
Desenvolvimento de um teclado virtual para comunicação por meio de gestos visuais

Rafael Augusto da Silva



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Uberlândia
2021

Rafael Augusto da Silva

**Desenvolvimento de um teclado virtual para
comunicação por meio de gestos visuais**

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Processamento Digital de Sinais

Orientador: Antônio Cláudio Paschoarelli Veiga

Uberlândia

2021

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S586 Silva, Rafael Augusto da, 1990-
2021 Desenvolvimento de um teclado virtual para comunicação
por meio de gestos visuais [recurso eletrônico] / Rafael
Augusto da Silva. - 2021.

Orientador: Antônio Cláudio Paschoarelli Veiga.
Tese (Doutorado) - Universidade Federal de Uberlândia,
Pós-graduação em Engenharia Elétrica.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.te.2021.19>

Inclui bibliografia.

1. Engenharia elétrica. I. Veiga, Antônio Cláudio
Paschoarelli, 1963-, (Orient.). II. Universidade Federal
de Uberlândia. Pós-graduação em Engenharia Elétrica.
III. Título.

CDU: 621.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Elétrica				
Defesa de:	Tese de Doutorado, 280, PPGEELT				
Data:	Onze de janeiro de dois mil e vinte e um	Hora de início:	09:00	Hora de encerramento:	12:30
Matrícula do Discente:	11613EEL008				
Nome do Discente:	Rafael Augusto da Silva				
Título do Trabalho:	Desenvolvimento de um teclado virtual para comunicação por meio de gestos visuais				
Área de concentração:	Processamento da informação				
Linha de pesquisa:	Processamento digital de sinais				
Projeto de Pesquisa de vinculação:					

Reuniu-se por meio de videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Elétrica, assim composta: Professores Doutores: Gilberto Arantes Carrijo - FEELT/UFU; Milena Bueno Pereira Carneiro - FEELT/UFU; Lacordaire Kemel Pimenta Cury - IFG; Luiz Fernando Batista Loja - IFG; Antônio Cláudio Paschoarelli Veiga - FEELT/UFU, orientador(a) do(a) candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, Dr(a). Antônio Cláudio Paschoarelli Veiga, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a).

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Lacordaire Kemel Pimenta Cury, Usuário Externo**, em 12/01/2021, às 16:29, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gilberto Arantes Carrijo, Professor(a) do Magistério Superior**, em 12/01/2021, às 17:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Luiz Fernando Batista Loja, Usuário Externo**, em 13/01/2021, às 15:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Milena Bueno Pereira Carneiro, Professor(a) do Magistério Superior**, em 13/01/2021, às 23:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Antonio Claudio Paschoarelli Veiga, Professor(a) do Magistério Superior**, em 14/01/2021, às 11:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2502516** e o código CRC **3C1253E1**.

*Este trabalho é dedicado aos vários mestres da minha vida,
que despertaram em mim a alegria de trabalhar e conhecer.*

Agradecimentos

Ao Programa de Pós-Graduação em Engenharia Elétrica, por possibilitar a realização de excelentes trabalhos de pesquisa. Em especial à Cinara pela paciência e pronto auxílio.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pelo apoio financeiro que possibilitou a realização deste trabalho.

A meu orientador **Prof. Dr. Paschoarelli**, pelo apoio, paciência e confiança.

Ao professor **Eduardo Naves**, pela confiança e cessão de equipamento essencial para o desenvolvimento da pesquisa.

A meus pais, **Romes** e **Maria Angélica**, pelo amor incondicional, apoio e incentivo que sempre ofereceram a nossa família. Obrigado pelo exemplo!

A meus irmãos, **Romes** e **Renato**, com quem eu posso sempre contar.

A minha avó **Aranista**, pelo imenso carinho e presença constante.

A meus colegas e amigos, pelo apoio e conselhos que tornaram possível a realização deste trabalho.

A meus amigos de longa data que me dão coragem para continuar sempre.

A DEUS, por tornar possível todas as conquistas.

*“A ciência realmente só tem alcançado tornar
mais intensa e forte uma certeza:
- a velha certeza socrática da nossa irreparável ignorância.
De cada vez sabemos mais - que não sabemos nada.”
(Eça de Queirós)*

Resumo

A produção de texto é uma das atividades mais frequentes ao se utilizar um computador, uma tarefa corriqueