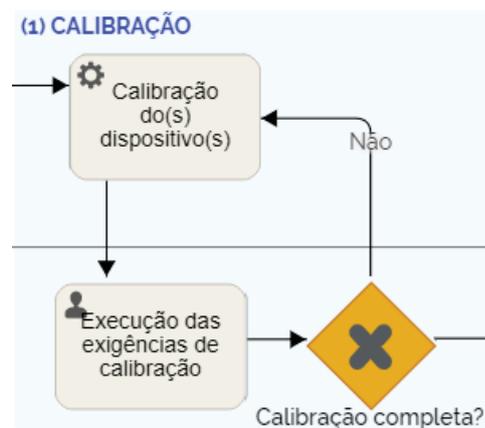


Figura 4.2: Estrutura definida na metodologia MyPGI para a Etapa 1 (Calibração).

Conforme previsto na Etapa 1 da metodologia (Figura 4.2), quando o sistema é iniciado, a câmera automaticamente é ligada e é solicitado ao usuário que realize o processo de calibração. A calibração tem como objetivo posicionar o usuário no centro de captura da câmera, facilitando assim a padronização de posturas para gravação e reconhecimento dos movimentos. Esse processo foi implementado levando em consideração que a ponta do nariz do usuário deve passar em um momento pelo centro da imagem capturada, identificado por duas linhas retas que se cruzam, como exemplificado na Figura 4.3. Para detecção da face e da ponta do nariz, foram utilizados os métodos disponíveis em "CascadeClassifier" da biblioteca OpenCV, que se baseiam em um algoritmo de classificação em cascata (*Haar Cascade*) proposto por Viola e Jones (2001). Esse algoritmo é amplamente usado em tarefas de reconhecimento de face pois apresenta taxas competitivas de detecção de objetos em tempo real (Peleshko e Soroka, 2013).

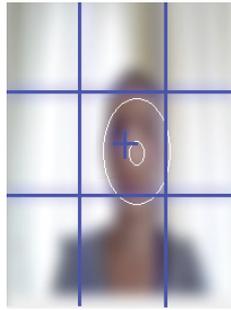


Figura 4.3: Imagem representativa do processo de calibração realizado pelo sistema PGCA, com o objetivo de posicionar o usuário no centro de captura da câmera. Esse processo exige que a ponta do nariz do usuário passe em um momento pelo centro da imagem capturada, identificado por duas linhas retas que se cruzam. O círculo maior apresentado na figura corresponde à região da face do usuário e o círculo menor corresponde à localização do nariz do usuário.

Depois de concluir esse processo, o cuidador pode começar a personalizar o sistema usando a guia na interface específica para isso: Área do Cuidador. O cuidador auxilia o usuário a registrar exemplos de gestos (amostras) que serão utilizados para treinar o sistema, e, posteriormente, como um modo de interação com a interface. Idealmente, qualquer usuário poderia personalizar o sistema por meio de gestos de forma independente. Contudo, para esta versão exploratória do sistema, é necessário a assistência do cuidador para a configuração inicial do sistema e para registrar e rotular os gestos com o usuário. Embora o apoio de um cuidador seja sempre necessário quando as pessoas têm deficiências extremas, principalmente as cognitivas que afetam a interação intencional, o sistema foi projetado para permitir sua configuração e uso por usuários que apresentem ao menos um movimento intencional padronizado e identificável.

4.3.2 Implementação da Etapa 2 - Criação de conjunto de dados personalizado



Figura 4.4: Estrutura definida na metodologia MyPGI para a Etapa 2 (Criação de conjunto de dados personalizado).

Para a aquisição de cada exemplo que irá compor o conjunto de dados, conforme previsto na Etapa 2 da metodologia (Figura 4.4), o movimento executado é capturado e representado como uma imagem de histórico de movimento (MHI). Esse processo é feito por meio de conversão

da imagem capturada em tons de cinza, obtenção da diferença entre *frames*, aplicação de *threshold* e atualização do histórico de movimento gerando uma imagem na qual tons de cinza claro representam a movimentação mais antiga, e tons de cinza escuro a movimentação mais recente. Para aplicação de *threshold*, foi utilizado no sistema PGCA um método de mesmo nome disponível na biblioteca OpenCV, com parâmetros *thresh (limiar) = 30*, *maxval = 1* e *type = 0*.

A imagem de histórico de movimento é usada principalmente para descrever a tendência do movimento relacionado à ação (gesto), incluindo descrição da informação temporal do movimento (Zhang et al., 2016). Duas técnicas para criar essa imagem foram empregadas no sistema PGCA, uma com base na definição de MHI convencional, e outra (OF-MHI) que agrega informação de velocidade via fluxo óptico à imagem MHI. A Figura 4.5 mostra dois exemplos de representação de um mesmo movimento usando MHI e OF-MHI. Detalhes sobre a implementação de cada uma dessas técnicas são descritos nas seções seguintes.

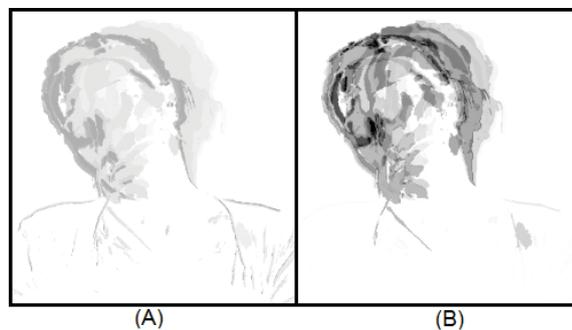


Figura 4.5: Exemplos de representação de movimento por meio de MHI (A) e OF-MHI (B) referente a um gesto executado com a cabeça.

4.3.2.1 Representação de movimento usando MHI convencional

Proposta originalmente por Davis e Bobick (Davis e Bobick, 1997; Bobick e Davis, 2001), MHI é uma representação espaço-temporal global do movimento que foi aplicada à análise de movimento e rastreamento para diferentes propósitos, como reconhecimento de gestos (Vafadar e Behrad, 2008) ou reconhecimento de ação humana (Huang et al., 2011; Zhang et al., 2016). O MHI converte as informações espaço-temporais 3D de uma sequência de vídeo em uma única imagem de intensidade 2D, incluindo informações de tempo e espaço e refletindo também a ordem dos movimentos.

No MHI, um valor de intensidade alta fixo é atribuído a um *pixel* em primeiro plano (objeto em movimento), enquanto o valor da intensidade é diminuído em uma pequena constante para um *pixel* em segundo plano (Tsai et al., 2015). O valor da intensidade no MHI registra o histórico de alterações temporais em cada localização de *pixel*. O MHI $H_\tau(x, y, t)$ é calculado a partir de uma função de atualização $\psi(x, y, t)$ descrita por Davis e Bobick (1997) e apresentada na Equação (4.1):

$$H_\tau(x, y, t) = \begin{cases} \tau, & \psi(x, y, t) = 1 (\in \text{primeiro plano}) \\ \max(0, (H_\tau(x, y, t-1)) - \delta), & \text{senão} \end{cases} \quad (4.1)$$

onde (x, y, t) são as coordenadas espaciais (x, y) de um *pixel* da imagem no tempo t (em termos do número do quadro da imagem). A duração τ determina a extensão temporal do

movimento em termos de quadros e δ é o parâmetro de decaimento. Foi utilizado no sistema PGCA $\tau = 3$ e $\delta = \text{timestamp}$ (incrementa a intensidade dos *pixels* em primeiro plano com o passar do tempo). $\psi(x, y, z)$ é definido na Equação (4.2), conforme descrito por Fan e Tjahjadi (2017):

$$\psi(x, y, z) = \begin{cases} 1, & D(x, y, t) \\ 0, & \text{senão} \end{cases} \quad (4.2)$$

onde $D(x, y, t)$ é uma imagem binária que compreende a diferença de intensidade de *pixel* dos quadros separados pela distância temporal Δ , definida na Equação (4.3):

$$D(x, y, z) = |I(x, y, t) - I(x, y, t \pm \Delta)| \quad (4.3)$$

onde $I(x, y, t)$ é o valor da intensidade do *pixel* (x, y) no quadro t_{th} da sequência da imagem. Nesse estudo, para calcular o MHI, foi utilizado a função "updateMotionHistory" disponível na biblioteca OpenCV.

4.3.2.2 Representação de movimento usando OF-MHI

Fluxo óptico (Horn e Schunck, 1981; Lucas e Kanade, 1981) denota uma mudança da mesma cena na sequência de imagens em um momento diferente no tempo, estimando o movimento em nível de *pixel* entre duas imagens. O fluxo óptico pode ser usado para agregar informações de velocidade do movimento ao MHI. No MHI convencional, para cada *pixel* de primeiro plano detectado é atribuído um valor de intensidade fixo. Ou seja, um movimento lento e um movimento rápido de diferentes partes do corpo terão a mesma força de movimento (Tsai et al., 2015), ou seja, diferenças de velocidade não são consideradas.

Diferentes propostas foram apresentadas para adicionar informação de velocidade em MHI usando fluxo óptico, como Tsai et al. (2015), Fan e Tjahjadi (2017) e Khalifa et al. (2018), e uma proposta semelhante foi usada neste estudo. No sistema PGCA, um algoritmo de rotulagem (função "connectedComponentsWithStats" disponível na biblioteca OpenCV) é aplicado a uma silhueta obtida pela diferença de quadros para identificar regiões conectadas. Posteriormente, o fluxo óptico de Lucas e Kanade (1981) é calculado para os *pixels* do centróide de cada uma dessas regiões, e esse valor de deslocamento é replicado para os outros *pixels* da mesma região para tornar o processo de rastreamento via fluxo óptico rápido. O valor de intensidade resultante indica o histórico da velocidade do movimento nesse local. Quando a opção "Aproximar" é marcada nas configurações do sistema, a área de captura da câmera é aproximada, e os pontos de referência faciais (*Landmarks*) são inseridos como pontos a serem rastreados.

No sistema PGCA foi empregada a função "calcOpticalFlowPyrLK" disponível na biblioteca OpenCV para calcular o fluxo óptico de Lucas-Kanade, com os parâmetros winSize (31×31), minEigThreshold (0.001) e os demais parâmetros com seus valores padrão. O MHI com base em fluxo óptico (OF-MHI) é definido na Equação (4.4), descrita por Fan e Tjahjadi (2017):

$$E(x, y, t) = s(x, y, t) + E(x, y, t - 1) \cdot \alpha \quad (4.4)$$

onde $s(x, y, t)$ representa o comprimento do fluxo óptico do *pixel* (x, y) no instante t , e α é a taxa de atualização usada ($0 < \alpha < 1$). A força do movimento é dada pelo comprimento do fluxo $s(x, y, t)$ para cada *pixel* individual (x, y) . A intensidade de um *pixel* é incrementada se ele é um *pixel* de primeiro plano. Um pequeno valor de α cria uma diminuição acelerada da força do movimento, e apenas os movimentos recentes de curto prazo são retidos no modelo temporal.

Valores maiores de α , por sua vez, originarão um histórico de longo prazo no modelo temporal. No sistema PGCA foi utilizado 0,85 para o parâmetro α .

A Figura 4.6 mostra os pontos usados para rastrear via fluxo óptico os movimentos de um gesto executado com a mão e de uma expressão facial. Ambos os gestos são representados em tons de cinza, onde tons mais escuros representam a movimentação mais recente, ou executada com maior velocidade.

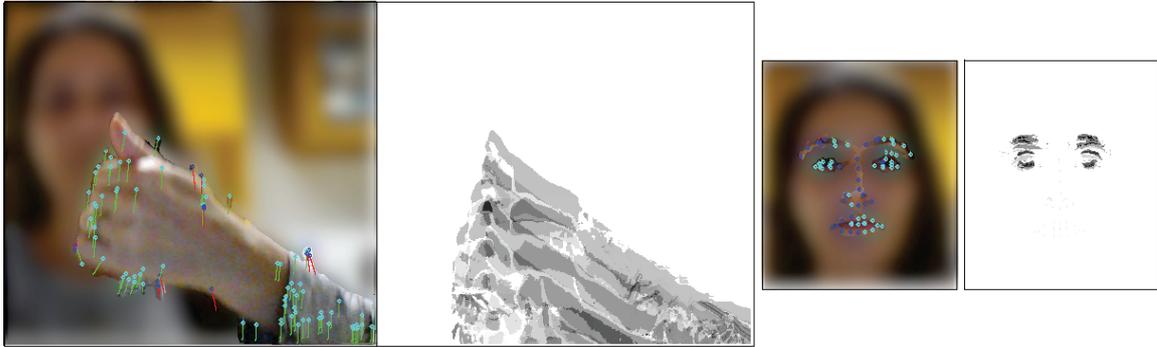


Figura 4.6: Pontos usados para rastrear e calcular o fluxo óptico de um gesto de mão (movimentos para cima e para baixo) e de uma expressão facial (usuário levanta as pálpebras em uma expressão de surpresa). As imagens em tons de cinza correspondem à representação do movimento por meio de OF-MHI.

4.3.3 Implementação da Etapa 3 - Treinamento para reconhecimento de gestos

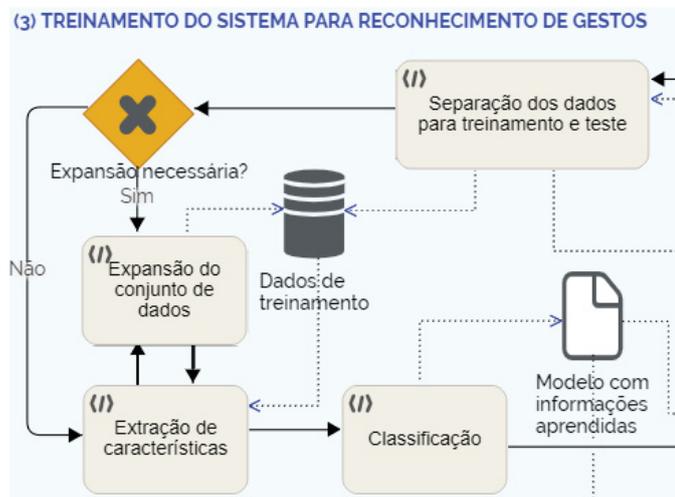


Figura 4.7: Estrutura definida na metodologia MyPGI para a Etapa 3 (Treinamento para reconhecimento de gestos).

O reconhecimento de gestos (Figura 4.7) foi implementado no sistema PGCA por meio de algoritmos de Aprendizado de Máquina. Dois classificadores foram implementados, utilizando diferentes abordagens, com o objetivo de comparar o desempenho de ambos no contexto desta proposta: Classificador com base em SVM e Classificador com base em CNN.

Independente do classificador empregado (indicado em configuração disponível no sistema) dois modelos para reconhecimento de gestos são gerados pelo sistema PGCA: o modelo para a etapa de avaliação; e o modelo para a etapa de interação do usuário com o sistema, que é gerado usando todas as amostras disponíveis no conjunto de dados. Nos dois modelos, o número de dados de treinamento é expandido via Aumento de Dados, onde amostras adicionais são criadas a partir dos dados existentes.

Após gerar alguns exemplos de imagens MHI ou OF-MHI (de acordo com a configuração utilizada) para cada rótulo de gesto a ser utilizado pelo usuário (pelo menos dez), o sistema realiza o processo de Aumento de Dados. Cada amostra do conjunto de dados de treinamento é aumentada, criando outras oito variações, girando e dimensionando a imagem original, com o objetivo de simular pequenas alterações no posicionamento da câmera ou na variação de distância que pode ocorrer quando os usuários interagem com o sistema. A Figura 4.8 mostra um exemplo de um gesto de mão representado por uma imagem OF-MHI, em que a imagem original (central) é usada para gerar outras oito imagens para ampliar o conjunto de dados usado pelo sistema. As variações empregadas foram -10 e 10 para ângulo e $0,9$ e $1,1$ para escala.

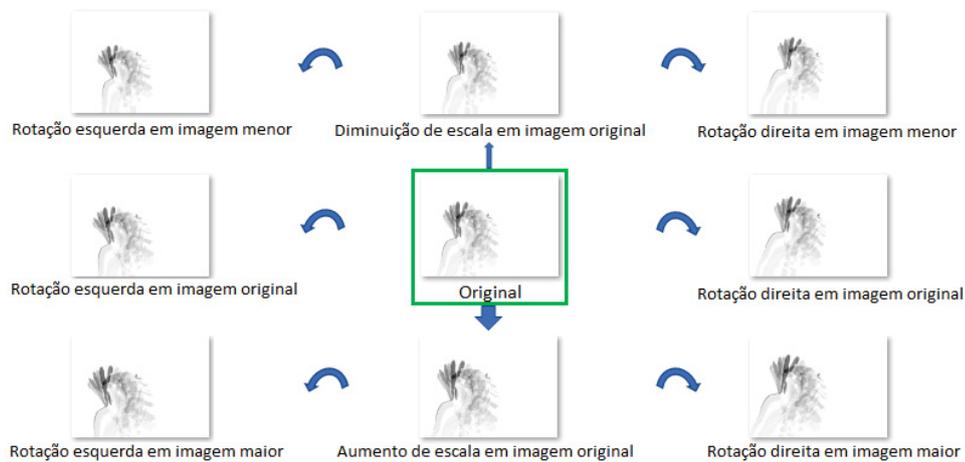


Figura 4.8: Exemplo de aplicação de Aumento de Dados. A partir de uma imagem representativa de gesto dinâmico (imagem original), outras oito variações são geradas por meio de operações de rotação e dimensionamento.

As informações aprendidas na etapa de treinamento são armazenadas, e posteriormente utilizadas na etapa de avaliação do conjunto de teste e também para reconhecimento dos gestos executados durante interação com a interface do sistema.

4.3.3.1 Classificador com base em SVM

Para a primeira versão do sistema PGCA, foi utilizado um classificador discriminante multiclasse SVM associado ao descritor de características HOG para reconhecimento de gestos. Classificadores discriminantes são treinados para separar dados em duas classes aprendendo um hiperplano em um espaço dimensional maior. O SVM (Salcedo-Sanz et al., 2014) é um classificador binário linear que atribui uma dada amostra a uma classe, de apenas duas classes possíveis (Mountrakis et al., 2011). O método "um-contra-todos" foi usado para tratar problemas de múltiplas classes: dado um problema de n -classes, para cada classe um modelo binário é construído, o conjunto de treinamento consiste em exemplos dessa classe como rótulos positivos e exemplos de outras classes como rótulos negativos.

HOG é um descritor de recurso usado para detecção de objetos, obtido com base no histograma de gradiente da imagem. Para esse classificador, o HOG foi extraído das imagens MHI ou OF-MHI completas (como exemplificado na Figura 4.9), redimensionadas para (64×48) , gerando um vetor de características com 1260 posições. Para extrair características usando o descritor HOG foi empregada a função "HOGDescriptor" disponível na biblioteca OpenCV e parâmetros incluem winSize $(64, 48)$, blockSize $(16, 16)$, blockStride $(8, 8)$, cellSize $(8, 8)$, nbins (9) , derivAperture (1) , winSigma (4) , histogramNormType (0) e outros parâmetros com seus valores padrão. A função "hog.compute" foi empregada na sequência, com parâmetros winStride (32×24) e padding $(0, 0)$.

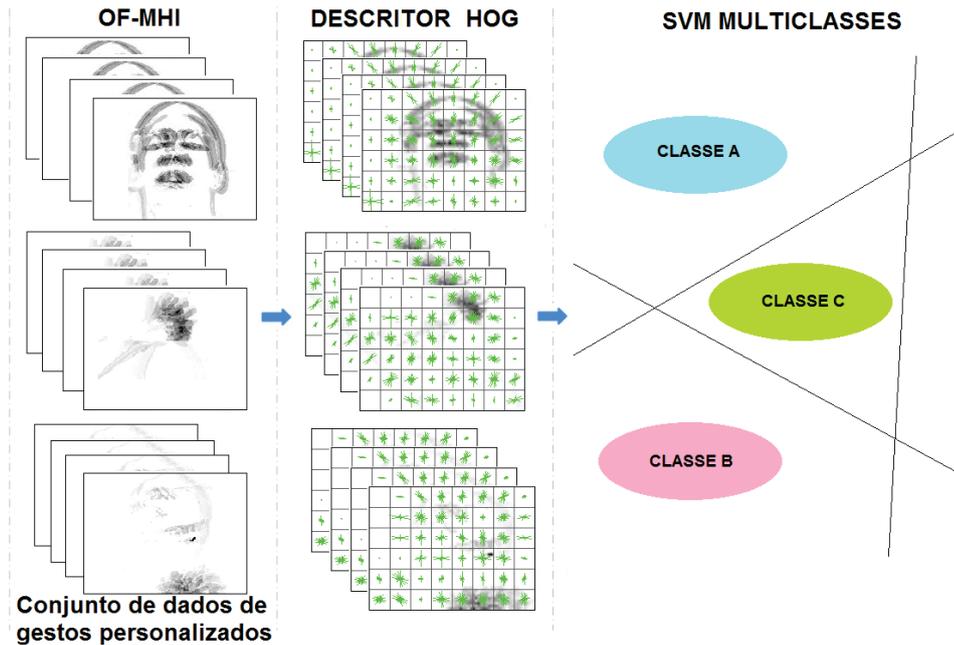


Figura 4.9: Esquema empregado pelo classificador com base em SVM para a etapa de treinamento do sistema PGCA. As amostras criadas de forma personalizada e representadas como imagens de histórico de movimento tem suas características descritas via descritor HOG, para então proceder a classificação dos gestos.

4.3.3.2 Classificador com base em CNN

Os descritores da CNN são capazes de gerar recursos mais sofisticados e difíceis de elaborar em meios descritivos, mas para isso demandam um conjunto de dados grande e variado para realizar o treinamento da CNN a partir do zero. No contexto desta tese, o número de amostras é limitado porque cada usuário criará seu próprio conjunto de dados. Assim, o conceito de Transferência de Aprendizado (*Transfer Learning*) foi considerado uma alternativa viável porque pode melhorar o conhecimento do aprendiz (ou seja, o mecanismo de aprendizado) sobre um domínio, transferindo informações aprendidas de um domínio relacionado (Weiss et al., 2016). Para isso, foi usado o *TensorFlow* (Abadi et al., 2016) *Inception V3* (Szegedy et al., 2016), um codinome para um modelo de arquitetura de CNN profundo, originalmente treinado no banco de dados *ImageNet* (Russakovsky et al., 2015), como base para treinar um conjunto personalizado de imagens. Aplicou-se Transferência de Aprendizado, retreinando a camada final do *Inception*, por 4000 etapas, com novas categorias, para construir um classificador de imagem personalizado de acordo com rótulos e gestos gerados pelo usuário do sistema em conjunto com seu cuidador. A Figura 4.10 mostra um resumo da arquitetura empregada para treinar o sistema usando o conceito de Transferência de Aprendizado no classificador com base em CNN.

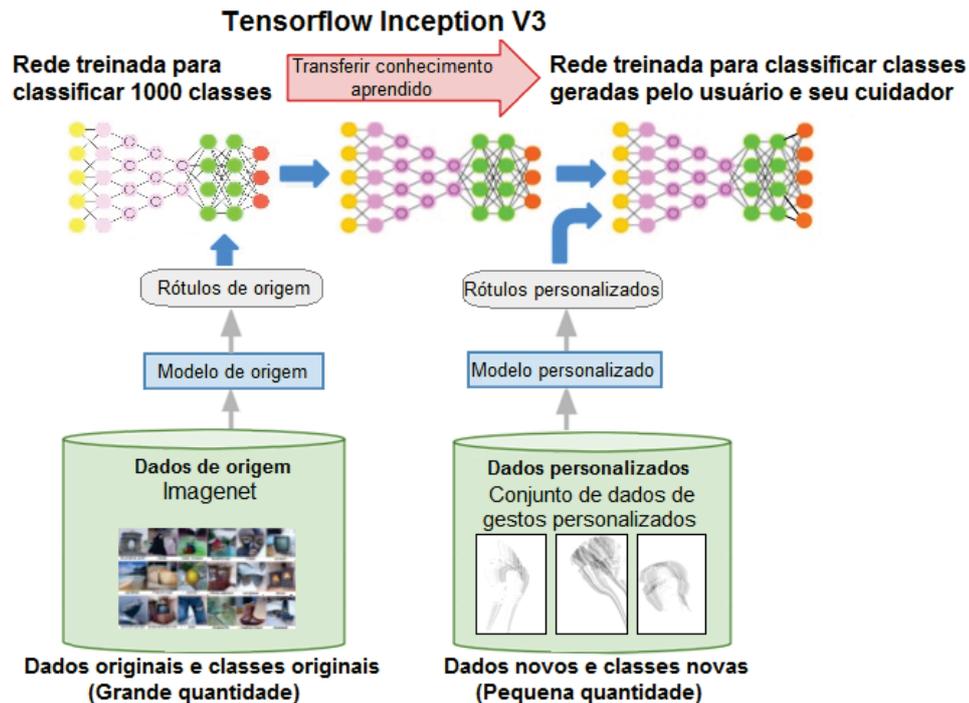


Figura 4.10: Esquema empregado pelo classificador com base em CNN e Transferência de Aprendizado para a etapa de treinamento do sistema PGCA. A rede neural *TensorFlow Inception V3* é treinada originalmente com o conjunto de imagens coloridas *Imagenet*. O conhecimento adquirido nesta etapa de treinamento é mantido, e posteriormente apenas a última camada da rede é retreinada usando o conjunto de imagens (em tons de cinza) criado pelo usuário do sistema em conjunto com seu cuidador. Assim a rede torna-se apta a classificar os gestos personalizados.

4.3.4 Implementação da Etapa 4 - Avaliação do desempenho do sistema



Figura 4.11: Estrutura definida na metodologia MyPGI para a Etapa 4 (Avaliação do desempenho do sistema).

Conforme a Etapa 4 prevista na metodologia (Figura 4.11), o sistema PGCA permite avaliar o desempenho do sistema usando dois métodos: Holdout (Kohavi, 1995) e validação cruzada K-fold (Geisser, 1975). O Holdout é um método básico de validação usado para estimar o desempenho do sistema no reconhecimento de gestos, tendo rápida execução. No sistema PGCA, para gerar o modelo que será usado na etapa de avaliação usando o método Holdout, o conjunto de dados original é dividido em dois subconjuntos, com 2/3 dos dados no conjunto de treinamento e 1/3 dos dados no conjunto de testes. O método de validação cruzada K-fold gera uma estimativa geral mais precisa sobre a capacidade do sistema de reconhecer os gestos capturados, mas demanda mais tempo de execução, dependendo do número de pastas (agrupamentos que separam as amostras dos gestos em parcelas de dados iguais) empregado. Para a validação cruzada K-fold, o sistema PGCA trabalha com dez pastas, criando dez modelos distintos, separando 90% dos

dados para treinamento e 10% para teste em cada modelo. A média ponderada do desempenho do sistema nas dez pastas indica o desempenho geral do sistema.

Após a execução de testes em todas as pastas, a precisão geral (média ponderada do número de acertos) é calculada. Normalmente, uma classificação perfeita produz um valor de precisão de 100%. Para o contexto desta tese, considerou-se desejável uma precisão de pelo menos 90% para que o modelo possa ser usado satisfatoriamente para interação gestual. Se o teste de precisão indicar um desempenho ruim no reconhecimento dos gestos, uma descrição dos gestos capturados é apresentada ao usuário, contendo a taxa de acertos de cada um e se há algum gesto que está sendo confundido com outro existente no conjunto de dados. Essas informações permitirão que o usuário realize uma nova coleta de dados ou exclua do conjunto de dados os gestos que não estão sendo reconhecidos corretamente para então realizar uma nova etapa de treinamento.

4.3.5 Implementação da Etapa 5 - Interação gestual personalizada com sistema de CAA

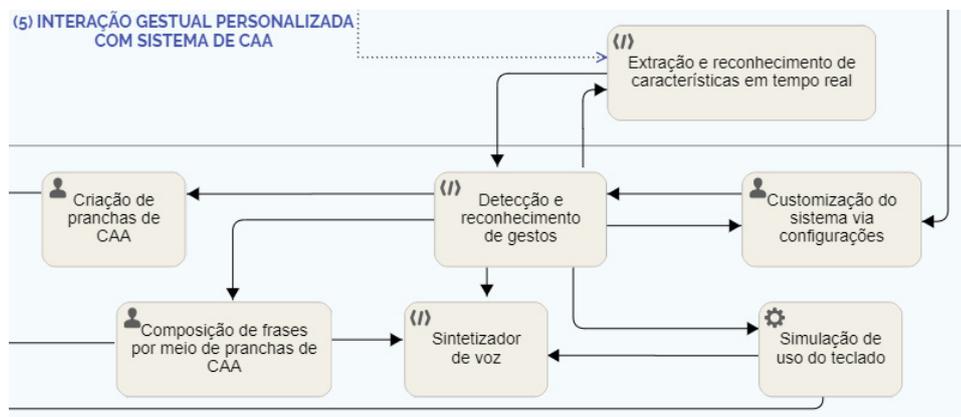


Figura 4.12: Estrutura definida na metodologia MyPGI para a Etapa 5 (Interação gestual personalizada com sistema de CAA).

O PGCA, sistema de CAA desenvolvido seguindo os pressupostos da metodologia MyPGI (Etapa 5 — Figura 4.12), possui interface composta por quatro áreas principais com finalidades específicas: 1. Área do Cuidador, onde o conjunto de gestos personalizado é criado; 2. Área do Usuário, onde o reconhecimento de gestos é usado para interagir com o sistema e com pranchas de comunicação; 3. Área de Pranchas de Comunicação, onde novas pranchas de comunicação de figuras podem ser geradas selecionando imagens quaisquer; e 4. Área de Jogo, onde o usuário pode interagir com a personagem de uma tartaruga movimentando-a por meio de gestos com o objetivo de levá-la até o mar.

No canto superior esquerdo da tela, existe um ícone de alto-falante, que, ao ser ativado, permite que um sintetizador de voz reproduza todas as mensagens apresentadas textualmente pelo sistema, com o objetivo de facilitar o entendimento para usuários ainda não alfabetizados. Em vários pontos do sistema, há um ícone de interrogação relacionado à função de ajuda, que descreve o funcionamento básico da área em questão. A interface do sistema possui uma borda em torno das guias principais, que é apresentada na cor amarela sempre que o sistema perceber alguma movimentação significativa feita em frente à câmera. Essa borda é apresentada na cor verde, sempre que um movimento é finalizado e o gesto correspondente é reconhecido pelo sistema.

Um breve resumo das funcionalidades disponíveis nas quatro áreas principais do sistema PGCA é apresentado na sequência.

4.3.5.1 Área do Cuidador

Na Área do Cuidador, os conjuntos de dados (compostos por gestos personalizados) são criados. Os gestos executados pelo usuário são associados a classes rotuladas com palavras que serão usadas para fins de comunicação ou interação (por exemplo, Oi, Tchau, Banheiro, Comida, Água, Confirmar, Desfazer). A Figura 4.13 apresenta a Área do Cuidador, exemplificando as telas apresentadas pelo sistema no momento da captura de amostras para compor o conjunto de dados, e também após a etapa de treinamento e avaliação do sistema.

Ao iniciar o sistema, o cuidador deve posicionar o usuário em frente ao computador e auxiliá-lo na realização do processo de calibração, que consiste em posicionar o usuário de forma centralizada em relação à câmera. Para isso, o sistema espera que a ponta do nariz do usuário passe em algum momento pelo centro da tela.

Após o requisito da etapa de calibração ser atendido, o sistema habilita a representação de movimento na forma de imagem MHI ou OF-MHI (indicado por meio de configuração), e então é possível iniciar a captura de amostras para representar cada gesto que irá compor o conjunto de dados do usuário em questão. Clicando no botão disponível na parte inferior direita da Área do Cuidador, pode-se configurar o sistema indicando o classificador, representação de movimento, método de avaliação e nível de confiança que serão utilizados, e se o sistema deve validar as amostras capturadas.

Para auxiliar na garantia de qualidade das amostras capturadas, o sistema PGCA utiliza um algoritmo de similaridade entre imagens. O primeiro movimento capturado para representar um determinado gesto é reproduzido para o usuário e então é solicitada uma confirmação de que o movimento realizado está correto. Após a confirmação do usuário/cuidador, o primeiro gesto é salvo como amostra base e será usado como referência para comparar os próximos movimentos que serão capturados para representar o mesmo gesto. A comparação das amostras capturadas é feita a partir de uma operação de subtração executada entre as imagens (MHI ou OF-MHI) correspondentes, e a área restante é analisada verificando se não é maior que a área original da amostra base. Além disso, o centroide de cada uma das imagens também é comparado para verificar se eles estão no mesmo quadrante na imagem ou em quadrante adjacente. Assim, sempre que o sistema identificar uma amostra muito diferente da amostra base (erroneamente gerada durante a realização de algum movimento involuntário), o usuário será informado de que a amostra foi considerada diferente e serão apresentadas as opções de descartá-la ou mantê-la, como pode-se visualizar na Figura 4.13 (B). A opção de configuração "Validar movimentos" disponível na Área do Cuidador permite que o usuário informe se deseja ou não usar esse recurso de validação de amostra.

O cuidador deve informar o nome do usuário, e na sequência, clicar no botão "Iniciar captura". Para cada gesto que será capturado o cuidador deve informar um rótulo/nome, e usar o botão disponível ao lado da caixa de texto para relacionar uma imagem ao gesto em questão. Na sequência, deve clicar no botão "Iniciar detecção automática de gesto" e solicitar ao usuário que realize o movimento correspondente ao rótulo informado. O sistema automaticamente identificará que o gesto foi concluído quando o usuário finalizar o movimento, ou seja, parar de se movimentar. E assim, o cuidador pode auxiliar o usuário a capturar várias amostras do mesmo gesto, sendo que o sistema prevê o mínimo de dez amostras disponíveis para cada gesto. Esse procedimento pode ser repetido para todos os gestos que usuário e cuidador entendam que sejam importantes constar no conjunto de dados. O número de amostras disponíveis para cada gesto deve ser padronizado, ou seja, se o conjunto de dados for composto por cinco gestos distintos, cada um dos cinco gestos precisa ter o mesmo número de amostras capturadas.

Depois de concluir a geração do conjunto de dados, o cuidador pode iniciar o treinamento do sistema clicando no botão correspondente para isso (Treinar sistema). Posteriormente, o

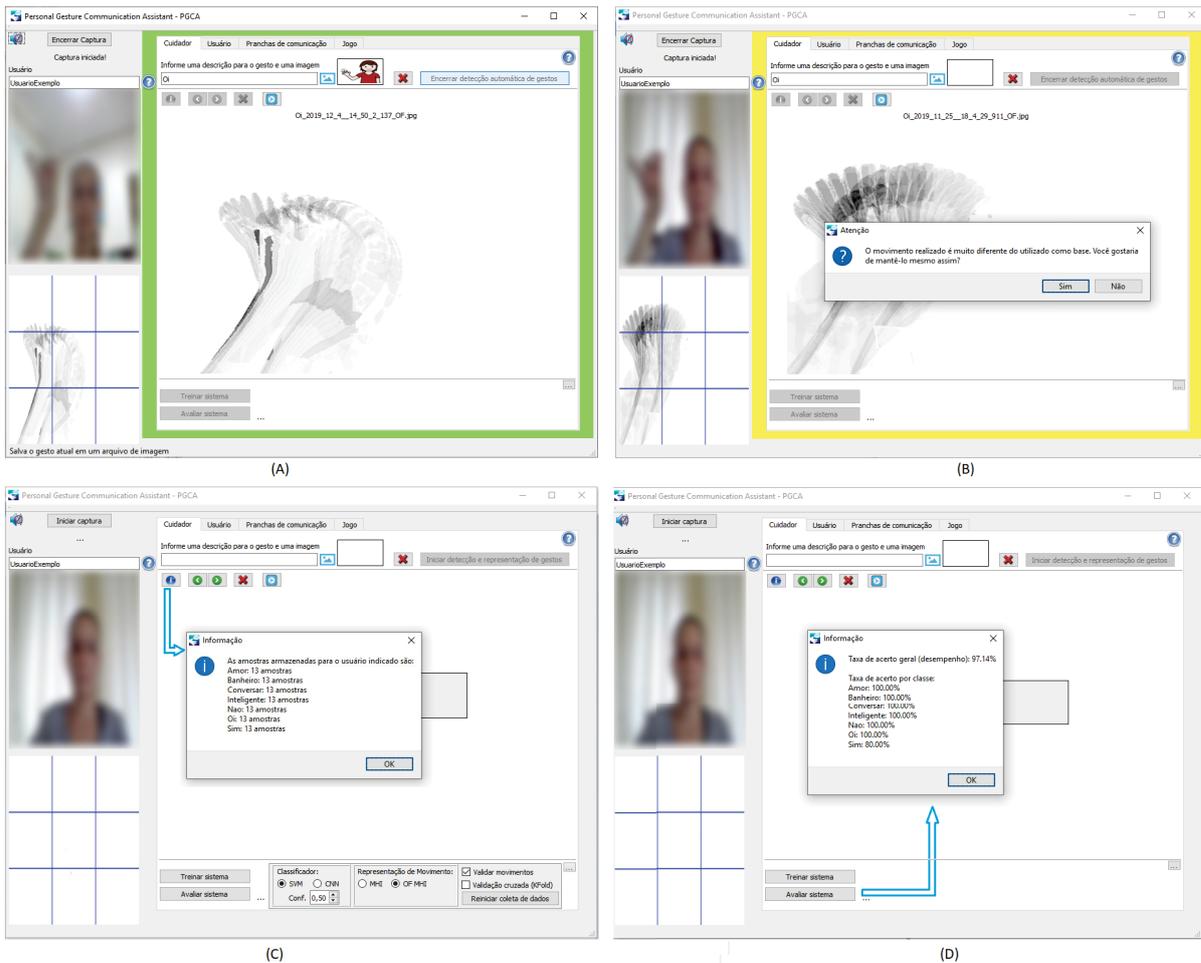


Figura 4.13: Área do Cuidador onde conjuntos de dados são criados e o sistema é treinado para o reconhecimento de gestos. (A) Exemplo de captura de uma amostra de gesto. (B) Exemplo de mensagem apresentada ao usuário após análise da amostra por algoritmo de similaridade. (C) Mensagem informativa sobre o número de amostras capturadas por gesto. (D) Apresentação do desempenho do sistema após executar a etapa de treinamento e posterior avaliação do sistema.

cuidador pode usar o botão "Avaliar sistema" para visualizar o desempenho do sistema no reconhecimento dos gestos capturados.

4.3.5.2 Área do Usuário

A Área do Usuário deve ser utilizada após a conclusão da etapa de treinamento do sistema. Essa área representa a interface principal com funções de CAA em que a interação gestual poderá ser utilizada para executar as seguintes funções: Ajuste de configurações; Detecção e representação de gestos; Navegação em pranchas de comunicação; Seleção de imagens; e Simulação do uso do teclado.

A Figura 4.14 mostra um exemplo de uso da Área do Usuário onde a execução de um gesto é reconhecida pelo sistema. Ao reconhecer um gesto, o sistema apresenta a identificação desse gesto (Rótulo) em uma barra de progresso, a figura correspondente (quando indicado na criação das amostras) é apresentada ao usuário, o nome da classe é inserido na caixa de texto, a borda da interface é apresentada na cor verde e o som correspondente é reproduzido por um sintetizador de voz.

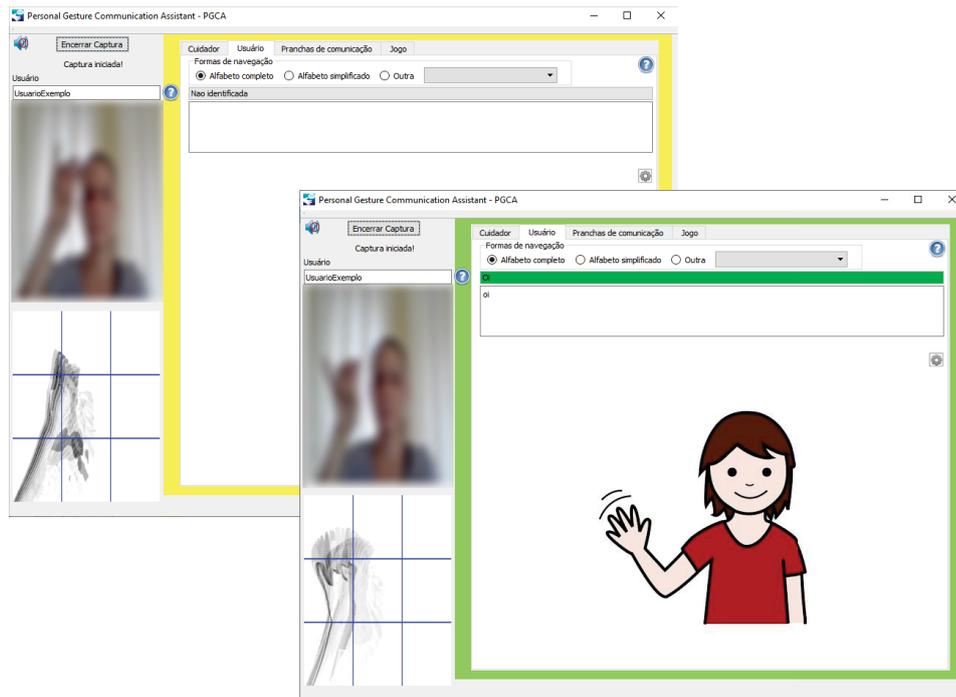


Figura 4.14: Exemplo de uso da Área do Usuário para o reconhecimento de gestos. A borda amarela indica que o gesto está em execução, e a borda verde indica que o gesto foi concluído e reconhecido pelo sistema. O rótulo associado ao gesto reconhecido é apresentado em uma barra de progresso, a figura correspondente é apresentada ao usuário, o nome da classe é inserido na caixa de texto, e o som correspondente é reproduzido por um sintetizador de voz.

Para customizar o sistema, o cuidador deve acessar a tela de configurações do usuário (Figura 4.15), que é oculta por padrão e exibida quando o usuário acessa a opção representada por um ícone de "engrenagem". Nas configurações o cuidador deve indicar como o usuário irá interagir com as funcionalidades do sistema.

A opção “Navegação automática” permite indicar como ocorrerá a interação com as pranchas de comunicação, se por meio de gestos ou tempo (segundos).

A opção “Sobrescrever textos” permite que o usuário sobrescreva ou não os textos já existentes na caixa de texto em que os rótulos dos gestos reconhecidos são inseridos. Para a composição de frases, essa opção deve ficar desmarcada.

A opção “Sensibilidade de captura” indica o tamanho mínimo (em *pixels*) da representação de movimento a ser considerada para que o sistema entenda que um gesto está sendo executado. Por padrão o sistema sugere valor 1000 para sensibilidade. Quando a opção "Aproximar" é marcada, o sistema aproxima a área de captura da câmera, com o objetivo de destacar e melhorar a percepção dos movimentos faciais.

A opção “Simular uso do teclado”, quando marcada, faz com que a inserção de rótulos (textos) ocorra no aplicativo (sistema computacional) que estiver em uso, simulando a digitação de caracteres ou comandos diretamente no teclado físico do computador. Dessa forma, é possível utilizar o reconhecimento de gestos, ou a seleção de opções em pranchas de comunicação para inserir dados na interface do sistema PGCA, em editor de texto, navegador de Internet ou outros aplicativos.

A opção “Gatilho duplo” indica se é necessário o usuário realizar cada gesto duas vezes para que uma ação seja iniciada. Por padrão essa opção é mantida desmarcada, mas pode ser útil em situações que o usuário tem dificuldade em realizar início e fim de movimentos de forma clara, gerando movimentos desnecessários que não precisariam ser interpretados pelo sistema.

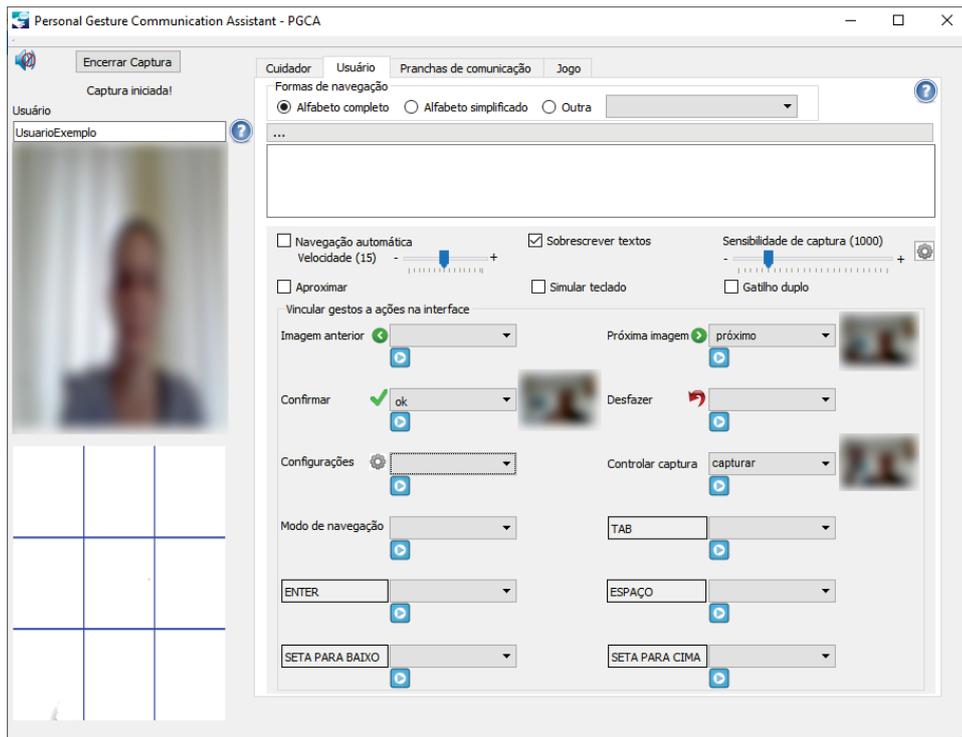


Figura 4.15: Tela de configurações da Área do Usuário, onde é possível customizar o funcionamento do sistema PGCA, relacionar gestos com a execução de gatilhos no sistema e com a simulação de uso de algumas teclas do teclado. Esta tela é oculta por padrão e exibida quando o usuário acessa a opção representada por um ícone de "engrenagem".

É possível relacionar os gestos utilizados para treinar o sistema com funcionalidades específicas de interação com o sistema PGCA, e com teclas do teclado. Quando o usuário relaciona um gesto a uma determinada funcionalidade do sistema, o movimento do gesto selecionado é reproduzido ao lado da caixa de seleção. Abaixo de cada caixa de seleção, há também um botão disponível para reproduzir o vídeo referente ao movimento que deu origem à amostra utilizada como base para o gesto selecionado.

O processo de configuração básica pode ser feito com o suporte do cuidador, e uma vez que o usuário conhece os gestos básicos relacionados às ações definidas nas configurações, o próprio usuário pode acessar a interface de configurações e personalizar o sistema usando gestos.

O sistema apresenta algumas opções de pranchas de comunicação por padrão, variando entre uma prancha de comunicação alfabética completa, simplificada, ou de figuras diversas, em que é possível realizar seleções em vários níveis de agrupamento. Na parte superior da Área do Usuário pode-se informar qual prancha de comunicação será utilizada. Além dessas pranchas padrão, marcando a opção outras, o usuário pode também utilizar as pranchas de comunicação criadas pelo cuidador na Área de Pranchas de Comunicação. A Figura 4.16 mostra um exemplo de navegação em imagens de uma prancha de comunicação de figuras disponibilizada pelo sistema como exemplo (composta por imagens disponíveis no portal ARASAAC (Rodrigo e Corral, 2013)).

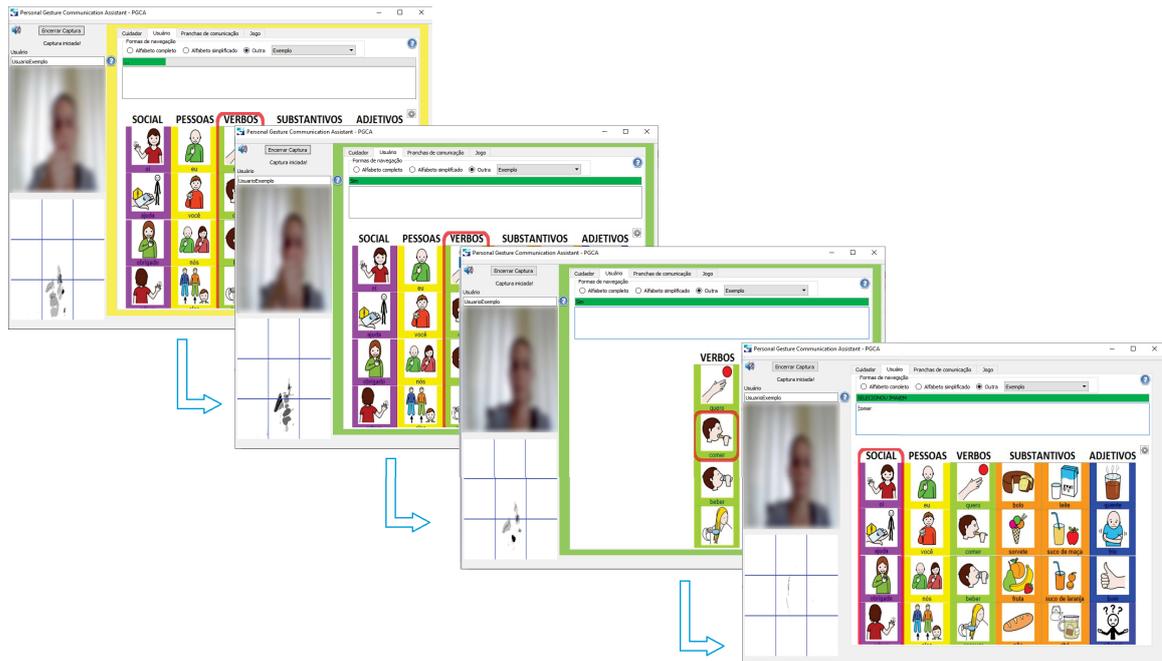


Figura 4.16: Exemplo de uso da Área do Usuário para a seleção de figuras em prancha de comunicação apresentada por padrão pelo sistema utilizando interação gestual. Ao selecionar uma opção no primeiro agrupamento apresentado pela prancha de comunicação, a navegação ocorre em um agrupamento menor, e então o usuário pode selecionar a figura referente à palavra que deseja escrever e o som correspondente é emitido.

O *layout* das pranchas alfabéticas apresentadas pelo sistema foi construído com base em um estudo desenvolvido por Liegel et al. (2008), como resultado de pesquisa feita com profissionais da área de Comunicação Alternativa. A Figura 4.17 apresenta o *layout* original, usado como base, e dois *layouts* adaptados criados a partir dele. A navegação nas pranchas alfabéticas foi inspirada no uso de *LetterBoards* que apresentam letras agrupadas e categorizadas por cores. Já o agrupamento utilizado na prancha de figuras (Figura 4.16) foi criado a partir de exemplos disponíveis em Sartoretto e Bersch (2014) como uma sugestão inicial, e pode ser ampliado e ajustado de acordo com as prioridades de cada usuário, desde que representadas na forma de imagens.

A partir da navegação entre agrupamentos específicos de cada prancha de comunicação apresentada por padrão pelo sistema, é possível selecionar uma letra ou figura, e, então, a descrição relacionada é inserida na caixa de texto, emitindo o som correspondente à palavra. A inclusão de palavras por meio de seleção também respeita as configurações do sistema para sobrescrever ou não o texto existente.

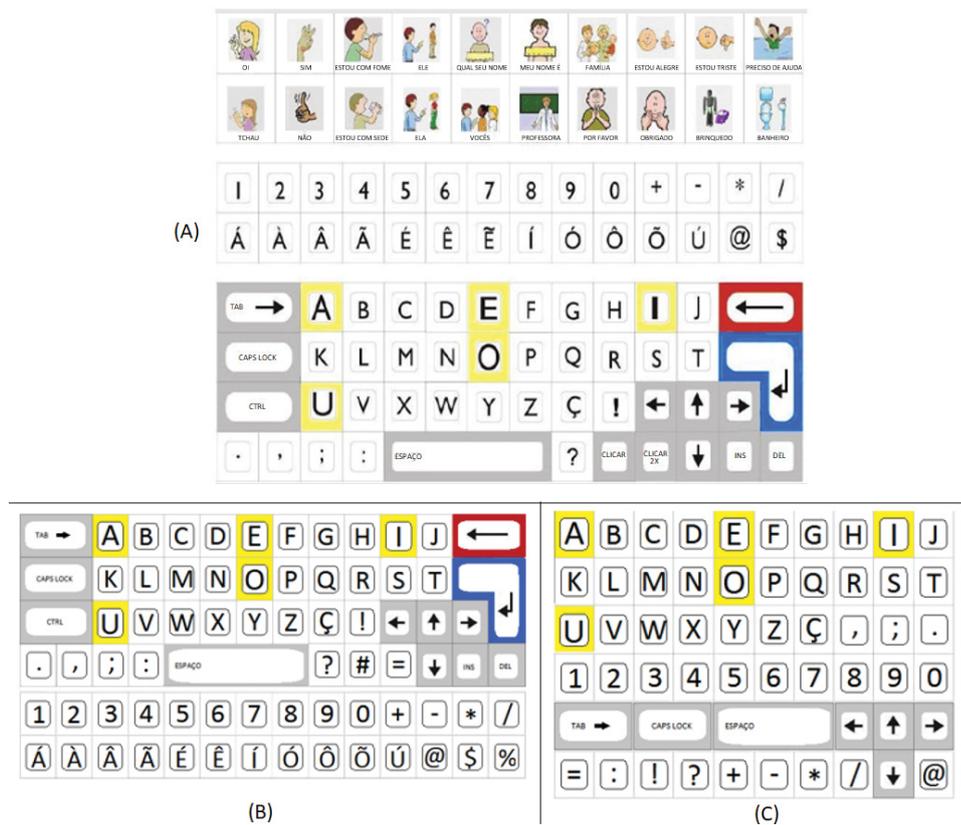


Figura 4.17: *Layout* de pranchas alfabéticas para CAA: (A) *Layout* utilizado como base para criação de pranchas adaptadas; (B) *Layout* de prancha alfabética adaptada completa, contendo inclusive caracteres especiais e comandos diversos; (C) *Layout* de prancha alfabética adaptada simplificada, contendo letras, números e comandos básicos.

Fonte: Liegel et al. (2008).

4.3.5.3 Área de Pranchas de Comunicação

A Figura 4.18 mostra a Área de Pranchas de Comunicação, onde o cuidador pode criar diferentes pranchas de comunicação simples compostas por imagens. A prancha criada deve ser identificada por uma descrição, assim como cada imagem utilizada para compô-la. A descrição utilizada para identificar cada imagem refere-se ao rótulo que será apresentado ao usuário quando esta imagem for selecionada via gestos na prancha de comunicação.

Posteriormente, na Área do Usuário é possível selecionar as pranchas de comunicação criadas pelo cuidador na Área de Pranchas de Comunicação, a partir da caixa de seleção apresentada ao lado da opção "Outra".

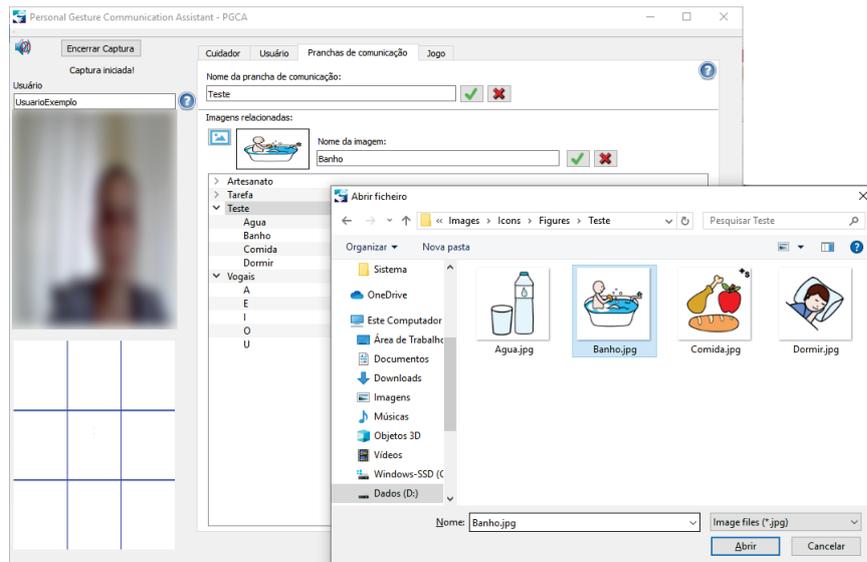


Figura 4.18: Área de Pranchas de Comunicação onde o cuidador pode gerar pranchas personalizadas usando imagens.

4.3.5.4 Área de Jogo

A Área de Jogo foi incluída com o objetivo de motivar ou incentivar o interesse dos usuários a interagir com o sistema. Considerando que os usuários do sistema PGCA podem apresentar habilidades motoras bastante restritas, foi criado um jogo simples, claro e objetivo que requer um pequeno número de gestos para seu uso. Assim, a tartaruga, que é a personagem principal do jogo, pode ser movida em quatro direções, mas com apenas uma ou duas direções já é possível atingir o objetivo do jogo.

Usando gestos, o usuário deve evitar os obstáculos na areia (uma concha e dois caranguejos) e levar a tartaruga para a região superior da imagem, onde fica o mar. Passar por qualquer um dos obstáculos não impede que a tartaruga se mova, apenas apresenta um som e imagem diferentes. Não há pontuações ou restrições de tempo de qualquer tipo, porque o jogo foi projetado principalmente para incentivar os usuários a realizar gestos repetidamente.

Para mover a tartaruga, é possível relacionar até quatro gestos com os movimentos de deslocamento para cima, para baixo, para a direita e para a esquerda. No topo da Área de Jogo, existem listas de seleção em que os gestos disponíveis (usados para treinar o sistema para o usuário em questão) podem ser relacionados aos movimentos. Para iniciar o jogo, a captura do sistema deve ser iniciada (botão "Iniciar captura"). Ao clicar no botão "Iniciar jogo", a tartaruga é posicionada no canto inferior esquerdo da tela, conforme apresentado na Figura 4.19 (A).

A Figura 4.19 (A) mostra a área de jogo e o estado inicial da tartaruga posicionada na praia. A Figura 4.19 (B) mostra um exemplo da mudança do estado da tartaruga durante o jogo, onde o movimento "para a direita" foi reconhecido e fez com que a tartaruga encontrasse o obstáculo do caranguejo. A borda verde indica que um gesto foi concluído e reconhecido pelo sistema. A Área de Jogo está disponível para uso após a execução do treinamento do sistema para o conjunto de dados criado.

A Tabela 4.1 apresenta a tartaruga em seu estado normal e, em seguida, em seu novo estado após tocar na concha ou em um dos caranguejos e depois de atingir a meta do jogo. Cada mudança no estado da tartaruga produz um som distinto. Um som de "salto" é emitido após o reconhecimento de um gesto associado a alguma execução de movimento. A Figura 4.20

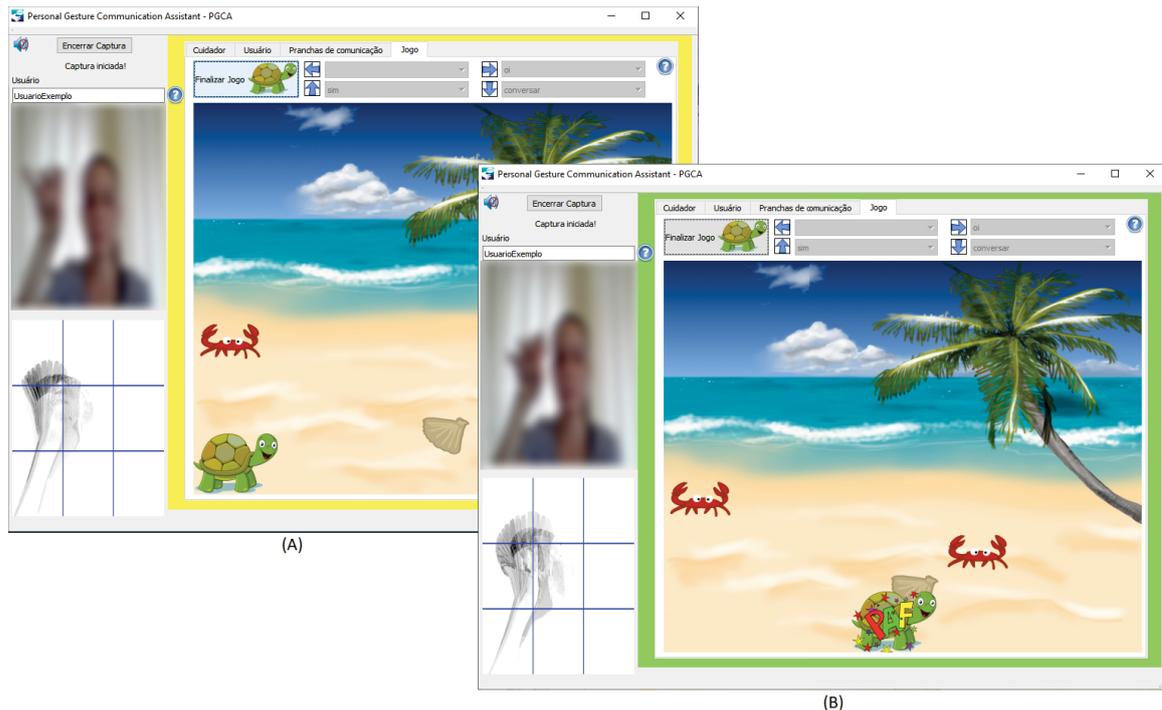
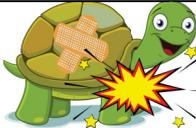


Figura 4.19: Área de Jogo, onde o usuário pode utilizar gestos para movimentar uma tartaruga com o objetivo de leva-la até o mar. (A) Tela inicial do jogo, a borda amarela indica que um movimento está sendo executado; (B) Exemplo de uso do jogo e alteração de estado da tartaruga, a borda verde indica que um gesto foi detectado e reconhecido.

mostra a imagem apresentada ao usuário após a conclusão do jogo, incluindo balões e confetes na paisagem da praia. Nesse caso, um som de "aplausos" também é emitido.

Tabela 4.1: Mudanças no estado da personagem tartaruga apresentadas durante a utilização do jogo desenvolvido.

Estado A Normal	Estado B Tocando no obstáculo da concha	Estado C Tocando nos obstáculos de caranguejos	Estado D Chegando ao mar
			

O funcionamento do jogo é descrito ao clicar no ícone de "?" disponível no canto superior direito da Área de Jogo. As imagens usadas no jogo foram obtidas pesquisando na Internet por conteúdo gratuito. A imagem da tartaruga foi obtida da Commons_Wikimedia (Commons, 2017) — repositório de mídia gratuito — e a imagem de fundo (já com a concha e um dos caranguejos) foi obtida do OpenGameArt (Day, 2015) — uma comunidade de desenvolvedores de jogos, que disponibiliza conteúdos gratuitos. As outras imagens usadas para fazer modificações nas imagens originais foram obtidas pesquisando imagens marcadas para reutilização no site Google Imagens (Google, 2019).

4.3.6 Considerações sobre as tecnologias empregadas

Em muitas situações, imagens com desfoque de movimento (um tipo comum de degradação causado pelo movimento relativo entre um objeto e a câmera) podem fornecer

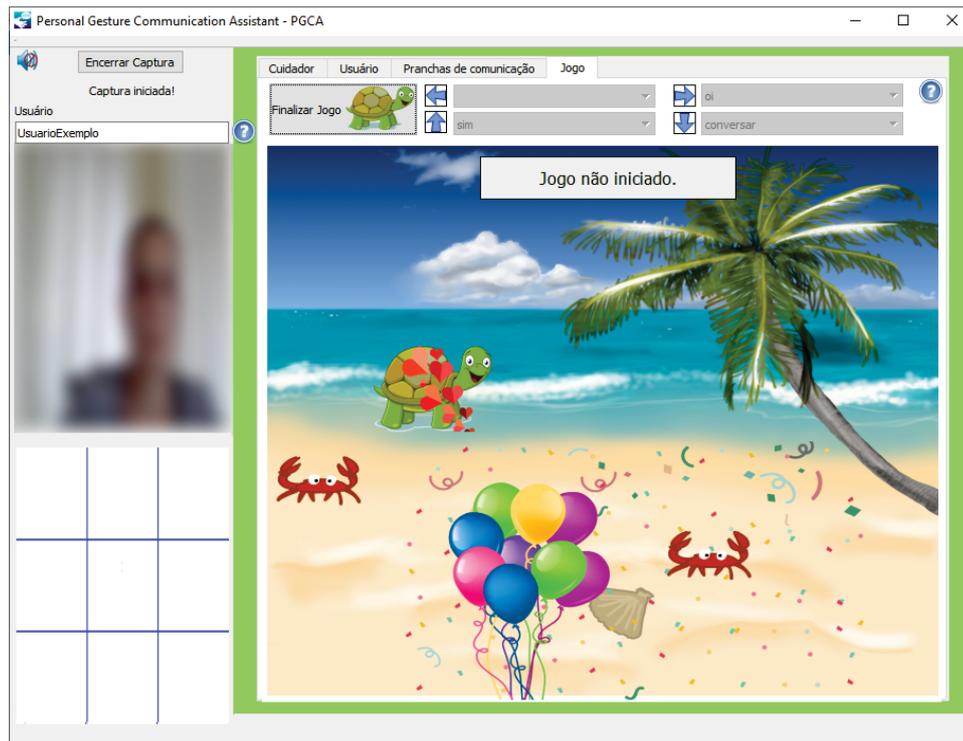


Figura 4.20: Tela apresentada ao usuário após a conclusão do objetivo do jogo, que é levar a tartaruga até o mar. Uma imagem com balões e confetes é apresentada ao usuário, o estado da tartaruga é alterado, e um som de aplausos é reproduzido.

resultados imprecisos, mas no sistema PGCA, como o movimento já é representado na forma de uma imagem desfocada composta por tons de cinza (MHI ou OF-MHI), o desfoque de movimento não interferiu significativamente na representação final do movimento, e não prejudicou o desempenho dos classificadores.

Oclusões podem prejudicar a compreensão de um gesto se interferirem significativamente na representação final do movimento gerado. Para utilização do sistema PGCA, o plano de fundo da cena deve ser estático, pois qualquer objeto ou pessoa que possa se mover atrás do usuário gerará representações de movimento imprecisas ou desnecessárias. O sistema PGCA demanda o uso de ambientes com fundo estático para o correto reconhecimento de gestos, mas não é necessário um fundo completamente homogêneo.

As condições de iluminação em diferentes ambientes de coleta de dados podem interferir na capacidade do sistema de capturar corretamente os movimentos executados pelos usuários. Portanto, para que o sistema tenha um desempenho melhor, é importante manter o mesmo padrão de iluminação durante a criação do conjunto de dados e durante a interação com o sistema. Preferencialmente, o usuário deve estar posicionado em frente a uma fonte de luz, como lâmpada ou janela.

Com relação aos algoritmos empregados, optou-se por usar algoritmos bem estabelecidos, embora algumas abordagens não sejam novas. Foram empregados métodos disponíveis na biblioteca OpenCV, por ela oferecer alta eficiência computacional e uso simples das infraestruturas de VC e Aprendizado de Máquina (Dall'Asta e Roncella, 2014). O uso de imagens MHI apresenta vantagens relacionadas à simplicidade, robustez na representação de movimento e baixa computação. O fluxo óptico de Lucas-Kanade possui cálculos muito rápidos e derivadas precisas de tempo (Noaman et al., 2017), e provou ser eficaz na agregação de informações de

velocidade do movimento nas imagens MHI, como observado nos trabalhos de Tsai et al. (2015); Fan e Tjahjadi (2017); Khalifa et al. (2018).

A escolha dos classificadores empregados se deu em função de bons resultados apresentados na literatura para a tarefa de classificação de imagens. Classificadores com base em SVM tem a vantagem de oferecer uma forte capacidade de generalização, arquitetura simples, bem como a capacidade de classificar poucas amostras (Liu et al., 2018; Mao et al., 2008). Classificadores com base em CNN mostraram recentemente popularidade explosiva, em parte devido ao seu sucesso na classificação de imagens e em outros campos da Visão Computacional (Dong et al., 2015; Wang et al., 2018). A Transferência de Aprendizado pode tirar proveito do conhecimento obtido de uma CNN profunda pré-treinada com um grande conjunto de dados para uma tarefa específica, melhorando o desempenho do reconhecimento de gestos (tarefa desta pesquisa) com um pequeno conjunto de dados, composto por um número restrito de amostras (contexto desta pesquisa).

4.4 ETAPAS DE AVALIAÇÃO

A natureza construtiva desta pesquisa requer uma estratégia progressiva e incremental, na qual o progresso é avaliado e informa as etapas adicionais de pesquisa e desenvolvimento. Assim, a metodologia MyPGI foi desenvolvida e revista com base nos resultados de experimentos conduzidos com diferentes objetivos, empregando o sistema PGCA desde a sua primeira versão.

O primeiro experimento foi conduzido com o objetivo de avaliar o protótipo do sistema PGCA em condições de laboratório, por profissionais que atuam com IHC (Experimento 1), utilizando um conjunto de gestos predefinidos para execução de algumas tarefas. Esse experimento permitiu identificar problemas de usabilidade e acessibilidade no sistema, bem como validar os requisitos do sistema antes da experimentação com voluntários.

Os resultados do primeiro experimento indicaram melhorias necessárias no sistema e então a segunda versão do sistema PGCA foi utilizada em um experimento realizado com voluntários sem deficiência motora e de fala (Experimento 2), cujo objetivo principal foi avaliar se o sistema seria capaz de reconhecer gestos criados de forma personalizada, sendo treinado com poucas amostras.

Posteriormente, com o objetivo de identificar as melhores estratégias para uma nova versão do sistema PGCA, dois classificadores de imagens e duas representações de movimento foram avaliados de acordo com seu desempenho no reconhecimento de gestos. Esse experimento (Experimento 3) foi conduzido usando o *Keck Gesture Dataset*, um conjunto de dados público disponível em Jiang et al. (2012). O uso de um conjunto de dados público teve como objetivo permitir a repetibilidade do experimento por outros pesquisadores, além de possibilitar a realização de testes com diferentes tecnologias antes que o sistema fosse disponibilizado para testes com o público-alvo.

Com o objetivo de conhecer melhor o público-alvo da proposta desse trabalho, visitas foram realizadas em escolas vinculadas a instituições coparticipantes (indicadas no projeto de pesquisa vinculado a esta tese de doutorado), onde existiam alunos com dificuldades motora e de fala matriculados. Nessas escolas, foram realizadas entrevistas com profissionais da área de educação especial (Entrevistas com especialistas do domínio) que compartilharam informações sobre a realidade dos alunos com dificuldades motora e de fala no ambiente escolar e iniciativas utilizadas para viabilizar a CAA. Os profissionais também indicaram alunos que seriam candidatos potenciais a participar dos experimentos seguintes.

Finalmente, após a análise dos resultados obtidos nos experimentos anteriores e as informações obtidas por meio de entrevistas com profissionais da área de educação especial, um

primeiro experimento com o público-alvo foi realizado (Experimento 4). O principal objetivo desse experimento foi verificar se o sistema PGCA apoiaria o público-alvo a gerar um conjunto de gestos personalizados, e se esses gestos poderiam ser utilizados para treinar um sistema no seu reconhecimento e posterior uso com propósitos de comunicação. Esse experimento permitiu avaliar a viabilidade técnica da proposta, e trouxe informações valiosas, mostrando um pouco da rotina dos alunos com deficiência motora e de fala no ambiente escolar.

Com o objetivo de revalidar algumas informações obtidas no Experimento 4, e aumentar o engajamento e motivação dos alunos para realizar as tarefas propostas, uma abordagem baseada em jogo interativo por gestos foi incluída no sistema PGCA e avaliada em um segundo experimento conduzido com o público-alvo (Experimento 5). Após a condução desse experimento, profissionais da área de educação especial que acompanharam a realização das tarefas pelos alunos preencheram um formulário de avaliação de usabilidade do sistema. Os resultados desse experimento permitiram analisar de maneira padronizada a precisão do sistema no reconhecimento de gestos e o esforço necessário para que cada usuário alcançasse o mesmo objetivo, cada um usando os gestos que é capaz de executar.

Detalhamentos sobre os resultados obtidos a partir de cada um dos experimentos conduzidos e entrevistas são apresentados no Capítulo 5.

Após a conclusão dos experimentos previstos para avaliação, a metodologia MyPGI foi utilizada para projetar uma solução em um contexto diferente do ambiente escolar, e foi avaliada também por um pesquisador externo.

4.5 CONSIDERAÇÕES

Este capítulo apresentou a metodologia MyPGI desenvolvida como resultado desta tese de doutorado, detalhando sua estrutura, características e possíveis cenários de aplicação. A metodologia foi utilizada como base para o desenvolvimento de um sistema inteligente criado para viabilizar a CAA usando interação gestual personalizada, chamado PGCA. O sistema desenvolvido foi descrito, detalhando a implementação de cada etapa prevista na metodologia, e também a forma de avaliação empregada para avaliação do sistema e da metodologia.

5 RESULTADOS

Este capítulo descreve os resultados obtidos a partir de diferentes experimentos conduzidos com o objetivo de avaliar a proposta dessa tese de doutorado. Também são apresentados resultados de entrevistas e avaliação de usabilidade, conduzidas com profissionais da área de educação especial.

Como o público-alvo dessa pesquisa envolve grupos de usuários considerados vulneráveis, um projeto de pesquisa foi submetido à avaliação do Comitê de Ética em Pesquisa com Seres Humanos do Setor de Ciências da Saúde da Universidade Federal do Paraná. A aprovação do projeto (CAAE 81279317.6.0000.0102) pelo comitê deu condições legais para a realização de testes com usuários vinculados a três instituições coparticipantes: Universidade Tecnológica Federal do Paraná (UTFPR - Câmpus Pato Branco), Escola Carlos Almeida Educação Básica na Modalidade Educação Especial (APAE Pato Branco) e escolas da rede estadual de educação vinculadas ao Núcleo Regional de Educação do município de Pato Branco/PR por meio da Secretaria Estadual de Educação do Estado do Paraná (SEED/PR).

Cinco experimentos foram conduzidos com o objetivo de avaliar a viabilidade técnica da metodologia proposta: Experimento 1 - estudo de caso exploratório com profissionais que atuam com IHC; Experimento 2 - criação de conjuntos de gestos personalizados por usuários sem dificuldades motoras e de fala; Experimento 3 - uso de um conjunto de dados público para avaliar o desempenho do sistema no reconhecimento dos gestos; Experimento 4 - uso do sistema por usuários com dificuldades motora e de fala; Experimento 5 - uso do sistema e também de um jogo por usuários com dificuldades motora e de fala, e uma avaliação de usabilidade do sistema. Além disso, entrevistas com profissionais da área de educação especial (especialistas de domínio) foram realizadas com o objetivo de conhecer melhor a realidade e características do público-alvo dessa pesquisa. Uma avaliação da metodologia em outro contexto também foi realizada. Todos os participantes dos experimentos assinaram um Termo de Consentimento Livre e Esclarecido (TCLE), e os responsáveis pelos alunos participantes assinaram um Termo de Assentimento Livre e Esclarecido (TALE).

A Tabela 5.1 resume as etapas de avaliação realizadas, estruturadas por identificação, características dos participantes, quantidade de participantes, objetivo e resultados. Detalhes dos resultados obtidos em cada avaliação são descritos na sequência.

Tabela 5.1: Resumo das etapas de avaliação realizadas para avaliar o sistema PGCA e consequentemente a metodologia MyPGI.

Identificação	Participantes	Quant.	Objetivo	Resultados
Experimento 1	Profissionais da área de IHC	12	Identificar problemas de usabilidade e acessibilidade no protótipo do sistema PGCA, bem como validar requisitos.	Identificação de não conformidades, melhorias e requisitos de sistema que precisavam ser refinados.
Experimento 2	Voluntários sem deficiência motora e de fala	5	Avaliar se o sistema seria capaz de reconhecer gestos criados de forma personalizada, sendo treinado com poucas amostras.	Capacidade do sistema em atender os pressupostos da metodologia. Permitiram identificar algumas melhorias necessárias ao sistema PGCA e à metodologia MyPGI.
Experimento 3	<i>Keck Gesture Dataset</i>	-	Permitir a repetibilidade do experimento por outros pesquisadores e realizar testes com diferentes tecnologias antes de disponibilizar o sistema para testes com o público-alvo.	Viabilidade das tecnologias implementadas (representações de movimento MHI e OF-MHI, classificadores com base em SVM e CNN).
Entrevistas	Profissionais da área de educação especial	8	Obter informações sobre a realidade dos alunos com dificuldades motora e de fala no ambiente escolar e iniciativas utilizadas para viabilizar a CAA.	Identificação de alguns desafios e possibilidades de novas melhorias no sistema. Identificação de alunos que seriam candidatos potenciais a participar dos experimentos seguintes.
Experimento 4	Alunos com deficiência motora e de fala	7	Verificar se o sistema PGCA apoiaria o público-alvo a gerar conjuntos de gestos personalizados, e se esses gestos poderiam ser utilizados para treinar um sistema e com propósitos de comunicação.	Avaliação da viabilidade técnica do sistema e da metodologia, e obter informações sobre a rotina dos alunos com deficiência motora e de fala no ambiente escolar e iniciativas utilizadas pelos professores.
Experimento 5	Alunos com deficiência motora e de fala	3	Revalidar algumas informações obtidas no Experimento 4, avaliar uma abordagem baseada em jogo interativo por gestos e avaliar a usabilidade do sistema.	Análise da precisão do sistema no reconhecimento de gestos e o esforço necessário para que cada usuário alcançasse o mesmo objetivo. Usabilidade do sistema foi considerada aceitável.
Outro Contexto	Pesquisador externo	1	Avaliar a metodologia considerando outro contexto e a percepção de outra pessoa.	Capacidade de replicabilidade da metodologia MyPGI considerando outros problemas, outros usuários e outros profissionais.

5.1 EXPERIMENTO 1: AVALIAÇÃO EXPLORATÓRIA DO SISTEMA POR PROFISSIONAIS QUE ATUAM COM IHC

Seguindo os pressupostos da metodologia proposta, um sistema piloto foi desenvolvido para permitir a criação de interação gestual personalizada com um sistema computacional de CAA. Nesse primeiro experimento, o sistema, que posteriormente foi chamado de PGCA, possuía apenas a implementação de MHI convencional para representação de movimento, e do classificador com base em SVM para reconhecimento de gestos.

No contexto de TA, torna-se crítico desenvolver sistemas pilotos e conduzir experimentos com especialistas antes de qualquer atividade envolvendo o público-alvo da solução¹. Caso contrário, corre-se o risco de submeter os participantes a erros e dificuldades que poderiam ter sido antecipados em análises em laboratório e com especialistas. Portanto, a primeira versão do sistema PGCA foi avaliada por profissionais que atuam com IHC a fim de identificar problemas de usabilidade e acessibilidade, melhorias e requisitos de sistema que precisavam ser refinados antes do sistema ser considerado adequado para experimentação com seu público-alvo. Participaram desse experimento alunos e professores vinculados ao Programa de Pós-Graduação em Informática da UFPR, envolvidos em pesquisas vinculadas à área de IHC e/ou acessibilidade.

Doze pessoas participaram da avaliação, sendo nove avaliadores profissionais que atuam com IHC (três duplas e um trio), dois observadores especialistas do domínio e um mediador. Após demonstração do sistema pela autora desta tese, os avaliadores exploraram o sistema individualmente para executar duas tarefas predeterminadas em uma dinâmica inspirada na técnica de *Thinking Aloud* (Nielsen, 1994) (protocolo verbal que consiste em observar o usuário utilizar uma interface enquanto verbaliza seus pensamentos), enquanto os observadores tomavam notas. As tarefas foram realizadas em explorações usando interação gestual previamente cadastrada pela pesquisadora, em função do tempo disponível não permitir que cada grupo criasse seu próprio conjunto de gestos. Em explorações que duraram cerca de cinco minutos, cada grupo deveria: (a) formar uma frase com três palavras usando os gestos predefinidos para descrever ações ou selecionando opções apresentadas pela prancha de comunicação de figuras; (b) escrever o próprio nome usando os gestos predefinidos para selecionar letras da prancha de comunicação alfabética, usando a navegação manual entre os agrupamentos ou a navegação automática. Após a exploração, os avaliadores elaboraram um relatório por dupla/trio com a indicação dos problemas encontrados e sugestões de melhorias. Foi utilizado um *template* de avaliação com as Heurísticas de Nielsen (Nielsen, 1994) e recomendações de avaliação simplificada de acessibilidade (Santana et al., 2018).

5.1.1 Resultado da avaliação

Como resultado da avaliação, três limitações técnicas e dez problemas diferentes foram apontados pelos avaliadores juntamente com mais de vinte sugestões de melhorias. Dessas sugestões, algumas foram implementadas, algumas foram consideradas não problemáticas, e outras foram registradas como novos requisitos. Os principais pontos considerados problemáticos ou com necessidade de melhorias foram trabalhados para gerar uma nova versão do sistema, que seria utilizada nos experimentos seguintes. Um dos pontos principais destacados pelos avaliadores foi a necessidade de reduzir ao máximo a mediação de um cuidador, aumentando a autonomia do usuário na configuração e utilização do sistema. As principais contribuições obtidas com a atividade de avaliação são apresentadas na sequência.

5.1.1.1 Sugestões implementadas

As sugestões listadas a seguir foram implementadas logo após a conclusão do experimento:

- apresentação da imagem do usuário de forma espelhada para que o usuário tenha a percepção real da direção dos movimentos realizados;

¹Testes e avaliações intensivos antes de avaliar um produto com o público-alvo são principalmente uma questão de ética, pois as pessoas com deficiência não podem ser tratadas como sujeitos de pesquisa.

- inclusão de quadrantes na imagem do usuário para destacar a visualização de movimentos em grade;
- realização de processo de calibração no sistema para ajuste do posicionamento inicial do usuário;
- possibilidade de aproximar a imagem, por meio de *zoom* para destacar e melhorar a percepção dos movimentos do rosto;
- uso de Aumento de Dados para aumentar o número de amostras no conjunto de dados, e gerar um conjunto de dados para testes automaticamente;
- uso de gestos para acessar as configurações do sistema, para iniciar e parar a captura de imagens, e para customizar o funcionamento do sistema por meio de seleção da prancha de comunicação a ser utilizada;
- melhoria na estética da interface;
- criação de uma prancha alfabética mais simples, sem caracteres especiais.

5.1.1.2 Problemas encontrados e corrigidos

Os problemas apontados pelos profissionais que foram corrigidos logo após a conclusão do experimento são apresentados a seguir:

- o estado do sistema é visível, mas pode ter uma sobrecarga cognitiva por apresentar muitas opções de interação, nem sempre muito coesos entre si no espaço que estão;
- a estética precisa ser melhorada/atualizada;
- *feedback* poderia ser melhor no momento de execução;
- faltam rótulos em algumas caixas (indicação para que serve cada coisa);
- difícil entender que tem erro (precisa ser mais chamativo);
- dificuldade em ver quando o sistema estava lendo gestos, sugere-se destacar campo de texto e barra de progresso;
- a imagem anterior continua aparecendo ao iniciar outro gesto;
- dar *zoom* ou maior destaque no agrupamento selecionado na prancha alfabética;
- pode não haver clareza nos símbolos nas ações de navegação;
- após identificar o primeiro movimento, os quadros continuam sendo trocados.

5.1.1.3 Situações consideradas não problemáticas

Algumas situações apontadas pelos profissionais como problemas, foram consideradas como não problemáticas pela autora desta tese, por se tratar de situações originadas possivelmente por instrução incompleta durante a demonstração do sistema. Essas situações são apresentadas a seguir:

- substituir o termo sobrescrever por palavra/frase (o termo sobrescrever foi considerado mais apropriado por estar vinculado à ação de substituir);
- usuário interagir sem mouse (na versão atual do sistema isso é possível após configuração inicial feita pelo cuidador);
- na apresentação automática os movimentos para os lados não estão configurados (não é obrigatória essa configuração pois a navegação nesse caso é baseada em tempo);
- não foi encontrado nenhuma mensagem (as mensagens são apresentadas na barra de *status* do sistema, essa informação deveria ter sido melhor destacada durante a demonstração);
- é preciso pensar em usuários que movimentem apenas um lado do corpo (o sistema já atende esses usuários).

5.1.1.4 Limitações técnicas identificadas

Algumas limitações técnicas foram identificadas durante o experimento e são apresentadas a seguir:

- a cor da roupa do usuário e diferenças bruscas de iluminação podem interferir no processo de captura dos movimentos dificultando o reconhecimento dos gestos pelo sistema;
- restringe muito a movimentos corporais repetidos de especificidade grande;
- não há aporte para pessoas com deficiência visual/auditiva.

5.1.1.5 Novos requisitos

Várias sugestões apresentadas pelos profissionais foram consideradas válidas, e foram classificadas como novos requisitos que poderiam ser avaliados para implementação futura no sistema PGCA. Essas sugestões são apresentadas a seguir:

- verificar se pode ser vantajoso trocar gatilho duplo por gesto "confirma" (ex.: Água + Confirma);
- possibilitar integração (uso em conjunto) com outras ferramentas como Head Mouse, etc.;
- deveria ser possível o acesso da própria interface pelos gestos;
- melhorar o acesso para pessoas com deficiência auditiva e visual;
- adaptar o sistema para uso de Libras (Língua Brasileira de Sinais);

- durante a composição de frases, remover letras que não fazem sentido na sequência;
- explorar interação multimodal, aliando movimento de cabeça e direção de olhar;
- utilizar o sistema ACAT (Denman et al., 2016) usado por Stephen Hawking para comunicação e interação com seu computador, como base, principalmente para montagem de frases;
- aliar movimentos com sons que o usuário seja capaz de emitir;
- permitir o uso de um acionador;
- usar o sistema em dispositivos móveis;
- apresentar por padrão um conjunto básico de ações predefinido;
- criar um *wizard* (*software* que guia os usuários por uma tarefa específica com uma série de instruções) para auxiliar no processo de treinamento sugerindo comandos;
- reconhecer erros em outros programas, quando utilizada a opção de simulação do teclado;
- permitir o uso de um navegador, por exemplo, sem perder a visualização do que o sistema captura;
- empregar autossugestão de palavras para tornar a composição de frases mais rápida.

5.1.2 Discussão

A atividade de avaliação com profissionais que atuam com IHC indicou melhorias necessárias para que um experimento em contexto real fosse possível e produtivo, ajudando a antecipar problemas que dificultariam o sistema de ser flexível e adaptável às características de cada usuário, ou mesmo que impediriam seu uso por pessoas com diferentes limitações. Com o sistema implementado e os resultados da avaliação realizada, foi possível verificar que a metodologia proposta é viável para apoiar o *design* de um sistema de CAA, e que, com as melhorias indicadas, o sistema desenvolvido teria potencial para ser utilizado em contextos reais por pessoas com deficiências.

Com relação a aspectos específicos da implementação, o uso de imagens MHI apresenta vantagens relacionadas a simplicidade, robustez na representação de movimentos e baixa computação. O HOG do MHI foi empregado visando aumentar a taxa de reconhecimento, mas exige que o ambiente de fundo seja estático, pois variações de iluminação e presença de movimentos da câmera ou ambiente desordenado podem dificultar a precisão no reconhecimento dos gestos.

O desempenho do sistema foi considerado satisfatório na execução de movimentos simples, principalmente devido ao desempenho apresentado pelo classificador no reconhecimento do conjunto de testes (91,25% usando o método Holdout). Somado a isso, vale destacar que o sistema foi capaz de reconhecer gestos executados pelos avaliadores na atividade de avaliação, mesmo tendo sido treinado com um conjunto de dados criado por outra pessoa (a autora desta tese).

Os resultados obtidos com esse experimento reforçaram a percepção de que uma metodologia voltada para o projeto de sistemas de CAA usando interação gestual personalizada é

viável e pode ser aplicada em um contexto assistivo, ampliando as possibilidades de pessoas com deficiência motora se comunicarem e interagirem com sistemas computacionais.

A partir desse primeiro experimento, uma nova versão do sistema passou a ser gerada, implementando as sugestões obtidas com a atividade de avaliação e que ainda não haviam sido contempladas, principalmente com relação a melhorias na interface do sistema. Além disso, para melhorar o reconhecimento de movimentos dinâmicos e mais complexos, foi considerado pertinente acrescentar no sistema o uso de informações relacionadas à velocidade dos movimentos realizados.

5.2 EXPERIMENTO 2: CRIAÇÃO DE CONJUNTOS DE GESTOS POR USUÁRIOS SEM DIFICULDADES MOTORA E DE FALA

Com o objetivo de analisar a capacidade do classificador no reconhecimento de gestos criados por diferentes usuários, professores de uma das instituições coparticipantes foram convidados a participar de um experimento. Cinco voluntários aceitaram o convite (pessoas sem comprometimento motor e de fala, servidores da UTFPR - Câmpus Pato Branco) e participaram da primeira etapa da coleta de dados. O objetivo desta etapa foi avaliar se o sistema seria capaz de reconhecer gestos personalizados por meio de treinamento, considerando conjuntos de dados com poucas amostras.

O experimento foi projetado para avaliar a precisão do classificador no reconhecimento de gestos personalizados, utilizando conjuntos de dados compostos por um número pequeno de amostras. A autora desta tese de doutorado conduziu o experimento e desempenhou o papel de cuidador. Os voluntários foram convidados a criar conjuntos de dados compostos por seis a oito gestos diferentes, com rótulos e movimentos definidos pelos próprios voluntários, capturados em diferentes ambientes.

Para esta etapa de avaliação, apenas a Área do Cuidador e a Área do Usuário estavam disponíveis para uso no sistema PGCA. Somente a representação de movimento por meio de MHI convencional havia sido implementada e não havia possibilidade de armazenar vídeos dos movimentos realizados. O sistema já possuía dois classificadores implementados: SVM e CNN. A Área do Cuidador não dispunha de validação das amostras capturadas, mas permitiu a criação de conjuntos de dados de gestos personalizados, treinamento e avaliação do sistema. A Área do Usuário foi usada apenas para testar o reconhecimento dos gestos para os quais o sistema foi treinado. Informações sobre os conjuntos de dados criados pelos voluntários nesse experimento são apresentados na sequência.

5.2.1 Conjuntos de dados criados

Os voluntários V1 e V2 criaram conjuntos de dados com oito classes distintas, registrando vinte amostras por classe. O voluntário V3 criou um conjunto de dados com oito classes distintas, registrando quinze amostras por classe. A Figura 5.1 apresenta exemplos de amostras geradas pelos voluntários V1, V2 e V3 para compor cada conjunto de dados, representadas por meio de MHI convencional. Os voluntários V1 e V2 utilizaram a opção "Aproximar", disponível nas configurações do sistema, que aproxima a área de captura da câmera para destacar e melhorar a percepção dos movimentos faciais.

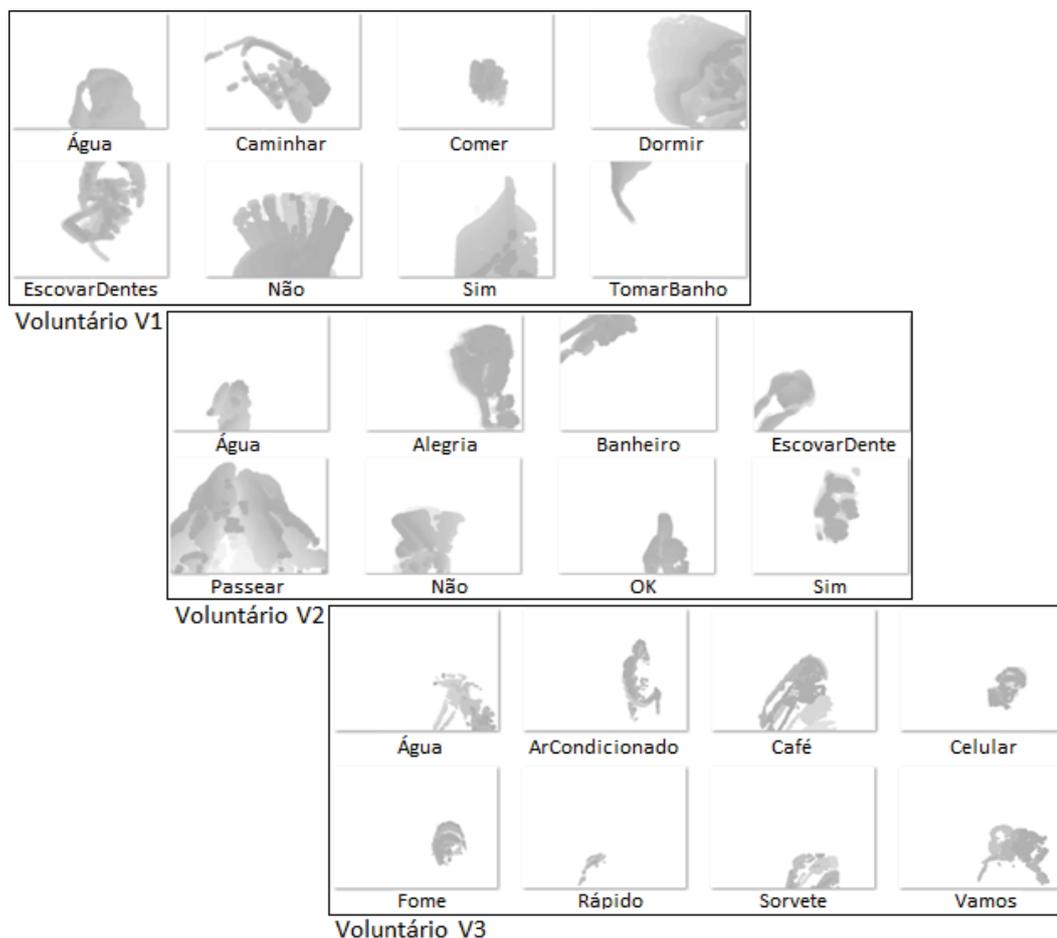


Figura 5.1: Exemplos de amostras de gestos executados pelos voluntários V1, V2 e V3 (pessoas sem deficiência motora e de fala), registradas durante a execução do Experimento 2 e representadas por meio de MHI convencional.

Os voluntários V4 e V5 criaram conjuntos de dados com seis classes distintas. O voluntário V4 registrou vinte amostras por classe e o voluntário V5 registrou quinze amostras por classe. A Figura 5.2 apresenta exemplos de amostras geradas pelos voluntários V4 e V5 para compor cada conjunto de dados, representadas por meio de MHI convencional.

Os conjuntos de dados criados pelos voluntários foram utilizados para realizar a avaliação de desempenho dos classificadores, cujos resultados são apresentados na sequência.

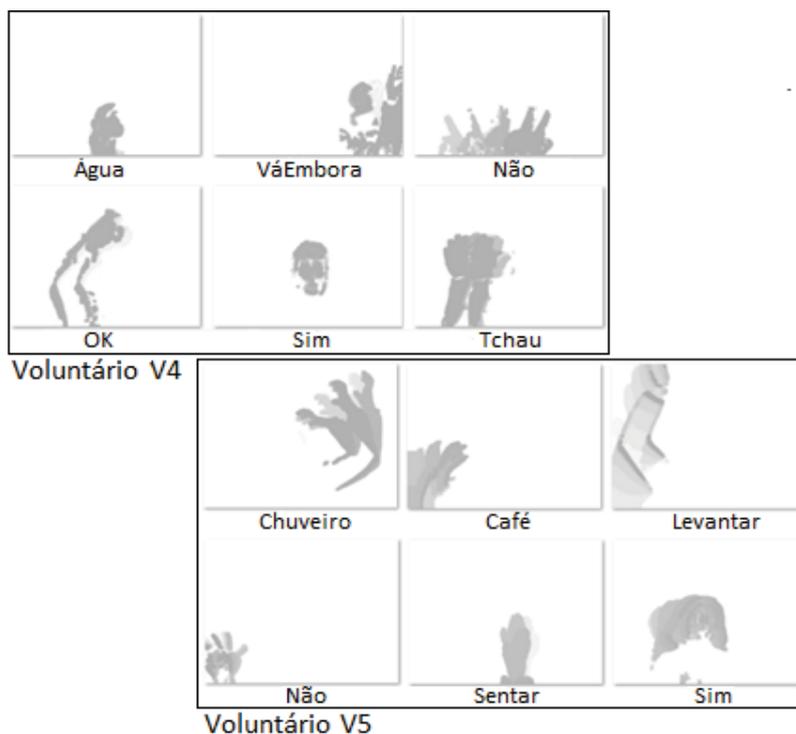


Figura 5.2: Exemplos de amostras de gestos executados pelos voluntários V4 e V5 (pessoas sem deficiência motora e de fala), registradas durante o Experimento 2 e representadas por meio de MHI convencional.

5.2.2 Avaliação de desempenho dos classificadores

Para avaliar o desempenho dos dois classificadores implementados no sistema PGCA, foi utilizada a validação cruzada K-fold, considerando dez pastas, separando 90% dos dados para treinamento e 10% para teste. O número de dados de treinamento foi expandido via Aumento de Dados, criando oito amostras adicionais a partir de cada amostra original.

Após a execução de testes em todas as pastas, a precisão geral (média ponderada), desvio padrão, variância e o Cohen Kappa (medida estatística de concordância entre avaliadores) foram calculados para cada um deles. SVM (com descritor HOG) e CNN (usando Transferência de Aprendizado) foram as técnicas de aprendizado de máquina utilizadas. Os resultados obtidos comparando os dois métodos de aprendizado para os conjuntos de dados gerados pelos cinco voluntários são apresentados na Tabela 5.2.

Tabela 5.2: Conjuntos de dados criados por voluntários no Experimento 2 — comparação de métodos de aprendizado de máquina: Precisão média, Cohen Kappa, Desvio padrão e Variância.

Voluntário	SVM				CNN			
	Prec.	Cohen k	Desv.	Var.	Prec.	Cohen k	Desv.	Var.
V1 - MHI	0.981	0.979	0.04	0.00180	0.994	0.993	0.02	0.0004
V2 - MHI	0.981	0.978	0.02	0.00090	0.987	0.985	0.02	0.0007
V3 - MHI	0.975	0.971	0.07	0.00620	0.941	0.936	0.07	0.0069
V4 - MHI	0.974	0.970	0.03	0.00160	0.974	0.970	0.03	0.0016
V5 - MHI	0.988	0.986	0.02	0.00006	1	1	0	0

5.2.3 Discussão

Normalmente, uma classificação perfeita produziria uma variação e desvio padrão de zero, e o valor de precisão e kappa de um. De acordo com os critérios de Landis e Koch (1977), para a interpretação do valor kappa: 0,0 a 0,2 = concordância leve, 0,2 a 0,4 = concordância justa, 0,4 a 0,6 = concordância moderada, 0,6 a 0,8 = concordância substancial e 0,8 a 1,0 = concordância quase perfeita. Nesse experimento, os classificadores apresentaram resultados satisfatórios, uma vez que a precisão média obtida em todos os conjuntos de dados foi alta (superior a 0,94), foram observados baixos desvios e variâncias e os valores de kappa indicaram concordância quase perfeita. O classificador com base em CNN apresentou uma precisão ligeiramente melhor em comparação ao classificador com base em SVM, em quatro dos cinco conjuntos de dados utilizados.

Durante a análise dos dados capturados nesse experimento, foi identificada a necessidade de fazer uma alteração na metodologia, que originalmente (Ascari et al., 2018b) previa a realização do processo de Aumento de Dados antes de separar os dados originais em dados de treinamento e teste. Essa situação pode gerar uma avaliação de desempenho muito otimista. Portanto, a metodologia foi atualizada para realizar primeiramente a divisão do conjunto de dados original em dados de treinamento e teste, para executar posteriormente o processo de Aumento de Dados apenas nos conjuntos de dados de treinamento. Essa alteração foi feita antes de avaliar o desempenho do sistema nos conjuntos de dados criados pelos voluntários.

Durante o experimento, foi possível perceber que as condições de iluminação nos diferentes ambientes trabalhados para coleta de dados podem interferir na capacidade do sistema em capturar corretamente os movimentos executados pelos usuários. Um dos voluntários foi posicionado ao lado de uma janela (fonte de luz) e foi possível perceber que a representação de gesto gerada quando o usuário foi posicionado com o corpo de lado para a janela é significativamente diferente da representação gerada quando o usuário foi posicionado com o corpo inteiro voltado para a janela. Essa condição pode ser observada principalmente no conjunto de gestos gerado pelo voluntário V3. Além disso, os gestos executados por esse voluntário foram em geral feitos com movimentos reduzidos, ou seja, gestos feitos com as mãos e braços em movimentos de pouca amplitude. Diferentes expressões faciais foram utilizadas, mas estas características não foram capturadas pelo sistema (o sistema ainda não realizava o rastreamento de pontos de referência faciais — *Landmarks*). Possivelmente essas situações contribuíram para que o desempenho apresentado pelos classificadores nesse caso tenha sido menor que os apresentados para os conjuntos de gestos gerados pelos outros voluntários.

Conforme observado nos conjuntos de gestos gerados pelos voluntários V1, V2 e V5, gestos executados com movimentos bastantes distintos, tendem a ser melhor classificados. Além disso, quanto maior o número de amostras coletadas para cada gesto, maior é a chance de o classificador aprender mais na etapa de treinamento e possivelmente apresentar melhores resultados na etapa de testes.

Com relação aos conjuntos de dados gerados pelos voluntários V1 e V2, a utilização da opção "Aproximar", não apresentou diferenças significativas na capacidade de classificação do sistema. Apenas observou-se que, nesses casos, tem-se melhor visualização de movimentos feitos com a região do rosto dos usuários. Em situações como a que ocorreu com o gesto rotulado como "EscovarDentes" executado pelo voluntário V2, percebeu-se que o classificador teve bastante dificuldade para identificar o gesto realizado. Em situações como essa, espera-se que o sistema possa alertar o usuário e seu cuidador de que o gesto em questão não deveria ser utilizado para interação com o sistema, demandando uma nova coleta de dados para ele, talvez com a realização de algum movimento diferente.

Os diferentes resultados apresentados pelo classificador podem indicar que a capacidade do sistema conseguir classificar corretamente os gestos para o qual foi treinado é bastante dependente das escolhas realizadas pelo usuário e cuidador, quanto ao movimento que irá caracterizar cada gesto, e também quanto ao ambiente utilizado para captura das imagens e uso do sistema. Uma característica considerada desejável futuramente para o sistema é a capacidade de interpretar os resultados obtidos pelo classificador, de forma a alertar o usuário e seu cuidador sobre quais gestos teriam maiores chances de serem corretamente classificados pelo sistema, e que poderiam, portanto, ser utilizados para execução de ações como forma de interação com a interface do sistema. Essa recomendação foi incluída na descrição da metodologia MyPGI.

Nesse experimento apenas as imagens MHI foram registradas. Como não foram armazenados vídeos referentes aos movimentos executados, não foi possível avaliar esses conjuntos de dados com a representação de movimento OF-MHI implementada posteriormente.

5.3 EXPERIMENTO 3: AVALIAÇÃO DE TECNOLOGIAS USANDO *KECK GESTURE DATASET*

Em uma nova etapa de avaliação, além de todos os recursos disponíveis anteriormente, a representação de movimento via OF-MHI foi incluída no sistema PGCA antes da realização dos testes previstos para esta etapa. O conjunto de dados público *Keck Gesture Dataset* (disponível em Jiang et al. (2012)) foi usado para avaliar o desempenho do sistema usando os dois classificadores (SVM e CNN) e duas representações de movimento implementados (MHI convencional e OF-MHI).

Keck Gesture Dataset é composto por quatorze gestos distintos, realizados por três pessoas na frente de um fundo estático. Para cada gesto, cada pessoa realiza três repetições. Exemplos de representações de movimento geradas para cada uma das quatorze classes de gestos disponíveis no *Keck Gesture Dataset* são apresentadas na Figura 5.3.

Para avaliação de desempenho, nove amostras estavam disponíveis para cada uma das quatorze classes existentes no conjunto de dados: um total de cento e vinte e seis amostras originais. A opção "Aproximar" foi usada para dar destaque às regiões onde os movimentos são realizados.

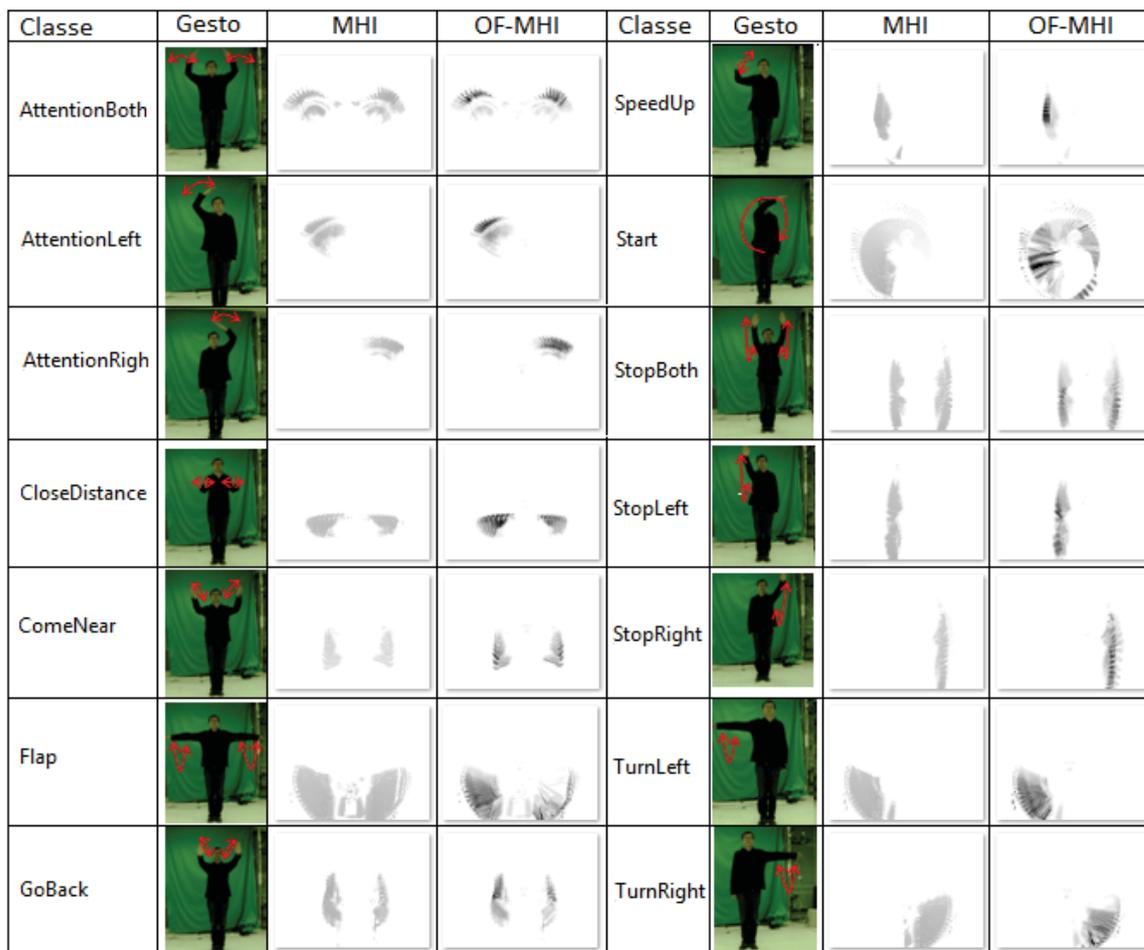


Figura 5.3: Amostras de gestos do *Keck Gesture Dataset* utilizado no Experimento 3, representados por MHI convencional e MHI com base em fluxo óptico (OF-MHI). As imagens em tons de cinza foram redimensionadas para dar ênfase às regiões em movimento.

5.3.0.1 Avaliação de desempenho das tecnologias empregadas

Os resultados obtidos a partir da avaliação de desempenho do sistema no reconhecimento dos gestos que compõem o conjunto de dados *Keck Gesture Dataset* são apresentados na Tabela 5.3. Foi empregada a validação cruzada K-fold com nove pastas, sendo que cada pasta continha uma amostra por classe para teste e setenta e duas amostras por classe para treinamento (após a execução do processo de Aumento de Dados).

Tabela 5.3: Uso do *Keck Gesture Dataset* no Experimento 3 — comparação de métodos de aprendizado de máquina: Precisão média, Cohen Kappa, Desvio padrão e Variância.

Keck Dataset	SVM				CNN			
	Prec.	Cohen k	Desv.	Var.	Prec.	Cohen k	Desv.	Var.
MHI	0.88	0.87	0.04	0.0025	0.90	0.89	0.05	0.0038
OF-MHI	0.89	0.88	0.04	0.0026	0.87	0.86	0.07	0.0060

Existem diferentes trabalhos que visam reconhecer gestos, ações ou imagens nas quais o *Keck Gesture Dataset* foi usado para avaliar a precisão dos classificadores ou métodos utilizados. Como nenhum estudo foi encontrado usando exatamente a mesma forma de avaliação empregada nesta tese (validação cruzada K-fold usando nove pastas), uma comparação direta do desempenho dos classificadores não foi realizada. Entretanto, é possível mencionar alguns

trabalhos considerados relevantes e os correspondentes resultados obtidos com o *Keck Gesture Dataset*. Por exemplo, no estudo de Pei et al. (2013), um algoritmo baseado em índice invertido rápido é introduzido para reconhecimento de ação em várias classes. Os resultados apresentados pelo método proposto por esses autores indicam uma precisão de até 89,88%. Fu et al. (2013) consideraram o problema do reconhecimento de ações com base na estrutura geométrica. O método proposto pelos autores utiliza uma estrutura de baixa dimensão no coletor Grassmanniano para representar sequências de vídeo usando a estrutura linear do espaço tangente e apresentou precisão de reconhecimento de 93,4%. Wan et al. (2014) apresentaram uma abordagem de aprendizado de dicionário específica por classe por meio da teoria da informação para reconhecimento de ação e gesto, e a precisão do reconhecimento alcançada é de até 95,1%. O estudo de Zhang et al. (2018) introduziu um modelo híbrido baseado em CNN para classificação de imagens, e os resultados indicam uma precisão de até 93,15%.

Neste Experimento 3, ambos os classificadores apresentaram resultados satisfatórios, utilizando o MHI e OF-MHI. Como apresentado na Tabela 5.3, a precisão apresentada pelo classificador com base em SVM utilizando a representação de movimento MHI foi de 88% e utilizando OF-MHI foi de 89%, já a precisão apresentada pelo classificador com base em CNN utilizando MHI foi de 90% e utilizando OF-MHI foi de 87%. A classificação usando os dois conjuntos de dados criados a partir do *Keck Gesture Dataset* apresentou precisão e dados estatísticos válidos com poucas variações. Portanto, os dois classificadores e as duas representações de movimento foram avaliados novamente nos experimentos seguintes, realizados com o público-alvo.

5.4 ENTREVISTAS COM ESPECIALISTAS DE DOMÍNIO

Depois de avaliar a interface do sistema em laboratório com profissionais que atuam com IHC e depois de avaliar diferentes tecnologias voltadas para reconhecimento de gestos e representação de movimento utilizando conjuntos de dados distintos, entrevistas com professores da área de educação especial foram realizadas.

O objetivo das entrevistas foi conhecer melhor a realidade dos alunos que possuem dificuldade motora e de fala no ambiente escolar, e identificar potenciais alunos para participarem de experimentos para avaliação do sistema PGCA. O uso de entrevistas foi motivado pela possibilidade de obter uma compreensão mais aprofundada sobre o tema de pesquisa, e obter um *feedback* para complementar os dados obtidos por meio de experimentos e comportamentos observados.

Foram conduzidas entrevistas semi-estruturadas, guiadas por um roteiro de perguntas abertas e fechadas. Esse tipo de entrevista é muito utilizado quando se deseja delimitar o volume das informações, obtendo assim um direcionamento maior para o tema, intervindo a fim de que os objetivos sejam alcançados (Boni e Quaresma, 2005). Além disso, entrevistas semi-estruturadas dão um certo nível de liberdade ao pesquisador, permitindo aproveitar oportunidades e imprevistos que possam ocorrer. As questões utilizadas para guiar as entrevistas com especialistas de domínio são apresentadas na Tabela 5.4.

Os participantes tiveram a possibilidade de discorrer sobre o tema proposto, seguindo o conjunto de perguntas previamente definidas, mas em um contexto muito semelhante ao de uma conversa informal. As respostas das perguntas foram registradas pela autora desta tese, que conduziu todas as entrevistas.

Tabela 5.4: Questões utilizadas para guiar as entrevistas semi-estruturadas com especialistas de domínio.

ID	Questão
Q1	Vínculo com o(s) aluno(s).
Q2	Tempo de experiência na educação especial.
Q3	Formas de comunicação utilizadas com o(s) aluno(s).
Q4	Já utilizou algum sistema de CAA para se comunicar com o(s) aluno(s)? Qual(is)? Algum sistema computacional?
Q5	Dificuldades do(s) aluno(s) em relação ao uso de sistemas computacionais.
Q6	Percepção pessoal em relação à comunicação social do(s) aluno(s).
Q7	Capacidade do(s) aluno(s) na realização de gestos voluntários.
Q8	Expectativas e experiências relacionadas ao uso de câmera e reconhecimento gestual para interação do(s) aluno(s) com sistemas computacionais.
Q9	Outras informações relevantes.

5.4.1 Participantes

Quatro escolas do município de Pato Branco/PR vinculadas a instituições coparticipantes foram visitadas. Uma das escolas é uma instituição educacional específica para alunos com necessidades especiais (Associação de Pais e Amigos do Excepcional - APAE), e as outras são escolas inclusivas onde alunos com necessidades especiais também podem frequentar as aulas regulares: Centro Estadual de Educação Básica para Jovens e Adultos (Escola Pública A), Colégio Estadual - Ensino Fundamental, Médio e Profissional (Escola Pública B), Colégio Estadual - Ensino Fundamental e Médio (Escola Pública C).

Nas quatro escolas visitadas, foram realizadas entrevistas com professores regentes, professores de apoio a comunicação, cuidador e tradutor e intérprete de língua de sinais (intérprete de Libras), para entender melhor as características do público-alvo da metodologia proposta. As entrevistas foram conduzidas com oito profissionais vinculados à educação especial. A Tabela 5.5 apresenta algumas informações sobre os profissionais entrevistados.

Tabela 5.5: Informações sobre os profissionais de educação especial (especialistas do domínio) entrevistados.

Profissional	Sexo*	Experiência	Vínculo	Instituição
P1	F	28 anos em educação especial	Professor regente	APAE
P2	F	19 anos em educação especial	Professor regente	APAE
P3	M	4 anos em educação especial	Professor itinerante	APAE
P4	F	11 anos em educação especial	Professor regente	APAE
P5	F	18 anos como tradutora e intérprete de língua de sinais	Tradutora e intérprete de língua de sinais (Libras)	Escola Pública A
P6	F	6 anos como assistente operacional	Cuidador	Escola Pública A
P7	F	14 anos em educação especial	Professor de apoio a comunicação	Escola Pública B
P8	M	7 anos em educação especial	Professor de apoio a comunicação	Escola Pública C

*Sexo: M - Masculino; F- Feminino.

5.4.2 Contribuições

Quando questionados sobre a principal forma de comunicação utilizada com os alunos que apresentam dificuldades motora e de fala, todos os oito profissionais relataram o uso da linguagem gestual. Os professores relataram que alguns alunos tem conhecimento de alguns sinais de Libras e os utilizam para comunicação, mas a maioria dos alunos tem seu próprio

vocabulário de sinais, ou seja, gestos usados no ambiente familiar que não seguem uma linguagem de sinais específica, também chamados de gestos caseiros.

Professores que trabalham na escola específica para alunos com necessidades especiais relataram que usam muita associação de conceitos com imagens, o que sugere que o uso de pranchas de comunicação com figuras para incentivar a interação dos alunos com o sistema é uma abordagem válida. Além disso, muitos alunos com necessidades especiais matriculados nessa escola ainda não são totalmente alfabetizados, o que impede o uso de pranchas alfabéticas. Como as pranchas disponíveis no sistema por padrão, poderiam não atender às demandas dos professores e seus alunos em todos os possíveis cenários, foi incluído no sistema PGCA a Área de Pranchas de Comunicação para oferecer suporte aos cuidadores na criação de pranchas de comunicação personalizadas.

O uso de sistema específico de CAA foi relatado apenas pelos dois professores que trabalham fornecendo apoio a comunicação dos alunos. Uma professora mencionou conhecer o sistema *Boardmaker* (Mayer-Johnson, 2004), mas alegou não o usar no ambiente escolar devido ao custo associado. Outro professor citou o aplicativo *Livox* (Pereira, 2019), apesar de nunca o ter usado com a aluna que acompanha em função da aluna não conseguir interagir com dispositivos por meio de toque. A profissional que trabalha como intérprete de Libras mencionou o uso de alguns *software* específicos para o ensino e reconhecimento da Libras, como o "Sinalário disciplinar de Libras" (SEED-PR, 2019b) e o "ICOM App" (SEED-PR, 2019a) para fazer videochamadas com surdos. Ela também usa o "livro com CDs" (Velo e Maia, 2011) para revisar conceitos, significados e representar movimentos.

Nenhum dos oito profissionais entrevistados usou *software* que usa câmera para reconhecimento de gestos. Alguns professores disseram que usaram câmera apenas para gravar vídeos dos alunos, gravar fotos e apresentá-los mais tarde na forma de slides em apresentações. A intérprete de Libras comentou já ter usado o *software* Windows Live Messenger™ com a câmera habilitada para se comunicar por meio de sinais de Libras com alguns alunos surdos. Na primeira conversa com um dos professores, ele estava preocupado se algo seria colocado no corpo do aluno. Segundo esse professor, a mãe de um aluno participante desta pesquisa já estava cansada de levá-lo para realizar testes com dispositivos vestíveis. Neste caso, o uso de uma câmera exclusivamente foi decisivo para a família concordar em autorizar a participação no experimento.

Os alunos com necessidades especiais matriculados nas escolas visitadas são em sua maioria oriundos de famílias carentes e, de acordo com seus professores, a simples inclusão no ambiente escolar desempenha um papel muito importante em seu desenvolvimento social. Além de aprender o conteúdo curricular, esses alunos também recebem muitas orientações relacionadas à execução de atividades diárias, a fim de motivar esses alunos a serem mais autônomos e funcionalmente independentes.

Embora o foco da pesquisa não estar direcionado ao processo de aprendizagem de crianças com necessidades especiais, é possível observar que o sistema desenvolvido tem potencial para servir de apoio aos professores na tarefa de alfabetização. Nesse contexto, alguns professores relataram que um método de alfabetização que tem mostrado bons resultados na alfabetização de crianças com necessidades especiais é o método Desafios da Aprendizagem —ABACADA (Silva, 2016). Esse método é destinado a estudantes com dificuldades de aprendizagem e deficiência intelectual e é baseado no Método Fônico (Capovilla et al., 2004) e Método Sodré (Sodre, 1986). O método ABACADA associa a sílaba inicial de cada palavra a uma figura correspondente (Figura 5.4 (A)), e poderia ser trabalhado com os alunos por meio de interação gestual com pranchas de figuras customizadas pelos professores.

Outro método citado foi o método Boquinhas fonovisuoarticulatório, desenvolvido por Jardini (2003) com o objetivo de alfabetizar e reabilitar distúrbios de leitura e escrita. Esse método baseia-se no pressuposto de que a integração sensorial de estímulos é benéfica para a consolidação da aprendizagem. A proposta consiste em estimular a alfabetização por meio da fala, do exercício da consciência fonológica e da consciência fonarticulatória. O sistema PGCA poderia ser utilizado para interagir com uma prancha de comunicação com fotos de bocas por meio de gestos, como exemplificado na Figura 5.4 (B).

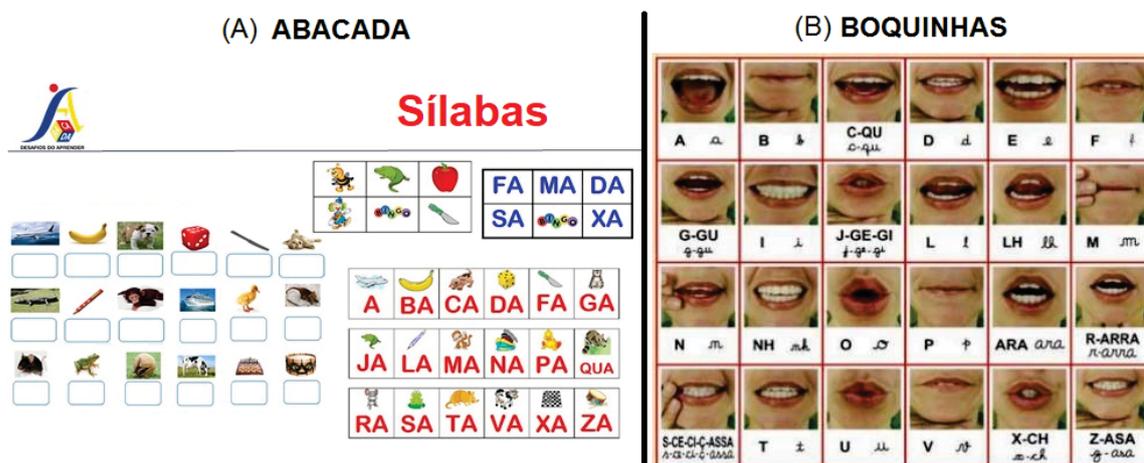


Figura 5.4: Métodos (A) ABACADA e (B) Boquinhas empregados na alfabetização de crianças com necessidades especiais, citados por profissionais em entrevistas.

Fonte: Silva (2019) e Jardini (2003); Pianca (2016)

Todos os oito professores consideraram a proposta desta tese interessante, válida, e mesmo com dúvidas sobre sua efetividade, ou aplicação, não mediram esforços para auxiliar na transferência de informações. A expectativa de um dos professores era de que nosso estudo pudesse fornecer alternativas para a comunicação com alunos que não realizam nenhum movimento voluntário ou não demonstram qualquer expressão apropriada às interações que são feitas com eles. Infelizmente, o estágio atual de nosso estudo não visa esse público, pois a solução desenvolvida pressupõe que o usuário tenha algum movimento padronizado que possa ser reconhecido e a capacidade de realizar interação intencional.

O sistema desenvolvido visa tanto se beneficiar do conhecimento dos cuidadores quanto contribuir com suas estratégias de comunicação, podendo promover a participação efetiva dos alunos em sala de aula e sua comunicação com professores e outras pessoas. Ao utilizar pranchas de comunicação e interagir com sistemas computacionais, os alunos podem expressar suas intenções de comunicação de maneira mais diversificada, não apenas utilizando os conceitos associados aos gestos que podem realizar.

Observando a rotina dos alunos que não falam no ambiente escolar e as informações obtidas a partir das entrevistas, foi possível perceber que os alunos que utilizam a comunicação gestual acabam tendo suas possibilidades de interação limitadas, pois seus gestos só podem ser interpretados se o professor estiver olhando para eles. No caso de alunos que realizam apenas movimentos de "Sim" e "Não", só é possível que o aluno interaja com o professor ou com os colegas de classe quando eles lhe direcionam perguntas afirmativas ou negativas. Portanto, o potencial do sistema proposto em permitir que os alunos interajam por meio de gestos personalizados com pranchas de comunicação foi visto como evidente. Tais possibilidades de interação permitem que os alunos expressem suas intenções de comunicação, chamando a atenção do professor quando necessário (usando sintetizador de voz), ganhando algum nível de autonomia e poder de interação no ambiente em que estão inseridos.

Os resultados obtidos por meio das entrevistas contribuíram para a identificação de algumas necessidades de melhorias no sistema PGCA que foram implementadas antes de conduzir os experimentos seguintes com o público-alvo, descritos na sequência. Alguns desafios também foram identificados nesta etapa, como o fato de alguns alunos ainda não serem totalmente alfabetizados, o que impediria o uso de pranchas de comunicação alfabética; a disponibilidade dos professores, que em alguns casos precisariam usar seu período de hora-atividade para acompanhar as sessões; dúvidas dos profissionais quanto à capacidade de compreensão de alguns alunos; e ainda a percepção de que alguns alunos poderiam não mostrar interesse em participar dos experimentos.

5.5 EXPERIMENTO 4: USO DO SISTEMA PGCA POR ALUNOS COM DIFICULDADES MOTORA E DE FALA

Antes de proceder a avaliação do sistema PGCA com o público-alvo, diversas melhorias foram implementadas: a) armazenamento do vídeo referente aos movimentos utilizados para criar os conjuntos de dados; b) inclusão de uma guia para a criação de pranchas de comunicação de imagens; c) inclusão de novas opções de configuração para simular o uso do teclado; d) visualização em vídeo do movimento relacionado aos gestos selecionados na tela de configuração; e) possibilidade de escolher diferentes pranchas de comunicação na área do usuário; f) registro de novas informações sobre as principais ações executadas na interface do sistema em um arquivo de *log*.

Além das melhorias implementadas no sistema PGCA, algumas possibilidades de melhorias na descrição da metodologia MyPGI também foram percebidas como importantes e motivaram novas implementações no sistema: a) previsão de gerar backup dos dados originais antes de executar as etapas de divisão (em dados de treinamento e testes) e Aumento de Dados; b) possibilidade de reiniciar a coleta de dados, restaurando o backup feito no início do processo de treinamento; c) sugestão para criar dois modelos para reconhecimento de gestos, um modelo para a etapa de avaliação e outro modelo para a etapa de interação do usuário com o sistema, contendo todas as amostras disponíveis no conjunto de dados.

Concluídas as implementações dessas melhorias no sistema, e após a autora desta tese de doutorado ter conhecido vários alunos com necessidades especiais matriculados nas quatro escolas visitadas, um grupo de sete alunos foi convidado a participar de um experimento para criar conjuntos de gestos personalizados e testar a interação gestual com o sistema PGCA. O grupo foi convidado com base nos resultados das entrevistas e recomendações dos professores que sugeriram alunos que teriam maior capacidade de compreensão e condições para realizar movimentos voluntários. O convite foi enviado para as famílias dos alunos sendo que um aluno não foi autorizado a participar. Portanto, esse experimento contou com a participação de seis alunos com dificuldades motora e de fala.

5.5.1 Participantes

Todos os alunos participantes do Experimento 4 são caracterizados como pessoas com paralisia cerebral, com diferentes níveis de necessidades especiais, conforme apresentado na Tabela 5.6.

Tabela 5.6: Características dos alunos com dificuldades motora e de fala participantes do Experimento 4.

Aluno	Sexo*	Idade	Relatório médico
A	M	18 anos	Paralisia cerebral devido a sequelas durante o parto.
B	F	29 anos	Dano cerebral e hidrocefalia discreta.
C	M	38 anos	Quadriplegia com componente de atetose, perda auditiva sensorioneural bilateral.
D	F	20 anos	Paralisia pseudobulbar. Hipotonia generalizada e hiperreflexia.
E	F	18 anos	Encefalopatia estática e quadriplegia espástica.
F	M	18 anos	Encefalite estática, epilepsia, e síndrome de Rubinstein Taybi.

*Sexo: M - Masculino; F- Feminino.

Cada aluno que participou do experimento foi acompanhado por um professor que desempenhou o papel de "cuidador" no sistema, informando os gestos que o aluno costuma usar para se comunicar e o significado de cada gesto. As tarefas esperadas para serem realizadas durante o experimento foram: 1. criação do conjunto de dados capturando gestos para treinar o sistema; 2. treinamento e avaliação do sistema; 3. uso do sistema para reconhecer gestos; 4. uso de gestos para selecionar imagens em prancha de comunicação; 5. uso de gestos para interagir com editor de texto ou navegador de Internet.

Nesse experimento, buscou-se posicionar os usuários na frente de uma fonte de luz, seja uma janela ou uma lâmpada comum. As tarefas conduzidas pelos professores permitiram avaliar o sistema na criação de um conjunto de dados com gestos personalizados para cada aluno. Detalhes sobre a execução dessas atividades são apresentados na sequência.

5.5.2 Execução de tarefas por participante

Para esse experimento com o sistema PGCA, o classificador com base em CNN foi usado para reconhecimento de gestos, no entanto, para cada gesto, o vídeo do movimento foi armazenado e, posteriormente, o reconhecimento de gestos foi simulado usando também o classificador com base em SVM. Todas as ações realizadas no sistema por meio de interação gestual foram armazenadas em um arquivo de texto (*log*) para registrar eventos na interação com a interface. As imagens referentes aos conjuntos de dados gerados pelos alunos são apresentadas apenas na forma de MHI ou OF-MHI para manter o anonimato dos participantes. Detalhes sobre como cada aluno executou as tarefas previstas são apresentados a seguir.

5.5.2.1 Aluno A

O Aluno A usa apenas dois gestos de cabeça no ambiente escolar para se comunicar com seus colegas e professores, que se referem a "Sim" e "Não". Esse aluno tem uma capacidade de compreensão muito preservada. A coleta de dados para o treinamento do sistema e os testes de interação com a interface foram realizados em duas sessões diferentes. Para esse aluno, o sistema foi configurado para usar a opção "Aproximar" para capturar melhor os movimentos faciais. O aluno A apresentou alguns movimentos involuntários durante a interação com o sistema, apoiando a perna ou o braço na mesa onde o computador e a câmera estavam dispostos, gerando algumas amostras com informações sobre o fundo da sala, consideradas como ruídos.

O conjunto de dados criado por esse aluno foi composto por duas classes; quinze amostras de cada classe foram consideradas válidas para treinar o sistema. Após o treinamento, a primeira etapa do teste de interação com a interface referia-se ao reconhecimento dos gestos para os quais o sistema foi treinado: os dois gestos foram corretamente identificados pelo sistema quando realizados voluntariamente pelo aluno. Posteriormente, o sistema foi configurado associando a classe "Sim" à opção de confirmação do sistema. Tal configuração permitiu testar o

uso de prancha de comunicação para escrever palavras relacionadas a solicitações específicas, como "Dormir", "Banheiro", "Comida" e outras.

A seguir, o sistema foi configurado associando a classe "Não" com a tecla ENTER, e a opção "simular teclado" foi habilitada para testar a interação com outros aplicativos. Essa configuração permitiu realizar a seleção de imagens em uma prancha de comunicação para escrever palavras diretamente em um editor de texto, simulando o pressionamento da tecla ENTER do teclado ao realizar o gesto "Não". Alguns falsos positivos foram observados nesse teste, e para minimizar essa situação, uma filtragem por nível de confiança maior que 0.5 (certeza de uma amostra pertencer a uma determinada classe) foi incluída no classificador com base em CNN. O classificador com base em SVM emprega um nível de confiança de pelo menos 0.5 por padrão. A Figura 5.5 (A) apresenta exemplos de amostras geradas pelo Aluno A para compor seu conjunto de dados.

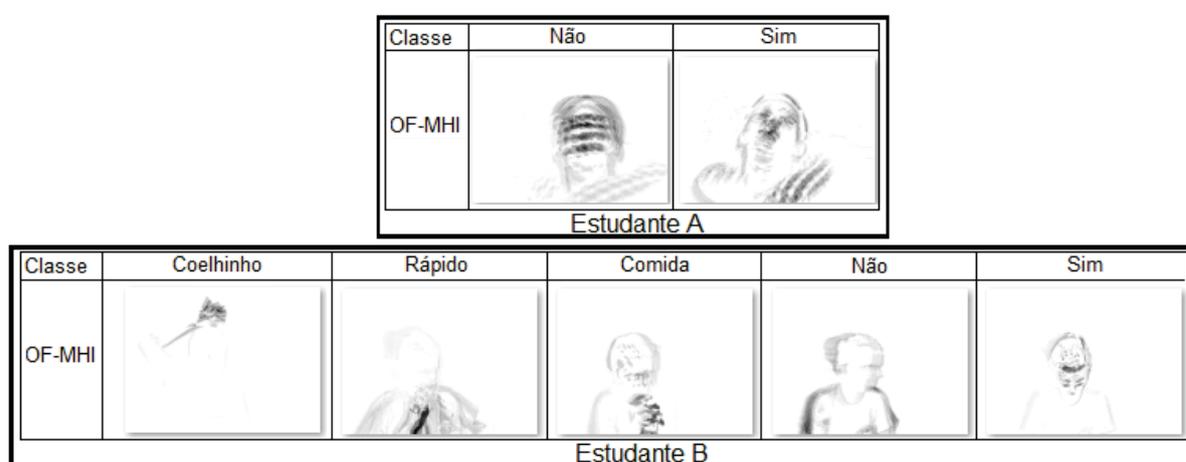


Figura 5.5: Exemplos de amostras de gestos executados pelos Alunos A e B (pessoas com dificuldades motora e de fala), registradas durante o Experimento 4 e representadas por meio de OF-MHI.

5.5.2.2 Aluna B

A Aluna B não apresenta comprometimento motor significativo e usa gestos de mão e cabeça no ambiente escolar para se comunicar, no entanto, evita interagir com pessoas desconhecidas. Ela tem algumas dificuldades na capacidade de compreensão e é muito tímida, exigindo constante incentivo da professora que a acompanhou para realização dos gestos. A coleta de dados para o treinamento do sistema foi realizada em uma sessão em um dia e os testes de interação com a interface foram realizados em outra sessão três dias depois. A opção "Aproximar" foi desativada no sistema. Seu conjunto de dados customizado foi composto de cinco classes com treze amostras cada, conforme apresentado na Figura 5.5 (B). Após o treinamento, no teste de interação, os gestos referentes a "Sim", "Rápido" e "Coelhinho" (referente ao Coelho da Páscoa) foram corretamente reconhecidos. Os gestos referentes a "Comida" e "Não" geraram algumas interpretações errôneas, mas isso não impediu a interação com o sistema. Posteriormente, o sistema foi configurado associando-se a classe "Sim" à opção de confirmação do sistema, e a aluna selecionou imagens em uma prancha de comunicação composta por imagens relacionadas a artesanato, pois seria um assunto de interesse da aluna, segundo sua professora. Essa prancha foi usada para escrever palavras associadas a cada figura na interface do sistema quando a aluna executou o gesto "Sim". Além disso, uma prancha de comunicação composta por imagens de vogais foi usada para a aluna indicar a primeira letra da palavra "Elefante". Com a ajuda da professora, a aluna fez o gesto "Sim" para selecionar a imagem referente à letra "E" quando

apresentada na interface do sistema e o som correspondente foi emitido. A Figura 5.6 apresenta representações de tarefas realizadas pelos alunos A e B. Como a aluna exigiu bastante intervenção da professora e teve pouca iniciativa na seleção de imagens por conta própria, nenhum outro teste de interação foi realizado.

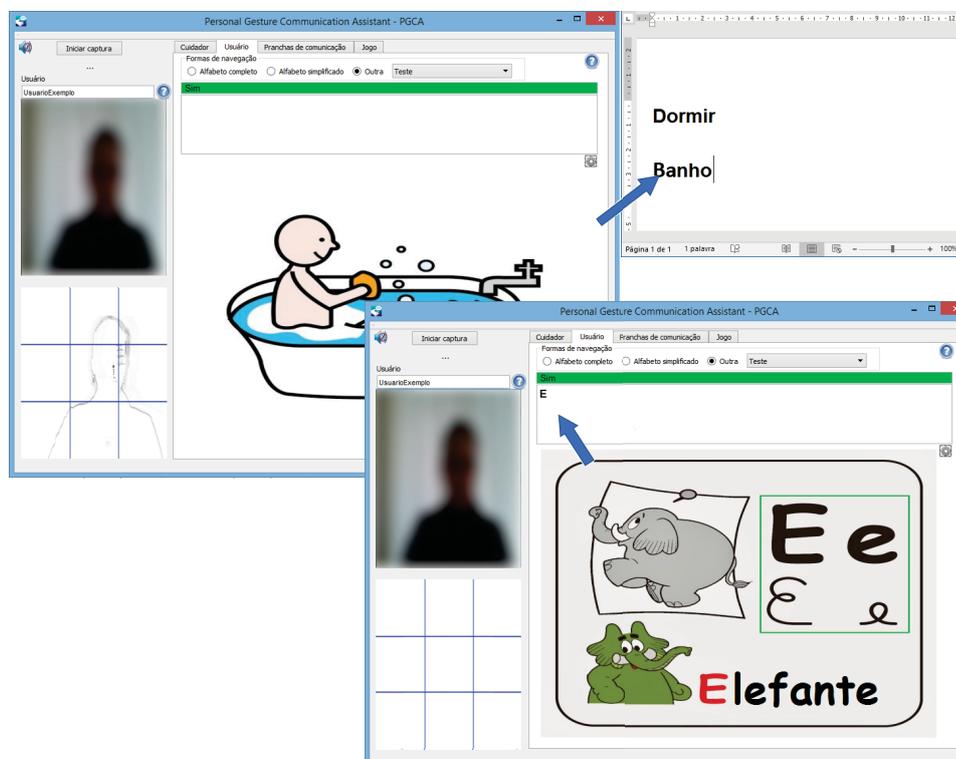


Figura 5.6: Pranchas de comunicação utilizadas pelos alunos com dificuldades motora e de fala durante o Experimento 4 para selecionar imagens e escrever em um editor de texto, e para indicar a letra inicial de uma palavra.

Fonte: Imagens utilizadas nas pranchas disponíveis em Rodrigo e Corral (2013), personagem de Sousa (2019) adaptado de PortalEscola (2014).

5.5.2.3 Aluno C

O Aluno C apresenta deficiência motora e de fala e perda auditiva severa, utilizando gestos de mão e cabeça no ambiente escolar para se comunicar. Esse aluno usa alguns gestos de Libras e sinais caseiros, mas devido ao comprometimento motor nas mãos, nem todos os intérpretes de Libras compreendem suas intenções de comunicação. O aluno C tem sua capacidade de compreensão bem preservada e para a execução do experimento foi acompanhado por sua cuidadora e por sua intérprete de Libras. A coleta de dados para o treinamento do sistema e os testes de interação foram realizados em duas sessões diferentes. A opção "Aproximar" foi desativada. Inicialmente, o conjunto de dados criado por esse aluno foi composto por dez classes, com doze amostras cada. A Figura 5.7 apresenta exemplos de amostras geradas. Após o treinamento, durante o teste de interação, alguns gestos caracterizados com movimentos semelhantes foram confundidos pelo classificador, como os gestos "Sim", "Eu", "Mãe" e "Água". Os vídeos armazenados e as amostras relacionadas a esses gestos mostraram que os movimentos para estes gestos são muito semelhantes entre si, variando apenas a posição dos dedos ou da mão e, provavelmente por isso, o treinamento realizado não foi suficiente para o sistema reconhecê-los. Portanto, em outra sessão, um mês depois, um segundo conjunto de dados foi criado contendo apenas sete gestos com dezessete amostras cada, omitindo os gestos "Eu", "Mãe" e "Água", e

mantendo o gesto "Sim". Com o novo conjunto de dados, a maioria dos gestos foi corretamente reconhecida pelo sistema, exceto o gesto referente a "Banheiro" que não foi reconhecido em algumas situações.

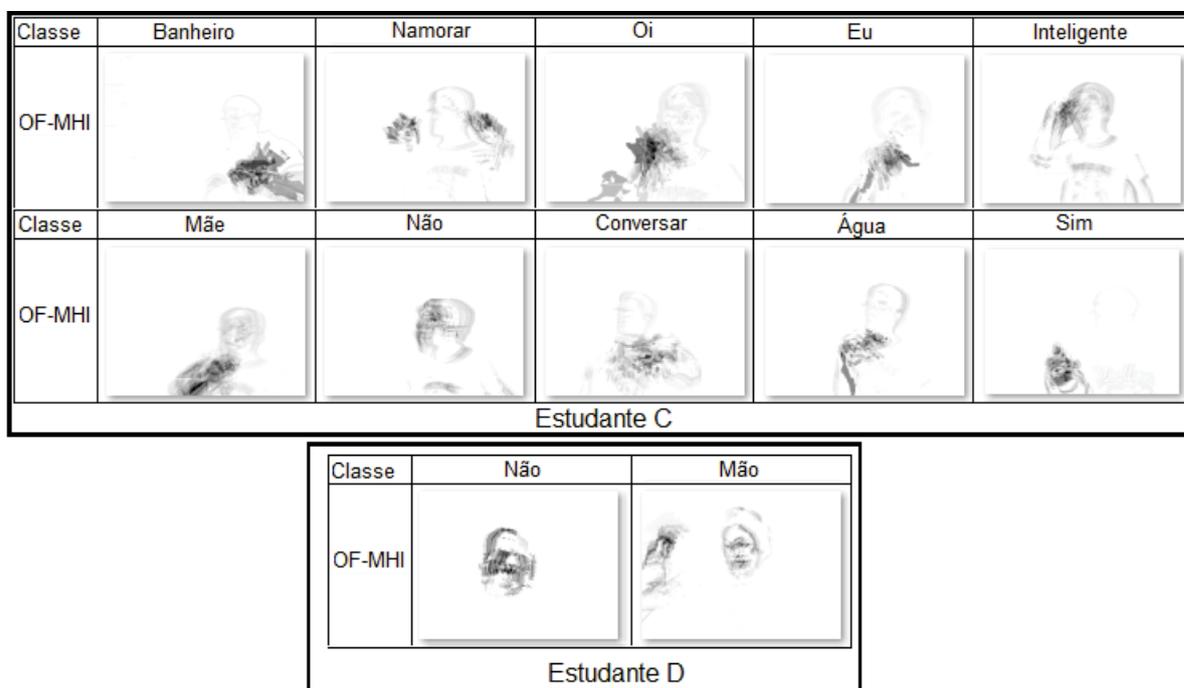


Figura 5.7: Exemplos de amostras de gestos executados pelos Alunos C e D (pessoas com dificuldades motora e de fala), registradas durante o Experimento 4 e representadas por meio de OF-MHI.

Posteriormente, o sistema foi configurado associando a classe "Sim" com a opção de confirmação do sistema, permitindo ao aluno testar a seleção de imagens nas pranchas de comunicação. Depois, o sistema foi configurado associando a classe "Inteligente" com a tecla ENTER, e a opção "simular teclado" foi habilitada. Em seguida, o aluno selecionou imagens em uma prancha de comunicação para escrever palavras em um editor de texto e simular o uso da tecla ENTER. Também foi possível selecionar uma figura em uma prancha de comunicação com opções de palavras-chave para pesquisar em um navegador da Internet, escrevendo diretamente na URL do navegador e simulando a tecla ENTER para procurar a palavra-chave (Figura 5.8). O teste de interação foi concluído nesta etapa. No entanto, seguindo a mesma estrutura de interação, outros gestos poderiam ser associados à tecla TAB para navegar entre os resultados da pesquisa e selecionar a página desejada usando a simulação da tecla ENTER.

5.5.2.4 Aluna D

A aluna D usa gestos de cabeça e expressões faciais no ambiente escolar para se comunicar, tem uma capacidade de compreensão muito preservada e pode mover voluntariamente o braço direito apesar de apresentar movimentos espásticos. O gesto que se refere a "Sim" é costumeiramente executado por meio de um movimento de sobrancelhas, no entanto, como a aluna apresenta muito movimento de cabeça associado, o sistema não conseguiu registrar corretamente o movimento dessa expressão facial. Para criar um conjunto de dados para essa aluna interagir com o sistema, a autora desse trabalho em conjunto com o professor que acompanhou o experimento optaram por capturar o movimento referindo-se a "Não" (movendo a cabeça para os dois lados) e o movimento referindo-se a "Mão" (movendo o braço direito).

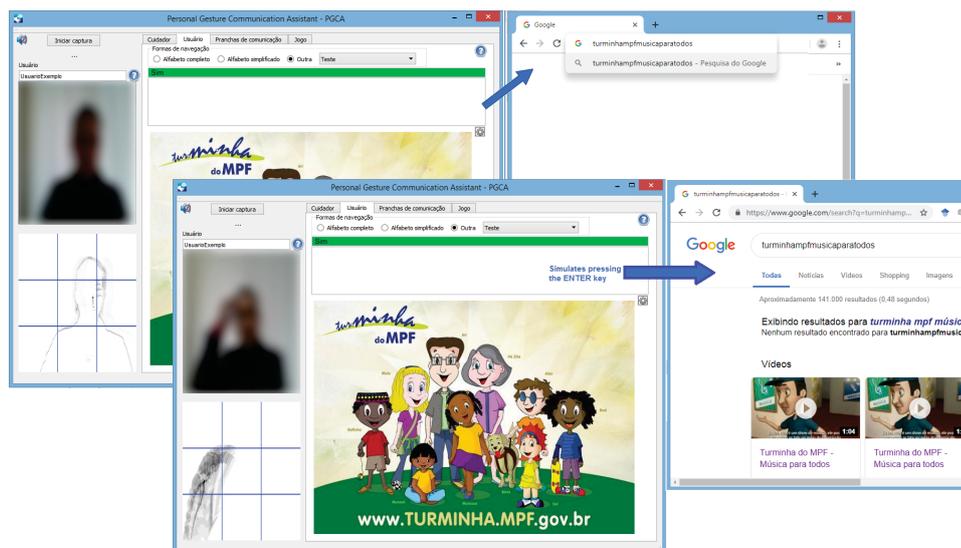


Figura 5.8: Prancha de comunicação usada para buscar um site do Ministério Público Brasileiro no navegador da Internet.

Fonte: Imagem utilizada na prancha disponível em Ministry (2019).

A coleta de dados foi realizada em uma sessão e os testes de interação em outra sessão cinco dias depois. Dez amostras foram consideradas válidas para cada classe no conjunto de dados. Após o treinamento, os dois gestos foram corretamente identificados pelo sistema no teste de interação. Posteriormente, o sistema foi configurado associando a classe "Mão" com a opção de confirmação do sistema e a classe "Não" com a tecla ENTER. Os mesmos testes de interação com o navegador da Internet e o editor de texto realizados pelos Alunos A e C foram realizados por essa aluna. No entanto, alguns movimentos involuntários ocorreram com o braço usado para fazer seleções e, inadvertidamente, o sistema selecionou itens nas pranchas de comunicação várias vezes. Segundo o professor acompanhante, o uso do braço para a aluna dar respostas afirmativas requer mais treinamento.

5.5.2.5 Aluna E

A Aluna E usa apenas pequenos gestos de cabeça e expressões faciais para expor suas intenções de comunicação no ambiente escolar. Os professores expressaram dúvidas sobre o nível de compreensão dessa aluna. Duas tentativas sem sucesso foram feitas para coletar dados com essa aluna em duas sessões diferentes, em dias diferentes. A estudante faz movimentos de cabeça muito restritos e com a mesma intenção de comunicar "Sim", em alguns momentos ela moveu a cabeça e em outros ela apenas sorriu. Segundo os professores, ao comer, a estudante expõe a língua para indicar que não quer uma comida específica. Mas durante o experimento, esse gesto referindo-se a "Não" não foi realizado pela aluna, mesmo após várias tentativas de a professora acompanhante fazer perguntas em busca de uma resposta negativa. Portanto, o experimento com a Aluna E foi encerrado sem a realização das tarefas previstas.

5.5.2.6 Aluno F

O Aluno F apresenta perda progressiva de suas funções motoras e utiliza gestos de mão e cabeça no ambiente escolar para se comunicar, apontando principalmente para objetos de interesse. Segundo os professores, o aluno tem uma capacidade de compreensão bem preservada. A autora desta pesquisa participou de algumas aulas com o aluno, observando a interação gestual

utilizada por ele. No entanto, não foi possível criar um conjunto de dados para esse aluno, embora duas tentativas tenham sido feitas para coletar dados em datas diferentes e com o apoio de diferentes professores. Em ambas as sessões, apesar de ter condições motoras, o aluno não demonstrou interesse em realizar os gestos quando solicitado. Após a segunda tentativa de captura dos gestos, a pesquisadora questionou o aluno se ele não queria ser filmado, e o aluno emitiu um som (considerando suas limitações de fala) que foi entendido como "Não". Assim, a sessão de coleta de dados foi encerrada sem a realização das tarefas previstas.

5.5.3 Avaliação de desempenho do sistema no reconhecimento de gestos

Nesse experimento, portanto, foram criados conjuntos de dados com gestos personalizados realizados por quatro alunos, e para avaliar o desempenho dos classificadores utilizando esses conjuntos de dados, aplicou-se o método K-fold Cross-Validation considerando 10 pastas, separando 90 % dos dados para treinamento e 10 % para teste. O número de dados de treinamento foi ampliado empregando Aumento de Dados, onde amostras adicionais foram criadas a partir de dados existentes. Depois de executar testes em todas as pastas, a precisão geral (média ponderada), desvio padrão, variância e o Cohen Kappa (medida estatística de concordância entre avaliadores) foram calculados para cada uma delas, como apresentado na Tabela 5.7. O número de amostras utilizadas para compor os conjuntos de dados dos alunos variou de acordo com o número de gestos capturados e considerados válidos pela autora desta pesquisa, que conduziu o experimento.

Tabela 5.7: Conjuntos de dados criados por alunos durante o Experimento 4 — comparação de desempenho de dois classificadores (SVM e CNN) e duas representações de movimento (MHI e OF-MHI): Precisão geral, Cohen Kappa, Desvio Padrão, Variância e Falsos Positivos.

Estudante		SVM				CNN			
		Prec.	Cohen k	Desv. Pad.	Var.	Prec.	Cohen k	Desv. Pad.	Var.
A	MHI	1	1	0	0	0.93	0.86	0.10	0.011
	OF-MHI	1	1	0	0	0.97	0.96	0.07	0.006
B	MHI	0.83	0.79	0.11	0.013	0.84	0.80	0.11	0.013
	OF-MHI	0.86	0.83	0.09	0.009	0.78	0.73	0.14	0.021
C1	MHI	0.84	0.82	0.13	0.020	0.68	0.64	0.15	0.026
	OF-MHI	0.87	0.86	0.13	0.019	0.73	0.70	0.10	0.012
C2	MHI	0.87	0.85	0.14	0.023	0.76	0.72	0.09	0.010
	OF-MHI	0.87	0.86	0.13	0.019	0.87	0.84	0.11	0.015
D	MHI	0.96	0.90	0.15	0.025	0.90	0.80	0.20	0.044
	OF-MHI	1	1	0	0	1	1	0	0

5.5.4 Discussão

O número de alunos que participou do experimento foi menor que o esperado, pois no primeiro contato com as escolas, vários alunos foram apresentados como possíveis candidatos (ou seja, pessoas com dificuldades motora e de fala). No entanto, à medida que a autora desta pesquisa conheceu melhor esses alunos, constatou que a minoria dos alunos com essas características não apresenta comprometimento cognitivo associado. Mesmo em situações em que há diagnósticos semelhantes (por exemplo, pessoas com paralisia cerebral), os próprios professores têm dúvidas sobre o nível de compreensão que cada aluno efetivamente possui. Então, muitos alunos inicialmente considerados potenciais participantes do experimento não

foram selecionados, pois era imperativo que o usuário tivesse um bom entendimento tanto para aceitar participar do experimento quanto para se beneficiar do uso intencional do sistema.

As observações obtidas a partir do experimento evidenciaram aspectos a serem aprimorados no sistema desenvolvido. A interface poderia ser melhorada adicionando uma sinalização mais visível (em comparação com a barra de progresso usada no momento) para indicar ao usuário se o sistema considerou o gesto como concluído ou se está tentando reconhecê-lo. As mensagens textualmente apresentadas pela interface poderiam ser traduzidas para Libras e apresentadas por meio de voz sintetizada, a fim de facilitar o entendimento por usuários surdos e por usuários não alfabetizados. Para os alunos que têm a capacidade de executar um número restrito de gestos voluntários, é necessário maior atenção do cuidador para iniciar e finalizar o processo de captura por meio da interface. Nos testes realizados com o Aluno A, quando ele estava observando os colegas na turma, movimentando a cabeça naturalmente sem a intenção de comunicar algo, o sistema acabou identificando o movimento como "Sim" ou "Não", gerando respostas desnecessárias e incorretas. Isso ocorreu porque o sistema permaneceu com o processo de captura iniciado mesmo depois do professor ter encerrado a interação com ele. Como o aluno não possui outros gestos que possam ser associados à funcionalidade de início e fim de captura, esse controle deveria ser feito pelo cuidador.

No caso do Aluno C, a quantidade distinta de sinais que ele é capaz de fazer é bastante grande, no entanto, alguns dos sinais geraram representações de movimento muito semelhantes que acabaram confundindo o classificador. Para interfaces baseadas em gestos, a precisão (altos positivos e baixos índices de falsos positivos) precisa ser assegurada, ao mesmo tempo em que mantém a sensação natural de comunicação interpessoal (Leo et al., 2017). Para minimizar falsos positivos, o sistema PGCA possui um nível de confiança que pode ser personalizado para minimizar atribuições falsas. Além disso, os usuários podem criar conjuntos de dados compostos por gestos que são bastante semelhantes entre si e, nesses casos, se o sistema não puder reconhecer esses gestos corretamente, o sistema pode sugerir uma nova coleta de dados para esses gestos ou sugerir que apenas um dos gestos semelhantes faça parte do conjunto de dados.

Como diferentes sessões ocorreram em dias diferentes, alguns alunos realizaram gestos com velocidade e intensidade diferentes, sugerindo que é interessante coletar amostras em dias diferentes para gerar mais informações a serem aprendidas pelo classificador. Além disso, considerando a situação observada na tentativa de capturar amostras do Aluno F, tornar a interface do sistema mais atrativa, talvez usando um jogo sério que exija a execução de movimentos semelhantes para seu uso, foi visto como uma possibilidade de estimular o interesse dos alunos e promover engajamento e motivação para treinar e usar o sistema.

Embora o processo de calibração tenha sido uma preocupação, pois requer um posicionamento específico dos usuários na frente da câmera, nenhum problema nessa etapa foi percebido. Todos os alunos tinham condições motoras para mover a cabeça e quando o nariz do usuário estava posicionado perto do centro da tela (mesmo em movimento), o requisito da etapa de calibração era facilmente alcançado.

A inclusão de pontos de referência faciais como pontos a serem rastreados via fluxo óptico na representação de movimento por meio de OF-MHI foi feita após a entrevista com alguns professores, que relataram que alguns dos alunos selecionados poderiam realizar movimentos faciais muito discretos, como levantar as sobrancelhas, piscar os olhos ou mostrar a língua. Essa situação não foi prevista na primeira versão do sistema, mas a mudança implementada possibilitou o rastreamento satisfatório desses pequenos movimentos.

A Transferência de Aprendizado provou ser eficiente para o estudo em questão, porque mesmo uma rede treinada com um conjunto de dados composto por imagens coloridas pôde

ser personalizado para reconhecer com sucesso as imagens em escala de cinza utilizadas pelo sistema PGCA. No entanto, o classificador com base em SVM associado à representação de movimento OF-MHI apresentou um desempenho ligeiramente melhor no reconhecimento de gestos dos conjuntos de dados dos alunos.

Esse experimento permitiu avaliar a viabilidade técnica da proposta e identificar ainda algumas melhorias necessárias para o sistema. A condução de novos experimentos com o público-alvo foi vista como importante para validar os resultados do Experimento 4 e também para avaliar a efetividade do uso do sistema no cotidiano de pessoas com necessidades de CAA. Após a condução dos experimentos, entendeu-se que o classificador que apresentar melhores resultados em novos experimentos é o candidato natural a ser adotado para a versão final do *software*.

O sistema PGCA em sua versão atual pode não produzir resultados satisfatórios para o reconhecimento de linguagens de sinais complexas, compostas de algumas centenas de gestos distintos, com poucas variações entre eles. No entanto, a metodologia proposta nesta tese de doutorado, que guiou o desenvolvimento do sistema PGCA poderia ser usada para desenvolver outras abordagens, levando em consideração diferentes estratégias de treinamento e reconhecimento, e as características dos gestos empregados como meio de comunicação pelo público-alvo.

5.6 EXPERIMENTO 5: USO DO SISTEMA PGCA POR ALUNOS COM DIFICULDADES MOTORA E DE FALA INCLUINDO ABORDAGEM BASEADA EM JOGO

Muitas pessoas com deficiência ainda são excluídas da participação total no mundo dos jogos. Jogadores com deficiência precisam usar os recursos da TA para jogar jogos acessíveis, e é importante destacar que não se trata de "apenas jogar um jogo" (que é divertido e oferece benefícios de aprendizado), mas uma questão de inclusão na vida real (Ossmann et al., 2008). Para contribuir nessa direção, foi incluída no sistema PGCA a Área de Jogo, que permite ao usuário interagir com um jogo simples por meio de gestos.

Um novo experimento foi conduzido com alunos que possuem dificuldades motora e de fala (Experimento 5), para avaliar as melhorias inseridas no sistema. O Experimento 5 teve como objetivo validar os dados obtidos em experimento anterior (Experimento 4), e agregar conhecimento sobre o processo de interação de pessoas com deficiência com um jogo. Esse experimento permitiu avaliar o sistema após a implementação das seguintes melhorias: a. possibilidade de reproduzir todas as mensagens apresentadas pelo sistema via sintetizador de voz; b. inclusão de bordas com cores distintas para tornar o estado atual do sistema mais visível; c. inclusão de botões de ajuda para descrever os principais recursos do sistema; d. inclusão de uma abordagem interativa de jogo baseado em gestos. Um dos principais objetivos deste experimento foi analisar de maneira padronizada a precisão do sistema no reconhecimento de gestos e o esforço necessário para que os usuários pudessem concluir as tarefas predefinidas, cada um usando os gestos que é capaz de executar.

Nesse experimento foi empregada apenas a representação de movimento OF-MHI por ter apresentado resultados melhores que MHI convencional no Experimento 4. Considerando as dificuldades identificadas e as lições aprendidas com a execução do primeiro experimento com o público-alvo, para o Experimento 5, foram selecionados apenas os alunos que apresentaram boa capacidade de compreensão e interesse em participar das tarefas no experimento anterior. Dos sete alunos selecionados previamente, um não foi autorizado pela família, dois não foram capazes de gerar conjuntos de dados na primeira coleta de dados (por diferentes razões) e um deles mostrou muita dificuldade em entender as tarefas solicitadas. Assim, para o Experimento

5, apenas três alunos com comprometimento motor e de fala foram convidados a participar, conforme descrito na seção a seguir.

5.6.1 Participantes

Todos os alunos participantes desse experimento participaram também do Experimento 4. São caracterizados como pessoas com paralisia cerebral, com diferentes níveis de necessidades especiais, conforme apresentado na Tabela 5.8.

Tabela 5.8: Características dos alunos com dificuldades motora e de fala participantes do Experimento 5.

Aluno	Sexo*	Idade	Relatório médico
A	M	18 anos	Paralisia cerebral devido a sequelas durante o parto.
B	M	38 anos	Quadriplegia com componente de atetose, perda auditiva sensorioneural bilateral.
C	F	20 anos	Paralisia pseudobulbar. Hipotonia generalizada e hiperreflexia.

*Sexo: M - Masculino; F- Feminino.

5.6.2 Execução de tarefas por participante

O principal objetivo desse experimento foi analisar a interação gestual personalizada fornecida pelo sistema PGCA para interagir com pranchas de comunicação e também para jogar um jogo simples. Os resultados desse experimento foram comparados aos resultados do experimento anterior (realizado com os mesmos alunos) a fim de validar os resultados obtidos.

Para testes de interação, o classificador com base em SVM foi usado porque apresentou resultados ligeiramente melhores que o classificador com base em CNN no experimento anterior. Mesmo assim, posteriormente, o classificador com base em CNN também foi avaliado.

A validação das amostras por meio de algoritmo de correspondência (similaridade) de imagens foi utilizada nesse experimento. Assim, ao identificar uma amostra possivelmente inválida, o algoritmo sugere ao usuário que o gesto não foi realizado de maneira semelhante ao primeiro (considerado como base), e o cuidador então deve decidir se a amostra deve ser mantida no conjunto de dados ou descartada. A autora desta pesquisa acompanhou o experimento e ajudou os cuidadores nessa etapa, filtrando as amostras geradas. Nesse experimento, para todos os alunos, o sistema foi configurado para usar a opção "Aproximar" que permite rastrear com maiores detalhes movimentos faciais.

No primeiro experimento com o público-alvo, conjuntos de dados foram criados em uma sessão e testes de interação em uma segunda sessão. Nesse segundo experimento com o público-alvo, uma única sessão foi suficiente. Cada conjunto de dados foi composto por quinze amostras originais por classe, a fim de padronizar o número de amostras disponíveis. Foram utilizadas as mesmas classes registradas pelos alunos no experimento anterior.

Para executar o teste de interação na mesma sessão em que o conjunto de dados foi criado, o sistema foi avaliado usando o método Holdout, por ser mais rápido e também por fornecer uma boa noção do desempenho do sistema no reconhecimento de gestos. Assim, foram gerados dois modelos para reconhecimento de gestos: (i) para a etapa de avaliação do sistema, onde o conjunto de dados gerado foi dividido em dois subconjuntos, com 2/3 dos dados no conjunto de treinamento e 1/3 dos dados no conjunto de testes; (ii) para a etapa de interação do usuário com o sistema, onde um modelo foi treinado usando todas as amostras disponíveis no conjunto de dados. Nos dois modelos, o número de dados de treinamento foi expandido via Aumento de Dados. Posteriormente, após o término da sessão (por demandar mais tempo de execução), foi utilizado o método de validação cruzada K-fold, considerando 10 pastas, separando 90 % dos dados para treinamento e 10 % para o teste, para gerar uma estimativa geral mais

precisa sobre a capacidade do sistema em reconhecer os gestos capturados. Detalhes da execução desse experimento com os alunos são apresentados a seguir. Algumas imagens apresentam os usuários interagindo com o sistema, mas o rosto dos participantes foi desfocado para preservar seu anonimato.

5.6.2.1 Aluno A

O aluno A está matriculado em uma instituição educacional específica para alunos com necessidades especiais (APAE). Esse aluno usa apenas dois gestos com a cabeça no ambiente escolar para se comunicar com seus colegas e professores, que se referem a "Sim" (movendo a cabeça para trás) e "Não" (movendo a cabeça para os dois lados). Portanto, o conjunto de dados criado foi composto de duas classes, com quinze amostras por classe.

Esse aluno tem dificuldade em manter o pescoço ereto e, no primeiro experimento, usou um travesseiro para apoiar o pescoço, porque era o apoio que ele vinha usando em sala de aula. No segundo experimento, o professor acompanhante ressaltou que a mobilidade física do aluno piorou um pouco (porque ele é capaz de segurar ainda menos o pescoço), e o travesseiro usado anteriormente não tem mais efeito. Assim, o aluno tem usado como suporte em sua cadeira de rodas um rolo de tecido, mas mesmo assim, durante o experimento, foi necessário ajustar a postura do aluno várias vezes. O professor comentou que, às vezes, para assistir filmes, por exemplo, tem usado um suporte de tecido preso à testa do aluno, mas isso bloqueia sua capacidade de interação. Possivelmente devido à situação descrita, as amostras capturadas nos Experimentos 4 e 5 para os mesmos gestos apresentam diferenças significativas. A Figura 5.9 (A) apresenta exemplos de amostras geradas pelo aluno A para compor seu conjunto de dados nos Experimentos 4 e 5.



Figura 5.9: Exemplos de amostras de gestos executados pelo Aluno A (pessoa com dificuldade motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.

Após o treinamento, o sistema foi avaliado usando a abordagem Holdout, mostrando 80% de precisão no reconhecimento de gestos. A primeira etapa do teste de interação com a interface foi referente ao reconhecimento dos gestos para os quais o sistema foi treinado, utilizando a Área do Usuário: o gesto "Não" foi corretamente identificado pelo sistema em todas as tentativas realizadas voluntariamente pelo aluno, mas o gesto "Sim" não foi identificado

algumas vezes, principalmente devido à dificuldade do aluno em se reposicionar corretamente após mover a cabeça para trás. Percebeu-se também que, às vezes, o usuário acabava abaixando a cabeça para visualizar a resposta da interface, gerando uma representação de movimento mais próxima do gesto "Não".

Posteriormente, foi utilizada a Área de Jogo, configurando a classe "Sim" para mover a tartaruga para a direita e a classe "Não" para mover a tartaruga para cima. Após algumas tentativas malsucedidas, a autora desta tese demonstrou o uso do jogo por meio de interação gestual, usando um conjunto de dados criado por ela e, em seguida, o aluno entendeu o funcionamento do jogo e conseguiu alcançar o objetivo esperado.

Na sequência, o sistema foi configurado associando a classe "Sim" com a opção "confirmar" do sistema, permitindo ao aluno realizar a seleção de imagens nas pranchas de comunicação disponíveis na Área do Usuário. A Figura 5.10 mostra um exemplo de teste de interação com uma prancha de comunicação de figuras, realizado por todos os alunos participantes.

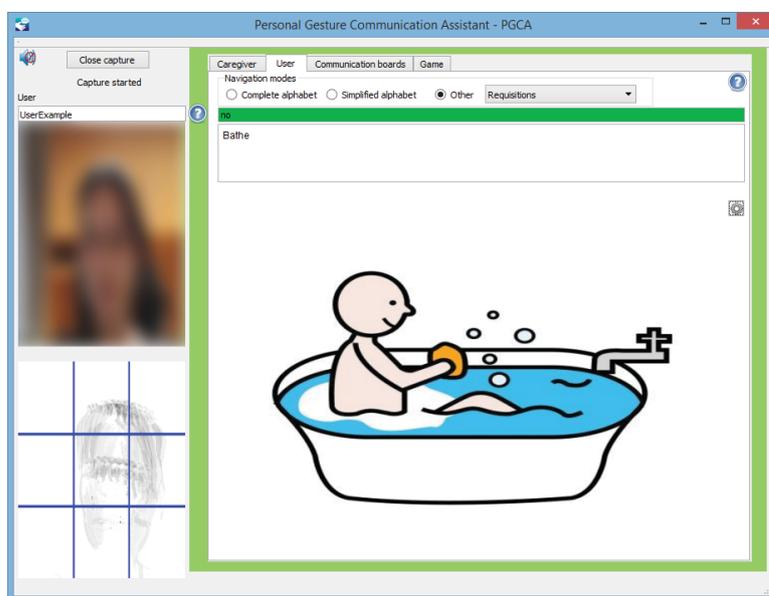


Figura 5.10: Exemplo de seleção de imagem em prancha de comunicação de figuras usando um gesto de cabeça. Esse teste de interação foi realizado por todos os alunos participantes do Experimento 5, cada um usando o seu conjunto de gestos personalizados.

Fonte: Imagem utilizada na prancha disponível em Rodrigo e Corral (2013).

A Figura 5.11 apresenta o Aluno A interagindo com a interface do sistema PGCA, jogando o jogo. O processo de coleta de dados para compor o conjunto de dados no Experimento 5 levou pouco mais de dez minutos para ser concluído.



Figura 5.11: Aluno A (pessoa com dificuldades motora e de fala) interagindo com a interface do sistema PGCA, jogando o jogo, como parte das tarefas executadas durante o Experimento 5.

5.6.2.2 Aluno B

O aluno B está matriculado em uma escola inclusiva, onde alunos com necessidades especiais também podem frequentar aulas regulares (Centro Estadual de Educação Básica para Jovens e Adultos). O aluno apresenta comprometimento motor e de fala e perda auditiva severa, utilizando gestos realizados com a mão e a cabeça no ambiente escolar para se comunicar. Para a execução do experimento, ele foi acompanhado por sua intérprete de Libras e sua cuidadora. O conjunto de dados criado por esse aluno foi composto por sete classes, referentes a gestos baseados em movimentos comumente usados pelo aluno no ambiente escolar, e que já haviam sido utilizados para criar um conjunto de dados no Experimento 4.

A Figura 5.12 mostra exemplos de amostras geradas pelo Estudante B para compor seu conjunto de dados nos Experimentos 4 e 5. No segundo experimento, o aluno apresentou um pouco de dificuldade em fazer movimentos mais amplos e em levantar os braços. Sua intérprete de Libras comentou que recentemente o aluno demonstrou fraqueza considerável e habilidades motoras diminuídas.

Após o treinamento, o sistema foi avaliado usando a abordagem Holdout, mostrando 74% de precisão no reconhecimento de gestos. Durante o teste de interação, alguns gestos foram facilmente reconhecidos pelo sistema, como "Oi", "Conversa", "Namorar", "Sim", no entanto, alguns gestos ("Não", "Inteligente", "Banheiro") tiveram que ser executados várias vezes antes de serem corretamente reconhecidos pelo sistema. Foi possível observar algum cansaço e dificuldade do aluno em executar os gestos da mesma forma como haviam sido executados durante a etapa de criação do conjunto de dados.

Posteriormente, o sistema foi configurado associando a classe "Sim" à opção "confirmar" do sistema, permitindo ao aluno testar a seleção de imagens em pranchas de comunicação disponíveis na Área do Usuário. A seguir, foi utilizada a Área de Jogo, configurando a classe "Oi" para mover a tartaruga para a direita e a classe "Sim" para mover a tartaruga para cima.

A autora desta tese conduziu o experimento e explicou ao aluno o funcionamento do jogo, usando gestos e a ajuda da intérprete de Libras. Como o aluno fez algumas tentativas e a tartaruga não se moveu, a autora desta tese reproduziu o gesto que estava sendo realizado pelo aluno, fazendo com que a tartaruga se movesse. A partir de então, o aluno começou a usar seus

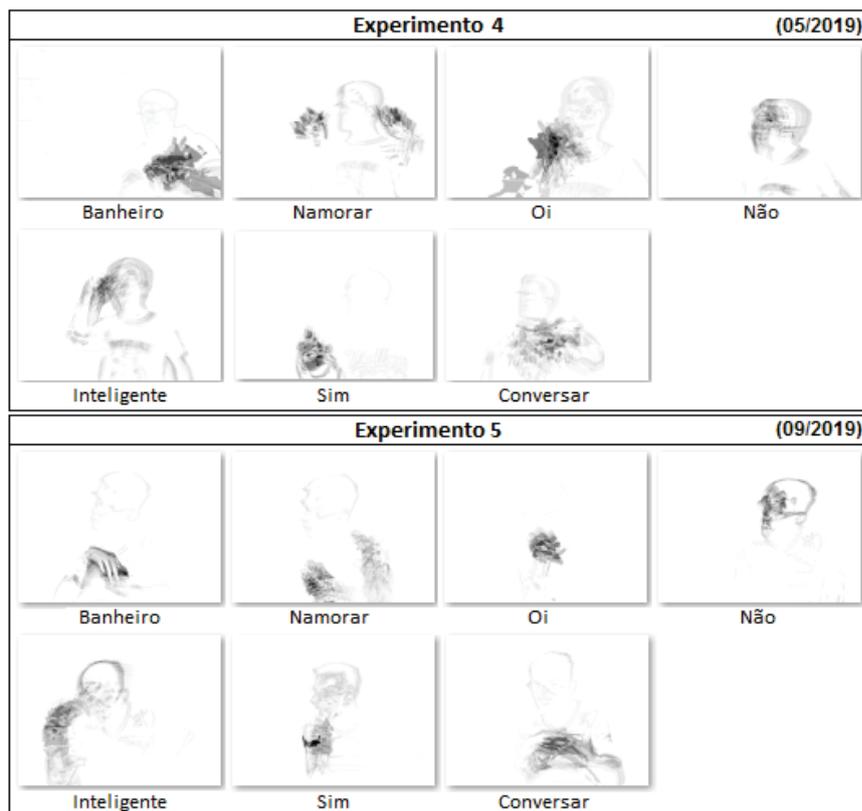


Figura 5.12: Exemplos de amostras de gestos executados pelo Aluno B (pessoa com dificuldades motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.

movimentos e conseguiu alcançar o objetivo do jogo. Durante a interação, o aluno se moveu de diferentes maneiras e diferentes gestos foram reconhecidos pelo sistema, além daqueles que foram configurados para mover a tartaruga.

A Figura 5.13 apresenta o aluno após concluir a interação com a interface do sistema e completar o objetivo do jogo. O processo de coleta de dados para compor o conjunto de dados do Aluno B no segundo experimento levou pouco mais de trinta minutos para ser concluído.

5.6.2.3 Aluna C

A aluna C está matriculada em uma escola inclusiva, onde alunos com necessidades especiais também podem frequentar aulas regulares (Colégio Estadual - Ensino fundamental e médio). Essa aluna usa gestos de cabeça e expressões faciais para expor suas intenções de comunicação no ambiente escolar. O gesto referente a "Sim" é realizado por uma elevação das sobrancelhas, no entanto, como a aluna possui muitos movimentos de cabeça associados, o sistema não conseguiu registrar corretamente o movimento dessa expressão facial. Assim, como já feito no Experimento 4, para criar um conjunto de dados para essa aluna interagir com o sistema, foram capturados o movimento referente a "Não" (movendo a cabeça para os dois lados) e o movimento referente a "Mão" (movendo o braço direito, apesar de apresentar muitos movimentos espásticos). A Figura 5.14 apresenta exemplos de amostras geradas pela Aluna C para compor seu conjunto de dados nos Experimentos 4 e 5.

Após o treinamento, o sistema foi avaliado usando a abordagem Holdout, mostrando 100% de precisão no reconhecimento de gestos. Os dois gestos foram identificados corretamente

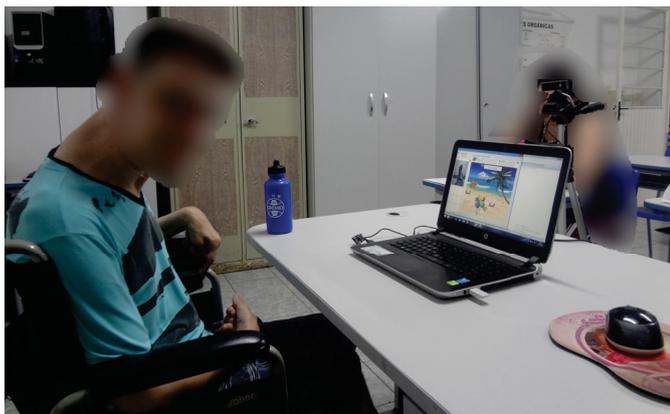


Figura 5.13: Aluno B depois de interagir com a interface do sistema PGCA, e atingir o objetivo do jogo. Esta é uma das tarefas executadas durante o Experimento 5.

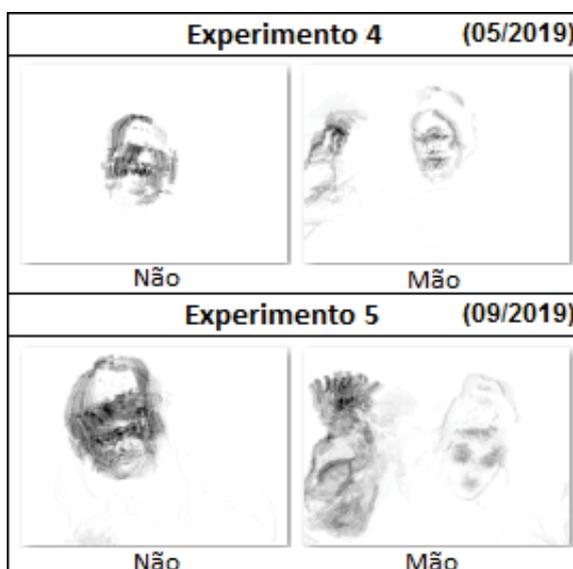


Figura 5.14: Exemplos de amostras de gestos executadas pela Aluna C (pessoa com dificuldades motora e de fala), registradas durante os Experimentos 4 e 5 e representadas por meio de OF-MHI. No canto superior direito é apresentado o mês e ano em que os experimentos foram conduzidos.

pelo sistema no teste de interação. Eventualmente, a aluna apresentou uma mudança significativa de postura, o que dificultou o reconhecimento de alguns movimentos pelo sistema.

A seguir, foi utilizada a Área de Jogo, configurando a classe "Não" para mover a tartaruga para a direita e a classe "Mão" para mover a tartaruga para cima. A pesquisadora que conduziu o experimento explicou à aluna o funcionamento do jogo, e a aluna jogou o jogo usando seus dois gestos personalizados, sem apresentar qualquer dificuldade para entender o funcionamento do jogo. No entanto, percebeu-se que o uso de movimentos do braço direito exigia um grande esforço físico da aluna.

Posteriormente, o sistema foi configurado associando a classe "Não" à opção de navegação "próxima" e a classe "Mão" à opção "confirmar" do sistema. A aluna conseguiu usar o gesto "Não" para navegar entre as imagens de uma prancha de comunicação e o gesto "Mão" para selecionar as imagens desejadas.

A Figura 5.15 apresenta a aluna C interagindo com a interface do sistema PGCA, jogando o jogo. O processo de coleta de dados para compor o conjunto de dados da Aluna C no segundo experimento levou pouco mais de sete minutos para ser concluído.

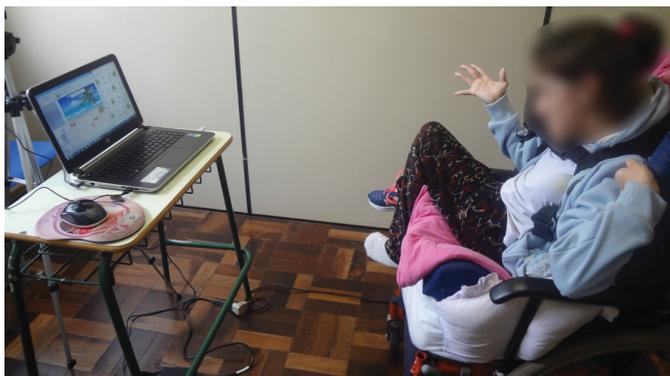


Figura 5.15: Aluna C interagindo com a interface do sistema PGCA, jogando o jogo, como parte das tarefas executadas durante o Experimento 5.

5.6.3 Avaliação de desempenho do sistema no reconhecimento de gestos

Nesse experimento, o foco da avaliação foi direcionado para os seguintes aspectos: 1) desempenho dos classificadores no reconhecimento de gestos personalizados e tempo de execução gasto para as etapas de treinamento e teste; 2) esforço necessário para que os usuários alcancem o objetivo do jogo; e 3) usabilidade do sistema.

A Tabela 5.9 mostra o tempo necessário para executar as etapas de treinamento e teste, considerando os métodos Holdout e K-Fold, pelos classificadores com base em SVM e CNN.

Tabela 5.9: Tempo aproximado de execução dos classificadores para as etapas de treinamento e teste, usando os métodos Holdout e K-Fold no Experimento 5.

Estudante	Classificador com base em SVM				Classificador com base em CNN			
	Holdout		K-Fold		Holdout		K-Fold	
	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
A	45s	6s	5m	16s	29m	58s	5h	3m
B	8m	7s	56m	26s	29m	5m	12h	9m
C	34s	4s	4m	4s	34m	51s	5h	3m

s: segundos; m: minutos; h: horas.

Para avaliar o desempenho dos classificadores no reconhecimento de gestos dos conjuntos de dados foi utilizado o método de validação cruzada K-fold e, após a execução de testes em todas as pastas, foi calculada a precisão geral, Cohen Kappa, desvio padrão, variância e falsos positivos, como apresentado na Tabela 5.10. Os classificadores consideraram decisões válidas quando tinham 50% de certeza ou mais de que uma amostra pertenceria a uma classe específica (nível de confiança).

Para avaliar o esforço necessário para os usuários atingirem o objetivo do jogo, observou-se o tempo necessário para cada usuário realizar a tarefa e o número de gestos realizados. As Tabelas 5.11, 5.12 e 5.13 mostram o progresso do jogo e os gestos reconhecidos durante a interação dos alunos com o jogo, considerando a interação na qual os usuários apresentaram melhor desempenho, segundo informações registradas no arquivo de *log*.

Tabela 5.10: Conjuntos de dados criados por alunos durante a execução dos Experimentos 4 e 5 — desempenho de classificadores usando validação cruzada K-fold: Precisão, Kappa de Cohen, Desvio padrão, Variância e Falsos positivos.

	Classificador	Método	Conjunto de dados do aluno		
			A	B	C
Experimento 4	Com base em SVM	Precisão	1	0.87	1
		Cohen k	1	0.85	1
		Desv. Pad.	0	0.080	0
		Variância	0	0.007	0
		Falsos pos.	0	0.002	0
	Com base em CNN	Precisão	0.97	0.82	1
		Cohen k	0.94	0.79	1
		Desv. Pad.	0.030	0.147	0
		Variância	0.001	0.024	0
		Falsos pos.	0.002	0.011	0
Experimento 5	Com base em SVM	Precisão	0.87	0.89	1
		Cohen k	0.73	0.87	1
		Desv. Pad.	0.200	0.136	0
		Variância	0.044	0.020	0
		Falsos pos.	0.004	0.006	0
	Com base em CNN	Precisão	0.90	0.69	0.92
		Cohen k	0.80	0.64	0.92
		Desv. Pad.	0.160	0.153	0.200
		Variância	0.028	0.026	0.044
		Falsos pos.	0.004	0.018	0.002

Durante a interação com a Área de Jogo, o tempo gasto pelo sistema para reconhecer cada gesto variou muito, pois depende das variações e limitações de cada usuário na execução dos movimentos. No entanto, a partir dos registros armazenados no arquivo de *log*, é possível ver que alguns gestos foram identificados em menos de dois segundos, o que mostra a eficácia da interface em fornecer uma interação gestual natural.

Tabela 5.11: Gestos reconhecidos pelo sistema durante a interação gestual personalizada do aluno A com o jogo e ações desencadeadas durante o Experimento 5.

Gesto	Mover tartaruga	Tempo (h:m:s:ms)	Estado da tartaruga	Som	Jogo
Sim	Para Direita	08:29:34:295	A	Salto	Em progresso
Não	Para Cima	08:29:48:646	A	Salto	Em progresso
*	-	08:29:57:163	A	-	Em progresso
Não	Para Cima	08:30:00:701	A	Salto	Em progresso
Não	Para Cima	08:30:03:942	C	Mordida	Em progresso
Não	Para Cima	08:30:08:900	C	Mordida	Em progresso
Não	Para Cima	08:30:13:077	D	Aplausos	Concluído

* Gesto não identificado.

Tempo gasto: 00:00:38:782

Tabela 5.12: Gestos reconhecidos pelo sistema durante a interação gestual personalizada do Aluno B com o jogo e ações desencadeadas durante o Experimento 5.

Gesto	Mover tartaruga	Tempo (h:m:s:ms)	Estado da tartaruga	Som	Jogo
*	-	14:28:20:050	A	-	Em progresso
Conversar	Para direita	14:28:30:236	A	Salto	Em progresso
*	-	14:28:31:947	A	-	Em progresso
*	-	14:28:32:934	A	-	Em progresso
*	-	14:28:34:672	A	-	Em progresso
*	-	14:28:36:197	A	-	Em progresso
Namorar	-	14:28:38:178	A	-	Em progresso
*	-	14:28:40:671	A	-	Em progresso
Namorar	-	14:28:42:120	A	-	Em progresso
*	-	14:28:43:921	A	-	Em progresso
Namorar	-	14:28:45:381	A	-	Em progresso
Sim	Para cima	14:28:48:481	A	Salto	Em progresso
Namorar	-	14:28:52:521	A	-	Em progresso
Namorar	-	14:28:54:394	A	-	Em progresso
Namorar	-	14:28:55:923	A	-	Em progresso
Sim	Para cima	14:29:00:924	A	Salto	Em progresso
Não	-	14:29:02:013	A	-	Em progresso
Sim	Para cima	14:29:04:549	C	Mordida	Em progresso
*	-	14:29:06:958	C	-	Em progresso
Sim	Para cima	14:29:09:471	C	Mordida	Em progresso
Sim	Para cima	14:29:11:454	D	Aplausos	Concluído

* Gesto não identificado. Tempo gasto: 00:00:51:404

Tabela 5.13: Gestos reconhecidos pelo sistema durante a interação gestual personalizada da Aluna C com o jogo e ações desencadeadas durante o Experimento 5.

Gesto	Mover tartaruga	Tempo (h:m:s:ms)	Estado da tartaruga	Som	Jogo
Mão	Para cima	08:18:07:324	A	Salto	Em progresso
Mão	Para cima	08:18:13:969	A	Salto	Em progresso
Mão	Para cima	08:18:22:322	C	Mordida	Em progresso
Mão	Para cima	08:18:26:228	C	Mordida	Em progresso
Mão	Para cima	08:18:29:653	D	Aplausos	Concluído

Tempo gasto: 00:00:22:329

5.6.4 Avaliação de usabilidade do sistema

As medidas de desempenho apresentadas para o sistema PGCA são complementadas por auto-relatórios coletados usando uma avaliação de usabilidade padronizada. Para entender melhor a percepção dos profissionais que acompanharam a execução do experimento, os professores participantes e acompanhantes, cuidadora e intérprete de Libras preencheram um questionário de avaliação de usabilidade do sistema, por meio do qual puderam opinar sobre o sistema e o interesse de usá-lo no futuro. Os profissionais participantes desta avaliação também participaram da etapa de entrevistas com especialistas do domínio. A Escala de Usabilidade do Sistema (SUS) (Brooke, 1996) foi empregada, por ser rápida, simples e devido à sua psicométrica desejável, incluindo alta confiabilidade e validade (Lewis e Sauro, 2009).

O questionário do modelo SUS possui 10 perguntas para medir a usabilidade do sistema, em que perguntas ímpares são enquadradas de forma positiva e as outras perguntas pares de forma negativa. Cada pergunta deve ser classificada entre 1 e 5 (escala do tipo Likert, onde concordo totalmente = 5 e discordo totalmente = 1) e a pontuação total de cada participante é

multiplicada por 2,5 para obter a faixa de pontuação entre 0 e 100. Finalmente, é considerada a pontuação média de todos os participantes. As pontuações totais são classificadas da seguinte forma quanto a usabilidade do sistema: (a) 0 a 64 = Não aceitável, (b) 65 a 84 = Aceitável e (c) 85 a 100 = Altamente aceitável.

A Tabela 5.14 apresenta o questionário SUS usado para avaliar a usabilidade do sistema desenvolvido. Uma questão discursiva foi incluída no questionário padrão para que os participantes pudessem expressar suas percepções ou sugestões, caso tivessem interesse.

Tabela 5.14: Questionário SUS usado para avaliar a usabilidade do sistema PGCA após a execução de tarefas pelo público-alvo no Experimento 5. O questionário foi preenchido por profissionais da área de educação especial que acompanharam a condução do experimento.

		Escala de Usabilidade do Usuário (SUS)				
		Discordo fortemente			Concordo fortemente	
		1	2	3	4	5
Q1	Gostaria de usar esse aplicativo frequentemente.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q2	Achei que o aplicativo era desnecessariamente difícil de usar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q3	Achei o aplicativo bem simples de se usar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q4	Penso que iria precisar de apoio técnico para usar o aplicativo.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q5	Achei as várias funcionalidades do aplicativo bem integradas.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q6	Penso que havia inconsistências no aplicativo.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q7	Imagino que a maioria das pessoas aprenda rapidamente a usar o aplicativo.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q8	Achei que o aplicativo não era trivial de usar.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q9	Senti-me muito confiante ao usar o aplicativo.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q10	Preciso aprender muito antes de poder usar este aplicativo.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q11	Você teria alguma sugestão de melhoria para o aplicativo ou algum comentário com relação às tarefas realizadas?					

A Figura 5.16 (A) demonstra a pontuação da Escala de Usabilidade do Sistema por cinco participantes e 5.16 (B) a pontuação geral (para 10 perguntas) de cada participante. Questões ímpares têm valores altos e questões pares têm valores baixos. A pontuação média do SUS de todos os participantes é 81, o que é aceitável conforme sugerido na descrição do SUS.

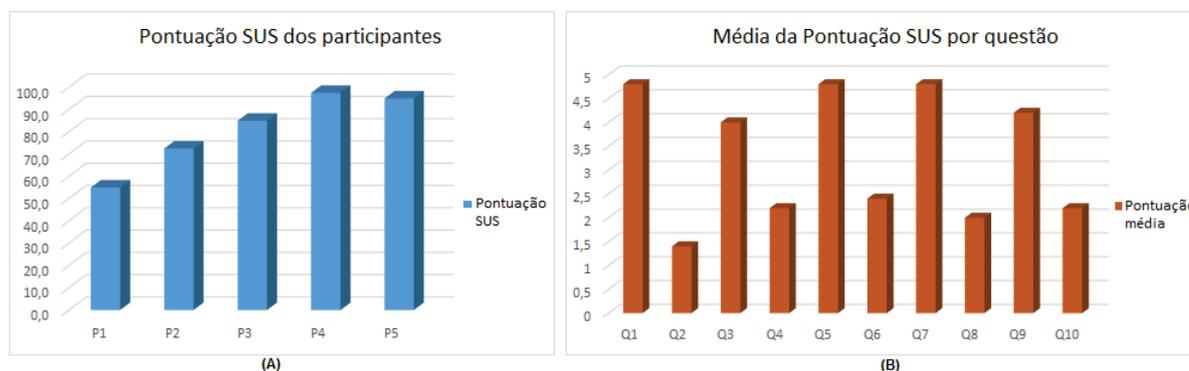


Figura 5.16: Pontuação obtida com a avaliação de usabilidade do sistema PGCA utilizando o questionário SUS: (A) Pontuação por participante (P1 a P5); (B) Pontuação média por questão (Q1 a Q10). A pontuação média do SUS de todos os participantes foi 81, que é aceitável.

Dois comentários escritos no questionário SUS, por professores participantes do experimento, mostram o interesse de ambos em usar a ferramenta no ambiente escolar: (i) "*Muito interessante, eu gostaria no futuro, se os resultados forem bons, de usar o sistema para trabalhar*

com os alunos da escola. Se o programa atingir seus objetivos."; (ii) *"Vejo que a aplicação contribuiria significativamente no processo de ensino-aprendizagem, sendo uma ferramenta de baixo custo para pessoas que usam esse tipo de recurso e extremamente importante para profissionais que trabalham com esse público, facilitando as adaptações de conteúdo".*

5.6.5 Discussão

Os experimentos conduzidos com o público-alvo (Experimentos 4 e 5) permitiram observar a viabilidade da proposta desta tese de doutorado. Ou seja, é possível treinar um sistema (criado a partir da metodologia MyPGI) usando dispositivos de baixo custo e um pequeno número de amostras geradas por pessoas com deficiência motora e de fala, para viabilizar a CAA usando interação gestual personalizada.

O desempenho apresentado pelo sistema PGCA no reconhecimento de gestos foi considerado satisfatório, em ambos os Experimentos 4 e 5, porque a precisão média obtida em todos os conjuntos de dados foi próxima ou superior a 0,9, principalmente quando utilizado o classificador com base em SVM. Baixo desvio padrão e variância foram observados, e os valores de kappa indicaram concordância substancial ou mais.

Ambos os classificadores avaliados apresentaram resultados satisfatórios, mas, em geral, o classificador com base em SVM teve melhor desempenho no reconhecimento de gestos e tempo de execução (etapas mais rápidas de teste e treinamento), sendo uma escolha eficiente e rápida para o uso padrão na versão final do sistema. Com o uso continuado do sistema, os profissionais de saúde poderiam capturar um número maior de amostras antes de realizar o treinamento do sistema, possivelmente obtendo resultados ainda melhores.

Em relação ao esforço necessário para os alunos atingirem os objetivos do jogo, foram percebidas diferenças significativas na interação de cada aluno. Os alunos A e C realizaram apenas dois gestos interativos e, apesar de não terem uma postura constante, são menos propensos a realizar gestos muito diferentes, o que possivelmente facilitou a condução rápida da tartaruga para o mar, usando poucos movimentos. O aluno B, por outro lado, por conseguir realizar vários gestos com características semelhantes, acabou se movendo bastante durante o uso do jogo (ainda que com dificuldades), gerando reconhecimento de gestos desnecessários, como o "Não" (realizado com a cabeça) e o gesto "Namorar" (realizado com as duas mãos), em vez de executar os gestos "Oi" e "Sim" (executados apenas com a mão direita). No entanto, pode-se esperar que, com o uso frequente, todos os alunos serão capazes de atingir os objetivos do jogo com menos esforço.

A usabilidade do sistema foi considerada adequada, de acordo com a avaliação dos profissionais da área de educação especial que acompanharam a execução dos experimentos.

A motivação para incluir a abordagem de jogo no sistema partiu da possibilidade de aumentar a imersão e o envolvimento dos alunos, e isso foi observado no Experimento 5. As tentativas de levar a tartaruga para o mar permitiram ao sistema capturar um grande número de amostras de gestos, rapidamente. No entanto, ao perceber que, tocando no caranguejo ou na concha, era possível continuar movendo a tartaruga sem prejuízos, os alunos não se preocuparam mais em evitar os obstáculos para alcançar o objetivo do jogo rapidamente. Talvez a inclusão de algum tipo de penalidade pudesse incentivar os alunos a traçar um caminho mais adequado para a tartaruga chegar ao mar, e utilizar assim uma variação maior de movimentos. Mesmo assim, foi observado o contentamento dos alunos em alcançar o objetivo do jogo, demonstrado por meio de sorrisos e comportamento eufórico. Foi possível observar também as possibilidades futuras que se abrem, permitindo uma interação mais natural e divertida para alunos que geralmente sofrem com tantas limitações.

Algumas situações observadas e lições aprendidas no primeiro experimento realizado com o público-alvo (Experimento 4) motivaram algumas escolhas de *design* e mudanças na execução do segundo experimento para gerar melhores resultados gerais. No Experimento 4, a webcam disponível no notebook em que o sistema estava sendo executado foi usada. Alguns usuários apresentaram movimentos involuntários durante a interação com o sistema, apoiando a perna ou o braço na mesa onde estavam dispostos o computador e a câmera, gerando algumas amostras com informações sobre o fundo da sala, consideradas ruidosas. Assim, no Experimento 5, optou-se pelo uso de uma *webcam* externa, apoiada em um tripé, mantendo uma posição próxima à *webcam* do notebook. Essa alteração se mostrou eficaz, pois não foram observadas amostras ruidosas neste experimento. Além disso, o número de amostras capturadas para cada gesto foi padronizado (quinze amostras originais).

Dois dos três alunos participantes do Experimento 5 mostraram uma redução em suas habilidades motoras desde o Experimento 4, realizado cerca de quatro meses antes. Infelizmente, conforme relatado pelos profissionais que acompanharam a execução das tarefas, essa situação é bastante comum e reforça nossa percepção de que os recursos de Tecnologia Assistiva precisam ser flexíveis e facilmente adaptáveis. Possivelmente, se houvesse uso contínuo do sistema por esses alunos desde o Experimento 4, o desempenho do sistema no reconhecimento de gestos diminuiria gradualmente, exigindo a criação de um novo conjunto de dados, já que a maneira como os gestos são realizados atualmente por alguns alunos já é bastante distinta.

A cada gesto realizado para interação com o sistema PGCA, é gerada uma imagem OF-MHI e também um arquivo de texto contendo a classe prevista pelo sistema e seu nível de confiança. Essas informações podem ser usadas em uma versão mais autônoma do sistema, no futuro, como uma maneira de recarregar o conjunto de dados com novas amostras, cuja classificação mostrou um alto nível de confiança. Isso pode ser importante no caminho da personalização, pois o sistema pode identificar, por exemplo, que ao longo do tempo um determinado gesto foi executado de maneira muito diferente da sua execução original quando o sistema foi treinado. Alcançando um nível de diferença muito significativo, o sistema poderia sugerir um novo treinamento usando amostras mais recentes.

5.7 AVALIAÇÃO DA METODOLOGIA EM OUTRO CONTEXTO

A metodologia MyPGI foi desenvolvida com o objetivo de guiar o desenvolvimento de soluções de CAA usando interação gestual personalizada, e os experimentos conduzidos pela autora desta tese, até o momento, permitiram avaliar a metodologia com o público-alvo em um ambiente escolar, utilizando o sistema PGCA.

A partir da demonstração de interesse de outro pesquisador (vinculado à área de pesquisa em CAA) em aplicar a metodologia MyPGI em um contexto diferente, a descrição da metodologia e o sistema PGCA, acompanhado de manual de instalação e utilização, foram disponibilizados para que o pesquisador pudesse projetar uma solução para atender seus objetivos. A solução em questão está em desenvolvimento, e como o pesquisador vivenciou e aplicou a metodologia, foi convidado a participar de uma atividade de avaliação, fornecendo um *feedback* de sua experiência até o momento, com o objetivo de avaliar a metodologia considerando outro contexto e a percepção de outra pessoa.

O participante aceitou participar de forma voluntária autorizando o uso dos dados para fins acadêmicos. O participante preencheu um questionário composto por perguntas referentes à aplicação, utilidade e facilidade da metodologia MyPGI, e intenção de uso futuro. O questionário utilizado foi inspirado no modelo de aceitação de tecnologia TAM 3 (*Technology Acceptance Model*) (Davis, 1989; Venkatesh e Bala, 2008) utilizando um formato não estruturado, e é

apresentado na Tabela 5.15. Buscou-se focar na utilidade percebida, (verificando o grau em que o usuário acredita que usando a tecnologia irá melhorar seu desempenho), facilidade de uso percebida (quão o usuário acredita que usando a tecnologia terá de realizar menos esforço para uma determinada atividade) e intenção de uso futuro (define o grau em que o usuário acredita que utilizaria a tecnologia em projetos futuros).

Tabela 5.15: Questionário usado para avaliar a metodologia MyPGI em um contexto diferente, por um pesquisador externo.

Q1	Em que contexto você vivenciou/aplicou a metodologia MyPGI? (Público-alvo, objetivo)
Q2	Com base na resposta da questão 1, considera que o uso da metodologia para guiar o projeto de uma solução voltada ao contexto citado foi útil? Por quê?
Q3	Como a metodologia apoiou o projeto da solução?
Q4	O que acha que não teria pensado ou conseguido fazer sem o uso da metodologia? O que ela fez perceber que não seria óbvio?
Q5	Quais as principais dificuldades encontradas ao seguir as etapas/direcionamentos descritos na metodologia? Por quê?
Q6	Alguma sugestão sobre como a metodologia poderia ser mais fácil de ser entendida e aplicada?
Q7	Você aplicaria a metodologia em outros projetos ou contextos? Pode citar algum exemplo?

5.7.1 Resultado da avaliação

O pesquisador participante dessa avaliação trabalha desde 2002 com paratletas (praticantes de atividade desportiva que possuem alguma deficiência, podendo ser esta física, visual, intelectual) de bocha paralímpica que possuem paralisia cerebral. Sua formação acadêmica compreende graduação em Educação Física, especialização em Medicina e Ciências do Esporte, e está cursando atualmente Mestrado em Engenharia Biomédica na UTFPR, onde realiza pesquisa vinculada à área de CAA. As respostas informadas pelo participante para cada questão do formulário de coleta de dados (Tabela 5.15) são descritas na sequência.

Quanto à questão "1) Em que contexto você vivenciou/aplicou a metodologia MyPGI? (Público-alvo, objetivo)", o participante respondeu que testou o sistema PGCA para ser utilizado com atletas de bocha paralímpica que possuem paralisia cerebral com o objetivo de facilitar a comunicação destes atletas durante as ações do jogo em treinamento.

Para a questão "2) Com base na resposta da questão 1, considera que o uso da metodologia para guiar o projeto de uma solução voltada ao contexto citado foi útil? Por quê?" A resposta dada foi "*Sim, porque o sistema mostrou-se bastante eficaz e fácil de ser implementado. As características do sistema facilitam a personalização da plataforma para o usuário de modo que possamos adequar os movimentos funcionais de cada usuário de forma independente*".

Sobre a questão "3) Como a metodologia apoiou o projeto da solução?" O participante informou que "*foi pela facilidade de traduzir um movimento específico em uma fala para ser utilizada no contexto da bocha paralímpica*".

Quanto à questão "4) O que acha que não teria pensado ou conseguido fazer sem o uso da metodologia? O que ela fez perceber que não seria óbvio?" O participante informou que "*estaria investindo em comunicação por sinais e pranchas de comunicação relacionados a bocha especificamente. A metodologia fez perceber que pode ser possível utilizar uma tecnologia de Visão Computacional para expressar informações curtas dos atletas*".

Sobre a questão "5) Quais as principais dificuldades encontradas ao seguir as etapas/direcionamentos descritos na metodologia? Por quê?" O participante indicou que "*as maiores dificuldades encontradas foram o treinamento e a avaliação do sistema, por ter demorado um*

pouco para compreender o funcionamento. Segundo ele, depois que a autora demonstrou a execução em etapas, tudo ficou mais claro".

Para a questão "6) Alguma sugestão sobre como a metodologia poderia ser mais fácil de ser entendida e aplicada?" O participante sugeriu algumas melhorias percebidas no uso do sistema PGCA: "*Como sugestão penso que a inclusão de caracteres especiais e a utilização de frases possam ser implementadas e com isso o sistema estará perfeito!!!*".

E por fim quanto à questão "7) Você aplicaria a metodologia em outros projetos ou contextos? Pode citar algum exemplo?" O participante respondeu que sim, em competições oficiais de bocha paralímpica.

Com base nas respostas informadas pelo participante da avaliação, por mais que a solução planejada ainda não tenha sido colocada em teste com o público-alvo em questão (atletas de bocha paralímpica), pode-se observar que a metodologia MyPGI tem potencial para ser empregada em diferentes contextos, guiando o desenvolvimento de soluções de TA personalizadas para pessoas com dificuldades motora e de fala.

5.8 CONSIDERAÇÕES

Esse capítulo apresentou a descrição de experimentos realizados com objetivo de analisar a viabilidade da proposta desta tese de doutorado.

Os testes realizados em condições de laboratório com profissionais que atuam com IHC (Experimento 1) permitiram identificar problemas de usabilidade e acessibilidade, melhorias e requisitos de sistema que precisavam ser refinados antes do sistema PGCA ser considerado adequado para experimentação com voluntários e com o público-alvo. Os resultados dessa avaliação permitiram verificar que a metodologia proposta é viável para guiar o desenvolvimento de sistemas de CAA usando interação gestual personalizada, e que, com as melhorias indicadas, o sistema PGCA teria potencial para ser utilizado em contextos reais com pessoas com deficiências, viabilizando a execução das tarefas previstas na metodologia.

O experimento com voluntários sem deficiência (Experimento 2) permitiu verificar a possibilidade de treinar um sistema com gestos personalizados, usando um número reduzido de amostras para treinamento, e permitiu a identificação de algumas melhorias necessárias ao sistema PGCA e à metodologia MyPGI.

O uso de um conjunto de dados público no Experimento 3 teve como objetivo permitir a repetibilidade do experimento por outros pesquisadores, além de possibilitar a realização de testes com diferentes tecnologias antes que o sistema fosse disponibilizado para testes com o público-alvo.

As informações obtidas por meio de entrevistas com profissionais da área de educação especial (especialistas de domínio) contribuíram para a identificação de alguns desafios e possibilidades de novas melhorias no sistema antes de iniciar os experimentos com o público-alvo. Além disso, as entrevistas permitiram identificar alunos com dificuldade motora e de fala que poderiam ser potenciais participantes dos experimentos seguintes.

Por fim, os Experimentos 4 e 5 trouxeram informações ricas, envolvendo representantes do público-alvo, mostrando um pouco da rotina dos alunos com deficiência motora e de fala no ambiente escolar e permitindo observar as iniciativas utilizadas pelos professores para se comunicar com esses alunos. A usabilidade do sistema foi avaliada por profissionais que acompanharam a condução dos experimentos e considerada adequada. De acordo com a percepção da autora desta tese, o emprego de uma abordagem baseada em jogos foi útil para estimular os alunos a participarem do Experimento 5 e poderia ser explorada no futuro também em outros momentos de interação com o sistema.

A avaliação da metodologia em outro contexto, ainda que tenha sido feita com base em uma solução em desenvolvimento, evidencia a capacidade de replicabilidade da metodologia MyPGI considerando outros problemas, outros usuários e outros profissionais. Os resultados obtidos por meio dos diferentes experimentos e avaliações conduzidas reforçam a percepção de que uma metodologia baseada em VC e Aprendizado de Máquina voltada para guiar o projeto de sistemas de CAA usando interação gestual personalizada é viável e pode ser aplicada em um contexto assistivo. Com isso, pode-se ampliar as possibilidades de pessoas com deficiência motora e de fala comunicarem-se por meio de sistemas de CAA.

O Capítulo 6 a seguir apresenta a conclusão desta tese de doutorado, por meio de uma discussão geral sobre os resultados obtidos a partir da pesquisa realizada, destacando trabalhos futuros e publicações geradas.

6 CONCLUSÃO

Este capítulo apresenta uma discussão sobre os experimentos e resultados obtidos, considerações finais sobre o projeto de pesquisa, trabalhos futuros e publicações geradas a partir desta tese de doutorado.

6.1 DISCUSSÃO

Neste documento foi descrita uma metodologia baseada em VC e Aprendizado de Máquina para guiar o projeto de sistemas de CAA usando interação gestual personalizada (MyPGI), e seu emprego no desenvolvimento de um sistema computacional de CAA. O sistema desenvolvido (PGCA) foi descrito e os métodos utilizados para sua implementação foram detalhados, assim como as tecnologias empregadas. O público-alvo desta tese são pessoas com dificuldades motora (independente da origem do comprometimento motor) e de fala, incapazes de interagir com sistemas computacionais por meio de dispositivos tradicionais.

Cinco experimentos foram conduzidos no decorrer da pesquisa realizada, para avaliar o sistema desenvolvido e conseqüentemente a metodologia proposta, contribuindo para a identificação de desafios e melhorias necessárias. Profissionais que atuam com IHC e voluntários sem dificuldades motora e de fala participaram de experimentos com o objetivo de antecipar problemas que poderiam surgir nos experimentos que seriam conduzidos posteriormente com representantes do público-alvo. Alunos com dificuldades motora e de fala fizeram uso do sistema PGCA em dois experimentos, permitindo observar o potencial e limitações do sistema desenvolvido em um contexto real. Profissionais da área de educação especial contribuíram para o amadurecimento da pesquisa por meio de entrevistas e avaliação de usabilidade do sistema.

A principal contribuição científica desta tese é o desenvolvimento da metodologia MyPGI que visa a promoção de autonomia para usuários de sistemas computacionais de CAA por meio do desenvolvimento de recursos de TA baseados em interação gestual personalizada. A efetividade dessa inovação foi avaliada por meio de experimentos conduzidos com diferentes objetivos, utilizando um sistema computacional de CAA desenvolvido como prova de conceito especialmente para esse propósito.

A contribuição tecnológica desta tese se refere à disponibilização do sistema PGCA como uma forma de aplicação da metodologia desenvolvida.

Os resultados obtidos permitiram responder as duas perguntas de pesquisa estabelecidas para esta tese. A primeira pergunta de pesquisa refere-se a verificar se é possível criar uma metodologia generalizável para o projeto de sistemas de CAA com base em interação gestual personalizada. A criação da metodologia se mostrou possível, uma vez que a metodologia MyPGI foi criada e descrita neste documento. Além disso, um sistema de CAA (PGCA) foi desenvolvido seguindo as etapas descritas na metodologia. Assim, foi possível observar que a metodologia criada é viável, e que um sistema desenvolvido a partir dela pode representar um recurso de TA promissor para apoiar a CAA de pessoas com dificuldades motora e de fala.

A segunda pergunta de pesquisa está relacionada a comprovar se, por intermédio de experimentos e utilizando-se da metodologia proposta, é possível auxiliar pessoas com comprometimento motor e de fala a se comunicarem por meio de TA baseada em interação gestual personalizada. Os experimentos conduzidos com alunos que possuem dificuldades motora e de fala permitiram observar na prática que a utilização do sistema PGCA pode auxiliar esses alunos a exercerem suas habilidades comunicativas, fazendo uso dos gestos que

costumeiramente já utilizam para se comunicar com seus cuidadores. Ao todo, quatro alunos com dificuldades motora e de fala criaram conjuntos de gestos personalizados, respeitando suas habilidades e limitações motoras. Os rótulos utilizados para as classes levaram em consideração o conhecimento dos cuidadores (profissionais que acompanharam a condução dos experimentos) no reconhecimento dos gestos executados pelos alunos. A execução de tarefas predeterminadas em diferentes sessões permitiu observar a viabilidade do uso do sistema PGCA e também do uso de interação gestual personalizada como um meio possível para permitir a comunicação e interação com sistemas computacionais, mesmo que algumas limitações no sistema desenvolvido tenham sido identificadas.

A interação gestual é uma prática comumente utilizada entre pessoas com dificuldades motora e de fala e seus cuidadores, como indicado por profissionais da área de educação especial em entrevistas. O observador familiar ou cuidador, normalmente compreende as intenções de comunicação da pessoa que cuida. Contudo, entre diferentes usuários, os gestos utilizados para comunicação podem variar consideravelmente, seja em função de limitações motoras, questões culturais, ou mesmo devido à maneira como os gestos foram ensinados. Assim, muitas vezes a linguagem gestual utilizada por uma pessoa com dificuldade motora e de fala, não segue modelos padronizados de linguagens de sinais, e sim um conjunto de gestos caseiros, personalizados. Nesse contexto, a metodologia MyPGI busca guiar o desenvolvimento de soluções de CAA voltadas para utilização por usuários e cuidadores, dando-lhes autonomia para a criação de uma linguagem gestual personalizada. Para isso é necessário que os desenvolvedores de *software* façam uma avaliação cuidadosa quanto a critérios de usabilidade e acessibilidade das soluções desenvolvidas. Afinal, os resultados provenientes dessas soluções serão bastante dependentes das escolhas feitas pelo usuário e cuidador, quanto aos gestos selecionados, ambiente utilizado para captura, e também, quanto à familiaridade de ambos (especialmente do cuidador) com o uso de sistemas computacionais. A descrição detalhada da metodologia em etapas visa compartilhar o conhecimento adquirido pela autora desta tese durante o desenvolvimento e avaliação da solução PGCA, com o intuito de direcionar a concepção de novas soluções de CAA baseadas em interação gestual personalizada.

Como observado nos trabalhos considerados correlatos, a CAA nem sempre é referenciada em pesquisas voltadas para viabilizar a comunicação de pessoas com deficiência. Independente disso, essa é uma área de conhecimento bastante abrangente e que vem sendo estimulada por pesquisas diversas, incluindo as baseadas em VC, como nos estudos de Chatteraj et al. (2017), Cristina e Camilleri (2016), Kurauchi et al. (2015) e Pal et al. (2017b). A metodologia MyPGI baseia-se em VC para explorar as vantagens inerentes ao uso contínuo de uma câmera monitorando as ações de usuários com dificuldade motora e de fala, no contexto de CAA. Os dados obtidos por meio de vídeo ou imagens capturadas são potencialmente úteis para melhor entender o comportamento dos usuários e suas intenções de comunicação. A crescente disponibilidade de poder computacional e os avanços da VC e Aprendizado de Máquina potencializam o desenvolvimento de sistemas inteligentes como um viés efetivo para auxílio na IHC.

Os jogos estão intimamente ligados ao entretenimento e ao prazer (Christou et al., 2014). Como já observado em outras pesquisas de TA (Foletto et al., 2018; Aced Lopez et al., 2015; Jiang et al., 2016, 2014), o uso de jogos tem demonstrado resultados positivos e isso foi observado também na abordagem baseada em jogo interativo empregada nesta tese. Os resultados obtidos com o último experimento conduzido com o público-alvo podem incentivar a reflexão positiva em relação às abordagens de intervenção baseadas em jogos.

Para a solução de CAA (sistema PGCA) criada e descrita nesta tese, buscou-se empregar tecnologias de baixo custo que permitissem a utilização de máquinas pessoais, sem demandar

investimentos em *hardware* adicionais e priorizando o uso de dispositivos não invasivos (sem contato com o usuário). O uso de imagens de histórico de movimento para representar os gestos executados, se mostrou uma alternativa adequada para os propósitos desta pesquisa. Apesar de algumas informações poderem ser perdidas ao calcular a projeção 2D de dados 3D, imagens de histórico de movimento são computacionalmente muito simples e eficientes (Tsai et al., 2015; Ahad et al., 2012). O uso de imagens também permitiu o emprego de um processo simplificado de Aumento de Dados, que foi essencial para acumular dados suficientes e dar condições para os classificadores atingirem um desempenho satisfatório. Aumento de dados é mais comumente utilizado no ramo de Aprendizado de Máquina que diz respeito à análise de imagens (Hauberg et al., 2016).

A análise de desempenho realizada a partir de diferentes conjuntos de gestos personalizados criados durante os experimentos, indicou resultados melhores no uso da representação de movimento OF-MHI, comparada ao uso de MHI convencional. Com base nisso, a inclusão do fluxo óptico para agregar informações de velocidade nas imagens de histórico de movimento se mostrou uma melhoria efetiva. Os dois classificadores implementados no sistema PGCA se mostraram capazes de aprender as informações dos conjuntos de dados (criados por voluntários e alunos) e puderam ser utilizados, portanto, para viabilizar a comunicação por meio de interação gestual personalizada. Contudo, o classificador com base em SVM mostrou desempenho ligeiramente melhor que o classificador com base em CNN, tanto em termos de precisão (na maioria dos resultados obtidos nos Experimentos 4 e 5), quanto em termos de tempo de execução (análise realizada no Experimento 5).

A usabilidade do sistema PGCA foi considerada aceitável (média SUS de 81), de acordo com avaliação realizada por profissionais da área de educação especial. O sistema demonstrou ser capaz de fornecer uma interação gestual natural, uma vez que reconheceu gestos em alguns casos em até dois segundos. Além de permitir a comunicação de pessoas com dificuldades motora e de fala usando linguagem gestual personalizada, inclusive no ambiente escolar, o sistema PGCA demonstrou potencial também para servir de apoio aos professores na tarefa de alfabetização.

Um dos principais objetivos da pesquisa em TA é desenvolver soluções tecnológicas que resolvam os problemas enfrentados pelas pessoas com deficiência (Kane et al., 2014). A participação de representantes do público-alvo no processo de desenvolvimento de recursos de TA é imprescindível para evitar o desenvolvimento de soluções desalinhadas com as necessidades de seus usuários. No entanto, é prudente analisar com cuidado o momento mais apropriado para incluir participantes com deficiências na etapa de avaliação das soluções. A realização de testes em condições de laboratório, ou com pessoas sem deficiência antes de iniciar os testes com o público-alvo é uma característica desejável para as pesquisas no contexto de TA. É importante ajustar todo o aparato técnico necessário e antever erros ou dificuldades que poderiam expor os usuários a situações desagradáveis ou desnecessárias. Assim, é possível desenvolver soluções de TA capazes de proporcionar uma experiência de uso mais adequada às exigências dos usuários, evitando frustrações. O adiamento de testes com o público-alvo pode não ser uma prática essencial, mas pode ser vista como uma questão de respeito com os usuários, principalmente porque as aspirações dos envolvidos pode gerar enorme expectativa.

O recrutamento de participantes da pesquisa pode ser difícil em todos os tipos de pesquisa sobre IHC e pode ser especialmente difícil ao realizar pesquisas com pessoas com deficiência. Localizar, recrutar e trabalhar com participantes com deficiência apresenta vários desafios (Kane et al., 2014), e isso também ocorre em pesquisas voltadas para CAA (Bhattacharya et al., 2008). O número reduzido de participantes envolvidos nos experimentos pode significar que a análise dos resultados é baseada em uma massa limitada de dados. Mas ainda assim, os dados obtidos a partir dos experimentos geraram informações encorajadoras, essenciais e

contribuíram efetivamente para o desenvolvimento da pesquisa vinculada a esta tese de doutorado. Ter um número reduzido de participantes reduz a capacidade de generalizar os resultados para uma população maior, mas de um modo geral, os resultados obtidos, mesmo que limitados em número, podem contribuir efetivamente para o conhecimento que vem sendo produzido na área de CAA.

Os estudos sobre adoção de tecnologia normalmente assumem que a percepção de usabilidade e utilidade da tecnologia por um usuário é central para sua adoção (Pal et al., 2017a). Uma questão a ser destacada nesse sentido, é que alguns dos alunos com dificuldades motora e de fala que participaram dos experimentos realizados, nunca haviam interagido com computadores. Então, além da dificuldade de se adaptar a uma nova solução, existia a dificuldade de visualizar ou entender as potencialidades associadas ao uso de sistemas computacionais (sua utilidade). Escrever algo em um editor de texto, ou navegador de Internet pode não ter um significado tão claro para quem nunca teve contato com esse tipo de tecnologia. Além disso, outra questão que pode demandar um período de adaptação no uso de soluções baseadas em interação gestual personalizada, refere-se à mudança de semântica dos gestos utilizados pelos usuários. Os gestos que antes eram executados apenas com o objetivo de transmitir intenções de comunicação, por meio de um sistema computacional como o PGCA, podem passar a ser utilizados para executar ações na interface (quando configurado para tal). Essas mudanças podem parecer estranhas para os usuários, mas tendem a ser minimizadas com o uso contínuo das soluções.

A metodologia proposta prevê algumas funcionalidades consideradas mínimas para sistemas de CAA desenvolvidos a partir dela, contudo, é possível que outros recursos e funcionalidades sejam necessários para que esses sistemas sejam ainda mais efetivos como apoio a comunicação. Espera-se que as soluções desenvolvidas a partir da metodologia MyPGI sejam adaptáveis, flexíveis e acessíveis, representando recursos de TA importantes para a promoção de autonomia a usuários de sistemas computacionais de CAA que possuem dificuldade motora e de fala.

6.2 TRABALHOS FUTUROS

A estratégia empregada para implementação de um sistema de CAA desenvolvido com base na metodologia MyPGI, deve levar em consideração as características do público-alvo, aprendendo suas ações e não apenas adaptando por limites as variantes de gestos de diferentes usuários. Como a metodologia está voltada para pessoas com dificuldades motora e de fala, devido às especificidades das deficiências de cada pessoa, o nível de compreensão ou consciência (habilidade cognitiva) e a habilidade motora dos usuários pode variar significativamente. Assim, um ponto que ainda pode ser melhor explorado a partir desta tese, é o nível de automatização do sistema, assim como o nível de interação exigida, que devem ser adequados às particularidades dos usuários.

Trabalhos futuros nessa área podem incluir investigação sobre uma abordagem personificada (representação simbólica), mais focada nas individualidades do usuário, podendo aprender e representar o usuário, indo além da adaptação e personalização da interface do sistema.

Conforme apresentado na Figura 6.1, partindo de um cenário ideal, no qual o usuário apresenta total habilidade cognitiva e habilidade motora para realização de gestos de maneira intencional, semelhante e repetidamente, um sistema pode aprender os gestos personalizados e realizar adaptações necessárias na interface. Contudo, considerando um cenário mais complexo, onde o usuário apresenta pouca habilidade cognitiva e pouca habilidade motora intencional, pode-se adotar uma abordagem personificada, concebendo recursos de Tecnologia Assistiva inteligentes que requerem intervenção mínima do usuário, capturando amostras continuamente

para interpretar padrões de movimentos realizados em frente à câmera. Essas amostras podem ser combinadas com outros dados, (por exemplo BCI) para treinar um sistema personificado, empregando também Aprendizado de Máquina, porém focando mais nas individualidades do usuário, podendo assim aprender e representar o usuário, superando os padrões de personalização.



Figura 6.1: Da personalização para a personificação: adequação da metodologia MyPGI levando em consideração a habilidade cognitiva e habilidade motora dos usuários, como uma possibilidade de trabalho futuro.

O sistema PGCA desenvolvido, demonstrou a partir dos experimentos conduzidos, estar apto para trabalhar com usuários que possuem pelo menos alguma habilidade cognitiva e alguma habilidade motora, atuando como uma abordagem de Tecnologia Assistiva mais próxima da personalização do que da personificação. Entretanto, a cada gesto executado para interação com o sistema PGCA, imagens MHI e OF-MHI são geradas, juntamente com arquivos de texto contendo a classe predita pelo sistema, e respectivo nível de confiança. Essas informações poderiam ser utilizadas em uma versão mais autônoma do sistema, como forma de realimentar o conjunto de dados com novas amostras, cuja classificação tenha apresentado um alto nível de confiança. Isso poderia ser importante no caminho da personificação, pois o sistema poderia identificar por exemplo que com o passar do tempo um determinado gesto vem sendo executado de maneira diferente do que foi executado originalmente quando o sistema foi treinado. E ao atingir um nível de diferença bastante significativo, o sistema poderia ser retreinado automaticamente utilizando amostras mais recentes.

Além disso, após um período longo de uso do sistema, um número bastante representativo de novas amostras é gerado. Isso pode viabilizar o treinamento completo de uma rede profunda como a CNN, permitindo analisar o desempenho obtido em relação ao uso de Transferência de Aprendizagem. Testes mais exaustivos podem ser conduzidos também com o objetivo de definir condições ideais para diferentes variações de fundo, iluminação escura, baixo contraste ou desfoque de movimento.

O diretor de uma das escolas visitadas comentou sobre o potencial do sistema para atuar como ferramenta de auxílio ao trabalho fisioterapêutico, pois, ao exigir a execução de um mesmo gesto repetidamente, estimula movimentos que podem ser benéficos para o usuário com deficiência motora. Por exemplo, no caso da aluna (participante dos Experimentos 4 e 5) que pode realizar movimentos com o braço direito, mas não estava habituada a usar esse gesto para fins de comunicação, o treinamento necessário poderia ser obtido em partes por meio de estímulos para interagir com o sistema, seja usando pranchas de comunicação ou jogando o jogo desenvolvido. Esse também poderia ser um trabalho de pesquisa futura relacionado ao

uso do sistema desenvolvido, analisando inclusive o esforço físico relacionado à execução de movimentos.

Observando o cotidiano dos alunos participantes dos experimentos, percebeu-se que para analisar a real efetividade de recursos de TA como solução para comunicação seria necessário treinamento e acompanhamento constantes, a longo prazo, conduzido possivelmente por profissionais da área de educação especial ou saúde. Converter o sistema PGCA para a plataforma *web* e disponibilizá-lo para acesso pode ser um ponto importante para viabilizar a disseminação do sistema e permitir a realização de novas avaliações por esses profissionais. Um manual de utilização e vídeo explicativo do sistema foram produzidos e podem ser úteis para familiarizar os futuros usuários e profissionais com as funcionalidades disponíveis.

O sistema desenvolvido em sua versão atual pode não produzir resultados satisfatórios para o reconhecimento de linguagens de sinais complexas, compostas por um grande número de gestos, com poucas variações entre eles. O sistema PGCA não foi desenvolvido com esse propósito, no entanto, a metodologia MyPGI poderia ser empregada para desenvolver outras abordagens, levando em consideração diferentes estratégias de treinamento e reconhecimento e as características dos gestos empregados na comunicação pelo público-alvo.

6.3 PUBLICAÇÕES GERADAS

Os resultados desta tese foram submetidos e divulgados, até o momento, no decorrer de seus experimentos, nas seguintes publicações ou apresentações:

- Rúbia Eliza de Oliveira Schultz Ascari, Roberto Pereira, e Luciano Silva. 2018. Mobile Interaction for Augmentative and Alternative Communication: a Systematic Mapping. *SBC Journal on 3D Interactive Systems* 9, 2 (2018), 105–118.
- Rúbia Eliza de Oliveira Schultz Ascari, Roberto Pereira, e Luciano Silva. 2018. Towards a Methodology to Support Augmentative and Alternative Communication by means of Personalized Gestural Interaction. In *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018)*. ACM, 38, 1-12. Belém/PA, Brasil. <https://doi.org/10.1145/3274192.3274230>.
- Rúbia Eliza de Oliveira Schultz Ascari, Luciano Silva, e Roberto Pereira. 2019. Personalized Interactive Gesture Recognition Assistive Technology. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems (IHC 2019)*. ACM, 38, 1-12. Vitória/ES, Brasil. <https://doi.org/10.1145/3357155.3358442>.
- Rúbia Eliza de Oliveira Schultz Ascari, Luciano Silva, e Roberto Pereira. 2019. Tecnologia Assistiva para auxiliar a comunicação de pessoas com dificuldades motora e de fala. Apresentação de pôster no Workshop Internacional sobre Tecnologias Assistivas. Fundação Araucária e SEJUF/PR. Curitiba/PR, Brasil.
- Rúbia Eliza de Oliveira Schultz Ascari, Luciano Silva, e Roberto Pereira. 2020. Personalized gestural interaction applied in a gesture interactive game-based approach for disabled people. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI 2020)*. ACM, 1-12. Cagliari, Itália. <https://doi.org/10.1145/3377325.3377495>.
- Rúbia Eliza de Oliveira Schultz Ascari, Roberto Pereira and Luciano Silva. 2020. Computer vision-based methodology to improve interaction for people with motor and speech impairment (Aceito para publicação). *ACM Transactions on Accessible Computing (ACM TACCESS)*. ACM, 1-33.

REFERÊNCIAS

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. et al. (2016). Tensorflow: a system for large-scale machine learning. Em *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, volume 16, páginas 265–283.
- Abascal, J. (2008). Users with disabilities: maximum control with minimum effort. *Articulated Motion and Deformable Objects*, páginas 449–456.
- Aced Lopez, S., Corno, F. e De Russis, L. (2015). Gnomon: Enabling dynamic one-switch games for children with severe motor disabilities. Em *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, páginas 995–1000. ACM.
- Adjouadi, M., Sesin, A., Ayala, M. e Cabrerizo, M. (2004). Remote eye gaze tracking system as a computer interface for persons with severe motor disability. Em *International Conference on Computers for Handicapped Persons*, páginas 761–769. Springer.
- Ahad, M. A. R., Tan, J. K., Kim, H. e Ishikawa, S. (2012). Motion history image: its variants and applications. *Machine Vision and Applications*, 23(2):255–281.
- Al Mahmud, A. (2012). *Iterative design to improve aphasic communication*. Tese de doutorado, Technische Universiteit Eindhoven.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.
- Alper, M. (2015). Augmentative, alternative, and assistive: Reimagining the history of mobile computing and disability. *IEEE Annals of the History of Computing*, 37(1):96–96.
- Alves, D. d. O., Gotti, M. d. O., Griboski, C. M. e Dutra, C. P. (2006). Sala de recursos multifuncionais: espaços para atendimento educacional especializado. *Brasília: Ministério da Educação, Secretaria de Educação Especial*.
- Alves, N., Blain, S., Falk, T., Leung, B., Memarian, N. e Chau, T. (2016). Access Technologies for Children and Youth with Severe Motor Disabilities. *Paediatric Rehabilitation Engineering: From Disability to Possibility*, página 45.
- Antunes, R. A., Palma, L. B., Coito, F. V., Duarteramos, H. e Gil, P. (2016). Intelligent human-computer interface for improving pointing device usability and performance. Em *12th IEEE International Conference on Control and Automation (ICCA)*, páginas 714–719. IEEE.
- Ascari, R. E. d. O. S., Pereira, R. e Silva, L. (2018a). Mobile Interaction for Augmentative and Alternative Communication: a Systematic Mapping. *SBC Journal on 3D Interactive Systems*, 9(2):105–118.
- Ascari, R. E. d. O. S., Pereira, R. e Silva, L. (2018b). Towards a Methodology to Support Augmentative and Alternative Communication by means of Personalized Gestural Interaction. Em *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems*, página 38. ACM.

- Ascari, R. E. d. O. S., Pereira, R. e Silva, L. (2020a). Computer vision-based methodology to improve interaction for people with motor and speech impairmen. *ACM Transactions on Accessible Computing (ACM TACCESS)*, 1(1):1–33. Aceito para publicação.
- Ascari, R. E. d. O. S., Pereira, R. e Silva, L. (2020b). Personalized gestural interaction applied in a gesture interactive game-based approach for disabled people). *Proceedings of the 25th International Conference on Intelligent User Interfaces*, páginas 1–11.
- Ascari, R. E. d. O. S., Silva, L. e Pereira, R. (2019). Personalized Interactive Gesture Recognition Assistive Technology. *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, páginas 1–12.
- Ashtiani, B. e MacKenzie, I. S. (2010). Blinkwrite2: an improved text entry method using eye blinks. Em *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, páginas 339–345. ACM.
- Azarian, A. e Siadat, A. (2011). A Synthesis of Software Evaluation Methodologies and the Proposal of a New Practical Approach. *Journal of Software (JSW)*, 6(11):2271–2281.
- Azmi, A., AlSabhan, N. M. e AlDosari, M. S. (2009). The Wiimote with SAPI: Creating an accessible low-cost, human computer interface for the physically disabled. *International Journal of Computer Science and Network Security*, 9(12):63–68.
- Bakheet, S. (2017). A Fuzzy Framework for Real-Time Gesture Spotting and Recognition. *Journal of Russian Laser Research*, 38(1):61–75.
- Baldner, F., Costa, P., Gomes, J. F. S. e Leta, F. R. (2017). *Metrologia por Imagem*. Elsevier Brasil.
- Ballard, D. H. e Brown, C. M. (1982). *Computer vision*. Prentice Hall.
- Baranauskas, M. C. C., de Souza, C. S. e Pereira, R. (2015). GrandIHC-BR–Grand Research Challenges for Human-Computer Interaction in Brazil. *Human-Computer Interaction Special Committee (CEIHC) of the Brazilian Computer Society (SBC)*.
- Berndt, D. J. e Clifford, J. (1994). Using dynamic time warping to find patterns in time series. Em *Workshop on Knowledge Discovery in Databases*, volume 10, páginas 359–370. Seattle, WA.
- Betke, M. (2008). Camera-Based Interfaces and Assistive Software for People with Severe Motion Impairments. Relatório técnico, Boston University Computer Science Department.
- Betke, M., Gips, J. e Fleming, P. (2002). The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10(1):1–10.
- Bez, M. R. (2014). *SCALA: Sistema de comunicação alternativa para processos de inclusão em autismo: uma proposta integrada de desenvolvimento em contextos para aplicações móveis e web*. Tese de doutorado, Universidade Federal do Rio Grande do Sul.
- Bhattacharya, S., Samanta, D. e Basu, A. (2008). Performance models for automatic evaluation of virtual scanning keyboards. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):510–519.

- Bhuiyan, M. e Picking, R. (2011). A gesture controlled user interface for inclusive design and evaluative study of its usability. *Journal of Software Engineering and Applications*, 4(09):513.
- Bian, Z.-P., Hou, J., Chau, L.-P. e Magnenat-Thalmann, N. (2016). Facial position and expression-based human-computer interface for persons with tetraplegia. *IEEE Journal of Biomedical and Health Informatics*, 20(3):915–924.
- Biswas, P. e Langdon, P. (2011). A new input system for disabled users involving eye gaze tracker and scanning interface. *Journal of Assistive Technologies*, 5(2):58–66.
- Biswas, P. e Langdon, P. (2013). A new interaction technique involving eye gaze tracker and scanning system. Em *Proceedings of the 2013 Conference on Eye Tracking South Africa*, páginas 67–70. ACM.
- Biswas, P. e Langdon, P. (2015). Multimodal intelligent eye-gaze tracking system. *International Journal of Human-Computer Interaction*, 31(4):277–294.
- Blank, M., Gorelick, L., Shechtman, E., Irani, M. e Basri, R. (2005). Actions as space-time shapes. Em *Tenth IEEE International Conference on Computer Vision*, volume 2, páginas 1395–1402. IEEE.
- Blignaut, P. (2017). Development of a gaze-controlled support system for a person in an advanced stage of multiple sclerosis: a case study. *Universal Access in the Information Society*, 16(4):1003–1016.
- Bobick, A. F. e Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267.
- Boni, V. e Quaresma, S. J. (2005). Aprendendo a entrevistar: como fazer entrevistas em Ciências Sociais. Em *Tese - Revista Eletrônica dos Pós-Graduandos em Sociologia Política da UFSC*, 2(1):68–80.
- Borges, L. C. L. d. F., Filgueiras, L., Maciel, C. e Pereira, V. (2013). A customized mobile application for a cerebral palsy user. Em *Proceedings of the 31st ACM International Conference on Design of Communication*, páginas 7–16. ACM.
- Bougharriou, S., Hamdaoui, F. e Mtibaa, A. (2017). Linear SVM classifier based HOG car detection. Em *18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, páginas 241–245. IEEE.
- Brasil (1999). Decreto 3.298 de 20 de dezembro de 1999. dispõe sobre a política nacional para a integração da pessoa portadora de deficiência, consolida as normas de proteção, e dá outras providências. *Diário Oficial da União*.
- Breiman, L. (2017). *Classification and regression trees*. Routledge.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Caminha, V. L. P. S. (2018). ADACA – Ambiente digital de aprendizagem para crianças autistas. <http://www.lncc.br/~alm/neupsico12/vera.pdf>. Acessado em 01/04/2020.
- Capovilla, A. G. S., Capovilla, F. C. et al. (2004). *Alfabetização: método fônico*. Memnon São Paulo.

- Capovilla, F. e Nunes, L. (2003). A memória de trabalho no paralisado cerebral: procedimento. *São Paulo, SP: Scortecchi, Fapesp, CNPq*, 1:151.
- Carroll, J. M. (2013). Human computer interaction-brief intro. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*
- Carvalho, A., Ponce de Leon, F. d. et al. (2006). Grandes desafios da pesquisa em computação no Brasil 2006–2016. *São Paulo: Sociedade Brasileira de Computação.*
- Cazzato, D., Leo, M. e Distante, C. (2014). An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation. *Sensors*, 14(5):8363–8379.
- Chattoraj, S., Vishwakarma, K. e Paul, T. (2017). Assistive system for physically disabled people using gesture recognition. Em *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, páginas 60–65. IEEE.
- Chau, M. e Betke, M. (2005). Real time eye tracking and blink detection with USB cameras. Relatório técnico, Boston University Computer Science Department.
- Chauhan, V. e Morris, T. (2001). Face and feature tracking for cursor control. Em *Proceedings of the Scandinavian Conference on Image Analysis*, páginas 356–362.
- Chen, C.-h. (2015). *Handbook of Pattern Recognition and Computer Vision*. World Scientific.
- Chen, W. (2013). Gesture-based applications for elderly people. Em *International Conference on Human-Computer Interaction*, páginas 186–195. Springer.
- Christou, G., Nardi, L. e Cheimonidou, A.-Z. (2014). Using video games for the rehabilitation of children with cerebral palsy: a pilot study. Em *International Conference on Human-Computer Interaction*, páginas 220–225. Springer.
- Coe, P. et al. (2015). Forcephone: new prototype with pressure and thermal feedback. Dissertação de Mestrado, University of Helsinki, Faculty of Science, Department of Computer Science.
- Cohen, L., Manion, L. e Morrison, K. (2000). Research methods in education [5 th edn] london: Routledge falmer. *Teaching in higher education*, 41:21.
- Commons, W. (2017). CommonsWikimedia. https://commons.wikimedia.org/wiki/File:Turtle_clip_art.svg. Acessado em 12/05/2020.
- Corno, F., Farinetti, L. e Signorile, I. (2002). A cost-effective solution for eye-gaze assistive technology. Em *IEEE International Conference on Multimedia and Expo*, volume 2, páginas 433–436. IEEE.
- Cortes, C. e Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- Cover, T. e Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.
- Cristina, S. e Camilleri, K. P. (2016). Model-based head pose-free gaze estimation for assistive communication. *Computer Vision and Image Understanding*, 149:157–170.

- da Silva, C. A., Fernandes, A. R. e Grohmann, A. P. (2014). STAR: Speech Therapy with Augmented Reality for Children with Autism Spectrum Disorders. Em *International Conference on Enterprise Information Systems*, páginas 379–396. Springer.
- Dalal, N. e Triggs, B. (2005). Histograms of oriented gradients for human detection. Em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, páginas 886–893. IEEE.
- Dall’Asta, E. e Roncella, R. (2014). A comparison of semiglobal and local dense matching algorithms for surface reconstruction. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 45.
- Das, S. (2016). Classification methods. Em *Data Science Using Oracle Data Miner and Oracle R Enterprise*, páginas 189–237. Springer.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, páginas 319–340.
- Davis, J. W. e Bobick, A. F. (1997). The representation and recognition of human movement using temporal templates. Em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, páginas 928–934. IEEE.
- Day, S. B.-S. (2015). OpenGameArt.org. <https://opengameart.org>. Acessado em 12/05/2020.
- De Faria Borges, L. C. L., Filgueiras, L. V. L. e Maciel, C. (2011). Towards a participatory development technique of assistive technology for mobility and speech impaired patients. Em *Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, páginas 247–256. Brazilian Computer Society.
- de Faria Borges, L. C. L., Filgueiras, L. V. L., Maciel, C. e Pereira, V. C. (2012). Customizing a communication device for a child with cerebral palsy using participatory design practices: contributions towards the PD4CAT method. Em *Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems*, páginas 57–66. Brazilian Computer Society.
- de Godoi, T. X. e Valentim, N. M. C. (2019). Towards an Integrated Evaluation of Usability, User Experience and Accessibility in Assistive Technologies. Em *Proceedings of the XVIII Brazilian Symposium on Software Quality*, páginas 234–239.
- Dehankar, A., Jain, S. e Thakare, V. (2017a). Using AEPI method for hand gesture recognition in varying background and blurred images. Em *Electronics, Communication and Aerospace Technology (ICECA), 2017 International Conference of*, volume 1, páginas 404–409. IEEE.
- Dehankar, A., Thakare, V. e Jain, S. (2017b). Detecting centroid for hand gesture recognition using morphological computations. Em *International Conference on Inventive Systems and Control (ICISC)*, páginas 1–5. IEEE.
- Deliberato, D. (2005). Speech and language therapy in the school: resources and procedures for augmentative and alternative communication. *Theoretical and methodological issues in research on augmentative and alternative communication. Canada: ISAAC*, páginas 116–125.

- Denman, P., Nachman, L. e Prasad, S. (2016). Designing for a user: Stephen Hawking's UI. Em *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops-Volume 2*, páginas 94–95. ACM.
- Diment, L. e Hobbs, D. (2014). A gesture-based virtual art program for children with severe motor impairments—development and pilot study. *Journal of Assistive, Rehabilitative & Therapeutic Technologies*, 2(1):23206.
- Dollár, P., Rabaud, V., Cottrell, G. e Belongie, S. (2005). Behavior recognition via sparse spatio-temporal features. Em *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, páginas 65–72. IEEE.
- Dong, C., Loy, C. C., He, K. e Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307.
- Drawdy, C. C. e Yanik, P. M. (2015). Gaze estimation technique for directing assistive robotics. *Procedia Manufacturing*, 3:837–844.
- Drigas, A. e Kokkalia, G. (2016). Mobile Learning for Special Preschool Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 10(1):60–67.
- Duda, R. O., Hart, P. E., Stork, D. G. et al. (1973). *Pattern classification*, volume 2. Wiley New York.
- Efron, B. (1992). Bootstrap methods: another look at the jackknife. Em *Breakthroughs in statistics*, páginas 569–593. Springer.
- Eidam, S., Garstka, J. e Peters, G. (2016). Towards regaining mobility through virtual presence for patients with locked-in syndrome. Em *Proceedings of the 8th International Conference on Advanced Cognitive Technologies and Applications. Rome, Italy*, páginas 120–123.
- El-Affifi, L., Karaki, M., Korban, J. e al Alaoui, M. A. (2004). 'Hands-free interface' - a fast and accurate tracking procedure for real time human computer interaction. Em *Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on*, páginas 517–520. IEEE.
- Epstein, S., Missimer, E. e Betke, M. (2014). Using kernels for a video-based mouse-replacement interface. *Personal and Ubiquitous Computing*, 18(1):47–60.
- Eroshkin, S. Y., Kameneva, N., Kovkov, D. e Sukhorukov, A. (2017). Conceptual system in the modern information management. *Procedia Computer Science*, 103:609–612.
- Escalera, S., Athitsos, V. e Guyon, I. (2017). Challenges in multi-modal gesture recognition. Em *Gesture Recognition*, páginas 1–60. Springer.
- Everts, I., Sebe, N., Jones, G. A. et al. (2007). Cooperative Object Tracking with Multiple PTZ Cameras. Em *International Conference on Image Analysis and Processing (ICIAP 2007)*, volume 7, páginas 323–330.
- Fan, X. e Tjahjadi, T. (2017). A dynamic framework based on local Zernike moment and motion history image for facial expression recognition. *Pattern Recognition*, 64:399–406.
- Federici, S. e Scherer, M. (2012). The assistive technology assessment model and basic definitions. *Assistive Technology Assessment Handbook*, páginas 1–10.

- Feil-Seifer, D. e Mataric, M. J. (2009). Human robotinteraction. *Encyclopedia of complexity and systems science*, 80.
- Fejtová, M., Figueiredo, L., Novák, P., Štěpánková, O. e Gomes, A. (2009). Hands-free interaction with a computer and other technologies. *Universal Access in the Information Society*, 8(4):277.
- Feng, W., Chen, M. e Betke, M. (2014). Target reverse crossing: a selection method for camera-based mouse-replacement systems. Em *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, páginas 1–4.
- Fernandes, L., Nunes, R. R., Matos, G., Azevedo, D., Pedrosa, D., Morgado, L., Paredes, H., Barbosa, L., Fonseca, B., Martins, P. et al. (2015). Bringing user experience empirical data to gesture-control and somatic interaction in virtual reality videogames: an exploratory study with a multimodal interaction prototype. Em *SciTecIn15 - Conferência Ciências e Tecnologias da Interação 2015*.
- Fine, G. e Tsotsos, J. K. (2009). Examining the feasibility of face gesture detection using a wheelchair mounted camera. Em *Proceedings of the 1st ACM SIGMM International Workshop on media studies and implementations that help improving access to disabled users*, páginas 19–28.
- Foletto, A. A., d’Ornellas, M. C. e Prado, A. L. C. (2018). Serious games for parkinson’s disease fine motor skills rehabilitation using natural interfaces. Em *MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics*, volume 245, página 74. IOS Press.
- Fu, X., McCane, B., Albert, M. e Mills, S. (2013). Action recognition based on principal geodesic analysis. Em *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, páginas 259–264. IEEE.
- Fu, Y. e Huang, T. S. (2007). hMouse: Head tracking driven virtual computer mouse. Em *Applications of Computer Vision, 2007. WACV’07. IEEE Workshop on*, páginas 30–30. IEEE.
- Gane, L., Power, S., Kushki, A. e Chau, T. (2011). Thermal imaging of the periorbital regions during the presentation of an auditory startle stimulus. *PLOS ONE*, 6(11):e27268.
- Gao, G. W. e Duan, X. Y. (2011). An overview of human-computer interaction based on the camera for disabled people. Em *Advanced Materials Research*, volume 219, páginas 1317–1320. Trans Tech Publ.
- Gao, X., Shi, L. e Wang, Q. (2017). The design of robotic wheelchair control system based on hand gesture control for the disabled. Em *2017 International Conference on Robotics and Automation Sciences (ICRAS)*, páginas 30–34. IEEE.
- Garbin, T. R. e Dainese, C. A. (2009). AmCARA - Ambiente e Comunicação Alternativo com Realidade Aumentada: O acesso do deficiente motor severo a softwares e Web. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Garbin, T. R. et al. (2008). *Ambientes de comunicação alternativos com base na realidade aumentada para crianças com paralisia cerebral: uma proposta de currículo em ação*. Tese de doutorado, Pontifícia Universidade Católica de São Paulo.

- García, L., Ron-Angevin, R., Loubière, B., Renault, L., Le Masson, G., Lespinet-Najib, V. e André, J. M. (2017). A comparison of a Brain-Computer Interface and an Eye tracker: is there a more appropriate technology for controlling a virtual keyboard in an ALS patient? Em *International Work-Conference on Artificial Neural Networks*, páginas 464–473. Springer.
- Gediga, G., Hamborg, K.-C. e Düntsch, I. (2002). Evaluation of software systems. *Encyclopedia of computer science and technology*, 45(supplement 30):127–53.
- Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350):320–328.
- Gevarter, C., O'Reilly, M. F., Rojeski, L., Sammarco, N., Lang, R., Lancioni, G. E. e Sigafos, J. (2013). Comparisons of intervention components within augmentative and alternative communication systems for individuals with developmental disabilities: A review of the literature. *Research in developmental disabilities*, 34(12):4404–4414.
- Ghanem, S., Conly, C. e Athitsos, V. (2017). A survey on sign language recognition using smartphones. Em *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, páginas 171–176.
- Gomez-Donoso, F., Cazorla, M., Garcia-Garcia, A. e Garcia-Rodriguez, J. (2016). Automatic Schaeffer's gestures recognition system. *Expert Systems*, 33(5):480–488.
- González, M., Mulet, D., Perez, E., Soria, C. e Mut, V. (2010). Vision based interface: an alternative tool for children with cerebral palsy. Em *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, páginas 5895–5898. IEEE.
- Gonzalez, R. C. e Woods, R. E. (2002). *Digital image processing*. Prentice hall.
- Google (2019). Google Image. <http://images.google.com>. Acessado em 12/11/2019.
- Graham-Knight, K. e Tzanetakis, G. (2015). Adaptive music technology using the kinect. Em *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, páginas 1–4.
- Grauman, K., Betke, M., Lombardi, J., Gips, J. e Bradski, G. R. (2003). Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2(4):359–373.
- Han, D., Liu, Q. e Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications*, 95:43–56.
- Hansen, J. P., Tørning, K., Johansen, A. S., Itoh, K. e Aoki, H. (2004). Gaze typing compared with input by head and hand. Em *Proceedings of the 2004 Symposium on Eye tracking Research & Applications*, páginas 131–138. ACM.
- Haugberg, S., Freifeld, O., Larsen, A. B. L., Fisher, J. e Hansen, L. (2016). Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. Em *Artificial Intelligence and Statistics*, páginas 342–350.
- Hemmingsson, H., Ahlsten, G., Wandin, H., Rytterström, P. e Borgestig, M. (2018). Eye-Gaze Control Technology as Early Intervention for a Non-Verbal Young Child with High Spinal Cord Injury: A Case Report. *Technologies*, 6(1):12.

- Henzen, A. F. e Nohama, P. (2017). Facial Movements Detection Using Neural Networks and Mpeg-7 Descriptors Applied to Alternative and Augmentative Communication Systems. Em *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*, páginas 626–629. Springer.
- Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. e Verplank, W. (1992). *ACM SIGCHI Curricula for Human-Computer Interaction*. ACM.
- Horn, B. K. e Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203.
- Hornof, A. J. e Cavender, A. (2005). EyeDraw: enabling children with severe motor impairments to draw with their eyes. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, páginas 161–170. ACM.
- Horprasert, T., Harwood, D. e Davis, L. S. (1999). A statistical approach for real-time robust background subtraction and shadow detection. Em *IEEE International Conference on Computer Vision (ICCV 99)*, volume 99, páginas 1–19. Citeseer.
- Huang, C.-P., Hsieh, C.-H., Lai, K.-T. e Huang, W.-Y. (2011). Human action recognition using histogram of oriented gradient of motion history image. Em *First International Conference on Instrumentation, Measurement, Computer, Communication and Control*, páginas 353–356. IEEE.
- Hunt, X. (2017). Machine learning concepts: styles of machine learning. <https://blogs.sas.com/content/subconsciousmusings/2017/09/25/machine-learning-concepts-styles-machine-learning/>. Acessado em 03/03/2020.
- Huo, F., Hendriks, E., Paclik, P. e Oomes, A. H. (2009). Markerless human motion capture and pose recognition. Em *Image Analysis for Multimedia Interactive Services, 2009. WIAMIS'09. 10th Workshop on*, páginas 13–16. IEEE.
- ISO/IEC (1998). 9241-11. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part II: Guidance on Usability. *The International Organization for Standardization*, 45:9.
- ISO/IEC (2001). 9126-1. Software engineering-product quality – Part 1: Quality model. *The International Organization for Standardization*, 1:21.
- Jacob, R. J. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2):152–169.
- Jain, L. C. e Kacprzyk, J. (2013). *New learning paradigms in soft computing*, volume 84. Physica.
- Jardini, R. S. R. (2003). *"Método das boquinhos": alfabetização e reabilitação dos distúrbios da leitura e escrita: livro I: fundamentação teórica*. Casa do Psicólogo.
- Jellinek, D. e Abrahams, P. (2012). Moving together: mobile apps for inclusion and assistance. *OneVoice for Accessible ICT*.
- Jha, V. (2017). Machine Learning Algorithm - Backbone of emerging technologies. <https://www.techleer.com/articles/203-machine-learning-algorithm-backbone-of-emerging-technologies/>. Acessado em 03/03/2020.

- Jiang, F., Ren, J., Lee, C., Shi, W., Liu, S. e Zhao, D. (2017). Spatial and temporal pyramid-based real-time gesture recognition. *Journal of Real-Time Image Processing*, 13(3):599–611.
- Jiang, H., Duerstock, B. S. e Wachs, J. P. (2014). An analytic approach to decipher usable gestures for quadriplegic users. Em *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, páginas 3912–3917. IEEE.
- Jiang, H., Duerstock, B. S. e Wachs, J. P. (2016). Variability analysis on gestures for people with quadriplegia. *IEEE Transactions on Cybernetics*, 48(1):346–356.
- Jiang, Z., Lin, Z. e Davis, L. (2012). Recognizing human actions by learning and matching shape-motion prototype trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):533–547.
- Ju, A. L. e Spasojevic, M. (2015). Smart Jewelry: The Future of Mobile User Interfaces. Em *Proceedings of the 2015 Workshop on Future Mobile User Interfaces*, páginas 13–15. ACM.
- Kakkoth, S. S. e Gharge, S. (2018). Real Time Hand Gesture Recognition & its Applications in Assistive Technologies for Disabled. Em *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, páginas 1–6. IEEE.
- Kane, S. K., Hurst, A., Buehler, E., Carrington, P. A. e Williams, M. A. (2014). Collaboratively designing assistive technology. *Interactions*, 21(2):78–81.
- Kane, S. K., Linam-Church, B., Althoff, K. e McCall, D. (2012). What we talk about: designing a context-aware communication tool for people with aphasia. Em *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, páginas 49–56. ACM.
- Kangas, K. e Lloyd, L. (1988). Early cognitive skills as prerequisites to augmentative and alternative communication use: What are we waiting for? *Augmentative and Alternative Communication*, 4(4):211–221.
- Karamchandani, H., Chau, T., Hobbs, D. e Mumford, L. (2015). Development of a low-cost, portable, tablet-based eye tracking system for children with impairments. Em *Proceedings of the International Convention on Rehabilitation Engineering & Assistive Technology*, páginas 1–4.
- Kbar, G., Bhatia, A., Abidi, M. H. e Alsharawy, I. (2017). Assistive technologies for hearing, and speaking impaired people: a survey. *Disability and Rehabilitation: Assistive Technology*, 12(1):3–20.
- Ke, S.-R., Thuc, H. L. U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H. e Choi, K.-H. (2013). A review on video-based human activity recognition. *Computers*, 2(2):88–131.
- Ke, Y., Sukthankar, R. e Hebert, M. (2007). Spatio-temporal shape and flow correlation for action recognition. Em *IEEE Conference on Computer Vision and Pattern Recognition*, páginas 1–8. IEEE.
- Khalifa, I., Ejbali, R. e Zaied, M. (2018). Hand motion modeling for psychology analysis in job interview using optical flow-history motion image: OF-HMI. Em *Tenth International Conference on Machine Vision (ICMV 2017)*. International Society for Optics and Photonics.

- Khan, A., Baharudin, B., Lee, L. H. e Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, 1(1):4–20.
- Kintsch, A. e DePaula, R. (2002). A framework for the adoption of assistive technology. *SWAAAC 2002: Supporting learning through assistive technology*, páginas 1–10.
- Kitchenham, B. (2004). Procedure for undertaking systematic reviews. *Computer Science Department, Keele University and National ICT Australia Ltd, Joint Technical Report*.
- Kocejko, T., Bujnowski, A. e Wtorek, J. (2009). Eye-mouse for disabled. Em *Human-Computer Systems Interaction*, páginas 109–122. Springer.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Em *Proceedings of Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, volume 14, páginas 1137–1145. Montreal, Canada.
- Krishnamurthi, R., Patan, R. e Gandomi, A. H. (2019). Assistive pointer device for limb impaired people: A novel Frontier Point Method for hand movement recognition. *Future Generation Computer Systems*, 98:650–659.
- Krueger, M. W., Gionfriddo, T. e Hinrichsen, K. (1985). VIDEOPLACE — an artificial reality. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 16, páginas 35–40. ACM.
- Kumar, S., Rai, A., Agarwal, A. e Bachani, N. (2010). Vision based human interaction system for disabled. Em *2010 2nd International Conference on Image Processing Theory, Tools and Applications*, páginas 441–446. IEEE.
- Kurauchi, A., Feng, W., Morimoto, C. e Betke, M. (2015). HMAGIC: head movement and gaze input cascaded pointing. Em *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, página 47. ACM.
- Kwan, C. e Betke, M. (2011). Camera Canvas: Image editing software for people with disabilities. Em *International Conference on Universal Access in Human-Computer Interaction*, páginas 146–154. Springer.
- Kwan, C., Paquette, I., Magee, J. J., Lee, P. Y. e Betke, M. (2011). Click control: improving mouse interaction for people with motor impairments. Em *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, páginas 231–232. ACM.
- Lafferty, J., McCallum, A. e Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Em *Proceedings of the Eighteenth International Conference on Machine Learning*, página 282–289.
- Lalanne, D., Nigay, L., Robinson, P., Vanderdonckt, J., Ladry, J.-F. et al. (2009). Fusion engines for multimodal input: a survey. Em *Proceedings of the 2009 International Conference on Multimodal Interfaces*, páginas 153–160. ACM.
- Lamb, K. e Madhe, S. (2016). Automatic bed position control based on hand gesture recognition for disabled patients. Em *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, páginas 148–153. IEEE.

- Landis, J. R. e Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, páginas 159–174.
- Laudan, L. (1978). *Progress and its problems: Towards a theory of scientific growth*, volume 282. Univ of California Press.
- LeCun, Y., Bengio, Y. e Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- Leite, D. Q., Duarte, J. C., Neves, L. P., de Oliveira, J. C. e Giraldo, G. A. (2017). Hand gesture recognition from depth and infrared Kinect data for CAVE applications interaction. *Multimedia Tools and Applications*, 76(20):20423–20455.
- Lemahieu, W. e Wyns, B. (2011). Low cost eye tracking for human-machine interfacing. *Journal of Eye Tracking, Visual Cognition and Emotion*, 1(1):1–12.
- Leo, M., Furnari, A., Medioni, G. G., Trivedi, M. e Farinella, G. M. (2018). Deep Learning for Assistive Computer Vision. Em *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Leo, M., Medioni, G., Trivedi, M., Kanade, T. e Farinella, G. M. (2017). Computer vision for assistive technologies. *Computer Vision and Image Understanding*, 154:1–15.
- Leung, B. e Chau, T. (2010). A multiple camera tongue switch for a child with severe spastic quadriplegic cerebral palsy. *Disability and Rehabilitation: Assistive Technology*, 5(1):58–68.
- Lewis, J. R. e Sauro, J. (2009). The factor structure of the system usability scale. Em *International Conference on Human Centered Design*, páginas 94–103. Springer.
- Liegel, L. A., Gogola, M. M. R. e Nohama, P. (2008). Layout de teclado para uma prancha de comunicação alternativa e ampliada. *Revista Brasileira de Educação Especial*, 14(3):479–496.
- Lin, C.-H., Hsu, F.-S. e Lin, W.-Y. (2010). Recognizing human actions using NWFE-based histogram vectors. *EURASIP Journal on Advances in Signal Processing*, 2010:9.
- Litslink (2019). An Introduction to Machine Learning Algorithms. <https://litslink.com/blog/an-introduction-to-machine-learning-algorithms>. Acessado em 03/03/2020.
- Littman, M. L. (2015). Reinforcement learning improves behaviour from evaluative feedback. *Nature*, 521(7553):445–451.
- Liu, Y., He, Y. e Cui, W. (2018). An improved SVM classifier based on multi-verse optimizer for fault diagnosis of autopilot. Em *IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, páginas 941–944. IEEE.
- Liu, Y., Lee, B.-S. e McKeown, M. J. (2016). Robust eye-based dwell-free typing. *International Journal of Human-Computer Interaction*, 32(9):682–694.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. Em *Proceedings of the seventh IEEE International Conference on Computer vision*, volume 2, páginas 1150–1157. IEEE.

- Lucas, B. D. e Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. Em *Proceedings of the 7th International Joint Conference on Artificial Intelligence*. Vancouver, BC, Canada.
- Lupu, R. G., Bozomitu, R. G., Păsărică, A. e Rotariu, C. (2017). Eye tracking user interface for internet access used in assistive technology. Em *E-Health and Bioengineering Conference (EHB), 2017*, páginas 659–662. IEEE.
- Ma, X., Tao, F., Zhang, M., Wang, T. e Zuo, Y. (2019). Digital twin enhanced human-machine interaction in product lifecycle. *Procedia CIRP*, 83:789–793.
- MacKenzie, I. S. e Ashtiani, B. (2009). Blinkwrite: efficient text entry using eye blinks. *Universal Access in the Information Society*, 10(1):69–80.
- Manresa-Yee, C., Ponsa, P., Varona, J. e Perales, F. J. (2010). User experience to improve the usability of a vision-based interface. *Interacting with Computers*, 22(6):594–605.
- Manresa-Yee, C., Varona, J., Perales, F. J. e Salinas, I. (2014). Design recommendations for camera-based head-controlled interfaces that replace the mouse for motion-impaired users. *Universal Access in the Information Society*, 13(4):471–482.
- Mao, X., Wang, L. e Li, C. (2008). SVM classifier for analog fault diagnosis using fractal features. Em *Second International Symposium on Intelligent Information Technology Application*, volume 2, páginas 553–557. IEEE.
- Marnik, J. (2014). BlinkMouse-On-Screen Mouse Controlled by Eye Blinks. Em *Information Technologies in Biomedicine*, volume 4, páginas 237–248. Springer.
- Martins, J. M., Rodrigues, J. M. e Martins, J. A. (2015a). Low-cost natural interface based on head movements. *Procedia Computer Science*, 67:312–321.
- Martins, P., Rodrigues, H., Rocha, T., Francisco, M. e Morgado, L. (2015b). Accessible options for deaf people in e-learning platforms: technology solutions for sign language translation. *Procedia Computer Science*, 67:263–272.
- Mauri, C., Granollers, T., Lorés, J. e García, M. (2006). Computer vision interaction for people with severe movement restrictions. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*.
- Mayer-Johnson, R. (2004). The picture communication symbols-PCS–Software Boardmaker. *Porto Alegre: Clik Tecnologia Assistiva*, 1.
- McCulloch, W. S. e Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.
- McMurrough, C., Rich, J., Metsis, V., Nguyen, A. e Makedon, F. (2012). Low-cost head position tracking for gaze point estimation. Em *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, páginas 1–4.
- Melo, A. M. e Baranauskas, M. C. C. (2006). An Inclusive Approach to Cooperative Evaluation of Web User Interfaces. Em *Proceedings of the 8th International Conference on Enterprise Information Systems (ICEIS)*, páginas 65–70.

- Memarian, N., Chau, T. e Venetsanopoulos, A. N. (2009a). Application of infrared thermal imaging in rehabilitation engineering: Preliminary results. Em *Science and Technology for Humanity (TIC-STH)*, 2009 IEEE Toronto International Conference, páginas 1–5. IEEE.
- Memarian, N., Venetsanopoulos, A. N. e Chau, T. (2009b). Infrared thermography as an access pathway for individuals with severe motor impairments. *Journal of Neuroengineering and Rehabilitation*, 6(1):11.
- Ministry, F. P. (2019). Turminha do MPF. <http://www.turminha.mpf.mp.br/>. Acessado em 01/04/2020.
- Missimer, E. e Betke, M. (2010). Blink and wink detection for mouse pointer control. Em *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments*, página 23. ACM.
- Moeslund, T. B. (2012). *Introduction to video and image processing: Building real systems and applications*. Springer Science & Business Media.
- Mohammad, U. e Anas, M. (2015). Design of a low cost DIY moving wheel chair using ATmega1284P based on retina movement for the persons disabled with quadriplegia. Em *2015 Annual IEEE India Conference (INDICON)*, páginas 1–4. IEEE.
- Mohammed, A. A. e Shereen, A. A. (2014). Efficient eye blink detection method for disabled-helping domain. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 5(5).
- Montanini, L., Cippitelli, E., Gambi, E. e Spinsante, S. (2015). Low complexity head tracking on portable android devices for real time message composition. *Journal on Multimodal User Interfaces*, 9(2):141–151.
- Moon, I., Kim, K., Ryu, J. e Mun, M. (2003). Face direction-based human-computer interface using image observation and EMG signal for the disabled. Em *Proceedings of the IEEE International Conference on Robotics Automation*, volume 1, páginas 1515–1520. IEEE.
- Moreira, E. A. e Baranauskas, M. C. C. (2018). Experiencing and delineating a vocabulary for a tangible environment to support alternative and augmentative communication. Em *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems*, páginas 1–10.
- Morrison, K. e McKenna, S. J. (2002a). Automatic visual recognition of gestures made by motor-impaired computer users. *Technology and Disability*, 14(4):197–203.
- Morrison, K. e McKenna, S. J. (2002b). Contact-free recognition of user-defined gestures as a means of computer access for the physically disabled. Em *Workshop on Universal Access and Assistive Technology*, páginas 99–103.
- Mountrakis, G., Im, J. e Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.
- Mouton, J. (1996). *Understanding social research*. Van Schaik Publishers.
- Munteanu, C., Oviatt, S., Penn, G. e Gomez, R. (2016). Designing Speech and Multimodal Interactions for Mobile, Wearable, and Pervasive Applications. páginas 3612–3619.

- Munzlinger, E., Narcizo, F. B. e de Queiroz, J. E. R. (2012). Sistematização de revisões bibliográficas em pesquisas da área de IHC. Em *Companion Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems*, páginas 51–54. Brazilian Computer Society.
- Murphy, K. P. e Russell, S. (2002). *Dynamic bayesian networks: representation, inference and learning*. Tese de doutorado, University of California, Berkeley.
- Myers, B. A. (1998). A brief history of human-computer interaction technology. *Interactions*, 5(2):44–54.
- Nakazawa, N., Aikawa, S. e Matsui, T. (2018). Development of Communication Aid Device for Disabled Persons Using Corneal Surface Reflection Image. Em *Proceedings of the 2nd International Conference on Graphics and Signal Processing*, páginas 16–20.
- Naqvi, R. A., Arsalan, M. e Park, K. R. (2017). Fuzzy system-based target selection for a NIR camera-based gaze tracker. *Sensors*, 17(4):862.
- Narasimhan, N. (2010). e-Accessibility Policy Handbook for Persons with Disabilities. *Nueva Delhi: ITU-G3ict*.
- Negin, F., Rodriguez, P., Koperski, M., Kerboua, A., González, J., Bourgeois, J., Chapoulie, E., Robert, P. e Bremond, F. (2018). PRAXIS: Towards Automatic Cognitive Assessment Using Gesture Recognition. *Expert Systems with Applications*.
- Nerişanu, R., Nerişanu, R.-A., Maniu, I. e Neamţu, B. (2017). Cerebral palsy and eye-gaze technology. Interaction, perspective and usability. A Review. *Acta Medica Transilvanica*, 22(4).
- Nguyen, H. D. e Poo, D. C. C. (2017). Unified Structured Framework for mHealth Analytics: Building an Open and Collaborative Community. Em *International Conference on Social Computing and Social Media*, páginas 440–450. Springer.
- Nielsen, J. (1989). Usability engineering at a discount. Em *Proceedings of the third International Conference on Human-Computer Interaction on Designing and using Human-Computer Interfaces and Knowledge based Systems (2nd ed.)*, páginas 394–401.
- Nielsen, J. (1994). Usability inspection methods. Em *Conference Companion on Human factors in Computing Systems*, páginas 413–414. ACM.
- Nister, D. e Stewenius, H. (2006). Scalable recognition with a vocabulary tree. Em *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, páginas 2161–2168. Ieee.
- Niu, S., Liu, L. e McCrickard, D. S. (2018). Tongue-able interfaces: Prototyping and evaluating camera based tongue gesture input system. *Smart Health*.
- Noaman, R. A., Ali, M. A. M. e Zainal, N. (2017). Enhancing pedestrian detection using optical flow for surveillance. *International Journal of Computational Vision and Robotics*, 7(1-2):35–48.
- Ossmann, R., Archambault, D. e Miesenberger, K. (2008). Accessibility issues in game-like interfaces. Em *International Conference on Computers for Handicapped Persons*, páginas 601–604. Springer.

- Oulasvirta, A. e Hornbæk, K. (2016). HCI research as problem-solving. Em *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, páginas 4956–4967. ACM.
- Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81.
- Pal, J., Viswanathan, A., Chandra, P., Nazareth, A., Kameswaran, V., Subramonyam, H., Johri, A., Ackerman, M. S. e O’Modhrain, S. (2017a). Agency in assistive technology adoption: visual impairment and smartphone use in bangalore. Em *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, páginas 5929–5940.
- Pal, S., Mangal, N. K. e Khosla, A. (2017b). Development of assistive application for patients with communication disability. Em *Proceedings of International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, páginas 1–4. IEEE.
- Paquette, I., Kwan, C. e Betke, M. (2011). Menu controller: making existing software more accessible for people with motor impairments. Em *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*, páginas 1–8.
- Parmar, K., Mehta, B. e Sawant, R. (2012). Facial-feature based Human-Computer Interface for disabled people. Em *International Conference on Communication, Information & Computing Technology (ICCICT)*, páginas 1–5. IEEE.
- Passerino, L. M. e Bez, M. R. (2013). Building an Alternative Communication System for literacy of children with autism (SCALA) with Context-Centered Design of Usage. Em *Recent Advances in Autism Spectrum Disorders-Volume I*. InTech.
- Patel, R. e Roy, D. (1998). Teachable interfaces for individuals with dysarthric speech and severe physical disabilities. Em *Proceedings of the AAAI Workshop on Integrating Artificial Intelligence and Assistive Technology*, páginas 40–47.
- Pauly, L. e Sankar, D. (2016). Non Intrusive Eye Blink Detection from Low Resolution Images Using HOG-SVM Classifier. *International Journal of Image, Graphics & Signal Processing*, 8(10).
- Pavliša, J. I., Ljubešić, M. e Jerečić, I. (2012). The use of AAC with young children in croatia—from the speech and language pathologist’s view. Em *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, páginas 221–230. Springer.
- Pei, L., Ye, M., Xu, P., Zhao, X. e Li, T. (2013). Multi-class action recognition based on inverted index of action states. Em *IEEE International Conference on Image Processing*, páginas 3562–3566. IEEE.
- Peleshko, D. e Soroka, K. (2013). Research of usage of Haar-like features and AdaBoost algorithm in Viola-Jones method of object detection. Em *12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, páginas 284–286. IEEE.
- Pereira, C. E. (2019). Livox - Software de comunicação alternativa. <http://www.livox.com.br>. Acessado em 01/04/2020.

- Perini, E., Soria, S., Prati, A. e Cucchiara, R. (2006). FaceMouse: A human-computer interface for tetraplegic people. Em *European Conference on Computer Vision*, páginas 99–108. Springer.
- Petersen, K., Vakkalanka, S. e Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.
- Pianca, S. (2016). Alfabeto - Método das boquinhas. <http://fonoaudiologiaemusica.blogspot.com/2016/07/alfabeto-metodo-das-boquinhas.html>. Acessado em 01/04/2020.
- Pirani, E. e Kolte, M. (2010). Gesture based educational software for children with acquired brain injuries. *International Journal in Computer Science and Engineering*, 2(3):790–794.
- PortalEscola (2014). Alfabeto de Parede - Atividades para imprimir. <https://www.portalescolar.net/2014/01/alfabeto-de-parede-atividades-para.html>. Acessado em 01/04/2020.
- Pradi, T. (2016). Ferramenta computacional para treinamento no reconhecimento e produção de expressões faciais por autistas. Dissertação de Mestrado, Pós-Graduação em Informática - Universidade Federal do Paraná.
- Pressman, R. e Maxim, B. (2016). *Engenharia de Software - 8ª Edição*. McGraw Hill Brasil.
- Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K. E. e Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3):171–193.
- Quintela, M. A., Mendes, M. e Correia, S. (2013). Augmentative and alternative communication: Vox4all® presentation. Em *8th Iberian Conference on Information Systems and Technologies (CISTI)*, páginas 1–6. IEEE.
- Rabiner, L.; Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3:4–16.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B. e Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21.
- Ravneberg, B. e Söderström, S. (2017). *Disability, society and assistive technology*. Taylor & Francis.
- Rebala, G., Ravi, A. e Churiwala, S. (2019). *An Introduction to Machine Learning*. Springer.
- Rivera, L. A. e DeSouza, G. N. (2014). Haptic and Gesture-Based Assistive Technologies for People with Motor Disabilities. Em *Assistive Technologies and Computer Access for Motor Disabilities*, páginas 1–27. IGI Global.
- Rodrigo, J. e Corral, D. (2013). ARASAAC: portal aragonés de la comunicación aumentativa y alternativa. Software, herramientas y materiales para la comunicación e inclusión. *Informática na Educação: Teoria & Prática*, 16(2).

- Rosales, C., Jácome, L., Carrión, J., Jaramillo, C. e Palma, M. (2017). Computer vision for detection of body expressions of children with cerebral palsy. Em *Ecuador Technical Chapters Meeting (ETCM), 2017 IEEE*, páginas 1–6. IEEE.
- Rosenberg, D. e Scott, K. (1999). *Use case driven object modeling with UML*. Springer.
- Roy, D. M., Panayi, M., Erenshteyn, R., Foulds, R. e Fawcus, R. (1994a). Gestural human-machine interaction for people with severe speech and motor impairment due to cerebral palsy. Em *Conference Companion on Human Factors in Computing Systems*, páginas 313–314. ACM.
- Roy, D. M., Panayi, M., Foulds, R., Erenshteyn, R., Harwin, W. S. e Fawcus, R. (1994b). The enhancement of interaction for people with severe speech and physical impairment through the computer recognition of gesture and manipulation. *Presence: Teleoperators & Virtual Environments*, 3(3):227–235.
- Rozado, D., Niu, J. e Lochner, M. (2017). Fast Human-Computer Interaction by Combining Gaze Pointing and Face Gestures. *ACM Transactions on Accessible Computing (TACCESS)*, 10(3):10.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Salcedo-Sanz, S., Rojo-Álvarez, J. L., Martínez-Ramón, M. e Camps-Valls, G. (2014). Support vector machines in engineering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(3):234–267.
- Saleh, S. e Berns, K. (2015). Nonverbal communication with a humanoid robot via head gestures. Em *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, páginas 1–8.
- Sambrekar, U. e Ramdasi, D. (2015). Human computer interaction for disabled using eye motion tracking. Em *International Conference on Information Processing (ICIP)*, páginas 745–750. IEEE.
- Sameshima, F. S., Rodrigues, I. B. e Deliberato, D. (2009). A parceria entre especialistas, professor e família no processo de implementação da comunicação alternativa: uma condição necessária. Em *Anais do V Congresso Brasileiro Multidisciplinar de Educação Especial*, páginas 379–388.
- Santana, V. F., Almeida, L. D. A. e Baranauskas, M. C. C. (2018). Websites atendendo a requisitos de acessibilidade e usabilidade. leanpub. <https://leanpub.com/warau>. Acessado em 26/06/2020.
- Sanz, C., Guisen, A., Baldassarri, S., Marco, J., Cerezo, E. et al. (2013). Games as educational strategy: A case of tangible interaction for users of Alternative and Augmentative Communication. Em *International Conference on Collaboration Technologies and Systems (CTS)*, páginas 377–381. IEEE.
- Sartoretto, M. L. e Bersch, R. (2014). Comunicação Alternativa. <http://www.assistiva.com.br/ca.html>. Acessado em 01/04/2020.

- Schrammel, J., Leitner, M. e Tscheligi, M. (2009). Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, páginas 2037–2040.
- Sedai, S., Bennamoun, M. e Huynh, D. (2009). Context-based appearance descriptor for 3d human pose estimation from monocular images. Em *Digital Image Computing: Techniques and Applications.*, páginas 484–491. IEEE.
- SEED-PR (2019a). Facilitando a comunicação entre surdos e ouvintes. www.icom-libras.com.br/. Acessado em 01/04/2020.
- SEED-PR (2019b). Sinalário Disciplinar em LIBRAS. <http://www.educadores.diaadia.pr.gov.br/modules/conteudo/conteudo.php?conteudo=1511>. Acessado em 01/04/2020.
- SEESP (2004). Portal de ajuda técnicas para educação: equipamento e material pedagógico para educação, capacitação da pessoa com deficiência física: recursos para comunicação alternativa. <http://portal.mec.gov.br/seesp/arquivos/pdf/comunicacao.pdf>. Acessado em 01/04/2020.
- Shah, D. e Philip, T. J. (2019). An Assistive Bot for Healthcare Using Deep Learning: Conversation-as-a-Service. Em *Progress in Advanced Computing and Intelligent Engineering*, páginas 109–118. Springer.
- Sharma, S., Varkey, B., Achary, K., Hakulinen, J., Turunen, M., Heimonen, T., Srivastava, S. e Rajput, N. (2018). Designing gesture-based applications for individuals with developmental disabilities: guidelines from user studies in india. *ACM Transactions on Accessible Computing (TACCESS)*, 11(1):1–27.
- Silva, C. M. d. (2016). Alfabetização e Deficiência Intelectual: uma estratégia diferenciada. *Semana Pedagógica*.
- Silva, C. M. d. (2019). Desafios do aprender. <https://abcclaudiamara.blogspot.com/?view=classic>. Acessado em 01/04/2020.
- Simpson, T., Broughton, C., Gauthier, M. J. e Prochazka, A. (2008). Tooth-click control of a hands-free computer interface. *IEEE Transactions on Biomedical Engineering*, 55(8):2050–2056.
- Smith, C. C. (2013). Using mobile technology to improve autonomy in students with intellectual disabilities in postsecondary education programs. Dissertação de Mestrado, The University of Tennessee, Knoxville.
- Sodre, B. S. (1986). *Cartilha Sodré*. São Paulo: Companhia Editora Nacional.
- Sousa, M. (2019). Jotalhao. <http://turmadamonica.uol.com.br/personagem/jotalhao/>. Acessado em 01/04/2020.
- Spain, V. U. e Spain, T. O. F. (2017). The Pictograms Room. www.pictogramas.org. Acessado em 01/04/2020.
- Stauffer, C. e Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. Em *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, páginas 246–252. IEEE.

- Stawicki, P., Gembler, F., Rezeika, A. e Volosyak, I. (2017). A novel hybrid mental spelling application based on eye tracking and SSVEP-based BCI. *Brain sciences*, 7(4):35.
- Steinfeld, E. e Maisel, J. (2012). *Universal design: Creating inclusive environments*. John Wiley & Sons.
- Story, M. F., Mueller, J. L. e Mace, R. L. (1998). The universal design file: Designing for people of all ages and abilities. *Raleigh, NC: North Carolina State University*.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. e Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Em *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 2818–2826.
- Tanner, M. A. e Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.
- Tavares, C. S. M., Anthero, F. e Scoz, M. (2018). Analysis of Methods for Evaluation of Assistive Technologies Focused on Computational Access of People with Cerebral Palsy. Em *Congress of the International Ergonomics Association*, páginas 1411–1419. Springer.
- Tomasi, C. e Kanade, T. (1991). Detection and tracking of point features. *Carnegie Mellon University Technical Report*.
- Torii, I., Ohtani, K., Shirahama, N., Niwa, T. e Ishii, N. (2012). Voice output communication aid application for personal digital assistant for autistic children. Em *11th International Conference on Computer and Information Science (ICIS) IEEE/ACIS*, páginas 329–333. IEEE.
- Torrey, L. e Shavlik, J. (2010). Transfer learning. Em *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, páginas 242–264. IGI Global.
- Toyama, K. (1998). Look, ma-no hands! hands-free cursor control with real-time 3d face tracking. *Workshop on Perceptual User Interfaces*.
- Triantafyllidis, A. K. e Tsanas, A. (2019). Applications of machine learning in real-life digital health interventions: Review of the literature. *Journal of Medical Internet Research*, 21(4).
- Tsai, D.-M., Chiu, W.-Y. e Lee, M.-H. (2015). Optical flow-motion history image (OF-MHI) for action recognition. *Signal, Image and Video Processing*, 9(8):1897–1906.
- Tu, J., Tao, H. e Huang, T. (2007). Face as mouse through visual face tracking. *Computer Vision and Image Understanding*, 108(1-2):35–40.
- Tuisku, O., Surakka, V., Rantanen, V., Vanhala, T. e Lekkala, J. (2013). Text entry by gazing and smiling. *Advances in Human-Computer Interaction*, 2013:1.
- Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189 – 195.
- Turk, Z. (2010). Communication Revolutions-How They Changed It All. Em *IKE*, página 491.
- Utaminigrum, F., Fauzi, M. A., Sari, Y. A., Primaswara, R. e Adinugroho, S. (2016). Eye Movement as Navigator for Disabled Person. Em *Proceedings of the International Conference on Communication and Information Systems*, páginas 1–5.

- Vafadar, M. e Behrad, A. (2008). Human hand gesture recognition using motion orientation histogram for interaction of handicapped persons with computer. Em *International Conference on Image and Signal Processing*, páginas 378–385. Springer.
- Varona, J., Manresa-Yee, C. e Perales, F. J. (2008). Hands-free vision-based interface for computer accessibility. *Journal of Network and Computer Applications*, 31(4):357–374.
- Veloso, É. e Maia, V. (2011). *Aprenda LIBRAS com eficiência e rapidez*. Curitiba: MãoSinais.
- Venkatesh, V. e Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315.
- Vidakis, N., Konstantinos, K. e Triantafyllidis, G. (2016). A Multimodal Interaction Framework for Blended Learning. Em *Interactivity, Game Creation, Design, Learning, and Innovation*, páginas 205–211. Springer.
- Vidhya, M. M., Devi, P. P., Emima, S. P. e Revathi, G. (2016). Implementation of Bidirectional Voice Communication between Normal and Deaf & Dumb Person. *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*.
- Viola, P. e Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Em *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1.
- Von Tetzchner, S. e Jensen, M. H. (1996). *Augmentative and alternative communication*. Whurr Publishers Ltda.
- Wan, J., Athitsos, V., Jangyodsuk, P., Escalante, H. J., Ruan, Q. e Guyon, I. (2014). CSMMI: Class-specific maximization of mutual information for action and gesture recognition. *IEEE Transactions on Image Processing*, 23(7):3152–3165.
- Wang, Y., Xu, C., Xu, C. e Tao, D. (2018). Packing convolutional neural networks in the frequency domain. Em *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Weiss, K., Khoshgoftaar, T. M. e Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1):9.
- Welch, G. e Bishop, G. (2006). An introduction to the Kalman filter, UNC-Chapel Hill. Relatório técnico, TR 95-041.
- Wilson, A. D. (2006). Robust computer vision-based detection of pinching for one and two-handed gesture input. Em *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, páginas 255–258. ACM.
- Xavier, K. F., da Costa, V. K., Cardoso, R. C., Peroba, J. A., Ferreira, A. O. L., Machado, M. B., Tavares, T. A. e Rodrigues, A. S. (2017). VisiUMouse: An Ubiquitous Computer Vision Technology for People with Motor Disabilities. Em *XIV Workshop de Trabalhos de Iniciação Científica (WTIC 2017). Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres*, páginas 115–118. Brazilian Computer Society.
- Xu, G., Wang, Y. e Zhang, X. (2010). Human computer interaction for the disabled with upper limbs amputation. Em *2nd International Conference on Advanced Computer Control*, volume 3, páginas 120–123.

- Yee, C. S. M., López, F. P. e Gómez, J. V. (2009). *Advanced and natural interaction system for motion-impaired users*. Tese de doutorado, Departament de Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears, Spain.
- Yoda, I., Ito, K. e Nakayama, T. (2017). Modular Gesture Interface for People with Severe Motor Dysfunction: Foot Recognition. *Studies in Health Technology and Informatics*, 242:725–732.
- Zander, T. O., Gaertner, M., Kothe, C. e Vilimek, R. (2010). Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *International Journal of Human-Computer Interaction*, 27(1):38–51.
- Zhang, J., Shao, K. e Luo, X. (2018). Small sample image recognition using improved convolutional neural network. *Journal of Visual Communication and Image Representation*, 55:640–647.
- Zhang, J., Zhuang, L., Wang, Y., Zhou, Y., Meng, Y. e Hua, G. (2013). An egocentric vision based assistive co-robot. Em *IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, páginas 1–7. IEEE.
- Zhang, S., Zhang, W. e Li, Y. (2016). Human Action Recognition Based on Multifeature Fusion. Em *Proceedings of Chinese Intelligent Systems Conference*, páginas 183–192. Springer.
- Zhang, X., Kulkarni, H. e Morris, M. R. (2017). Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities. Em *Proceedings of the CHI Conference on Human Factors in Computing Systems*, páginas 2878–2889. ACM.
- Zhao, Q., Tu, D., Xu, S., Shao, H. e Meng, Q. (2014). Natural human-robot interaction for elderly and disabled healthcare application. Em *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, páginas 39–44. IEEE.
- Zhao, T., Nevatia, R. e Lv, F. (2001). Segmentation and tracking of multiple humans in complex situations. Em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2. IEEE.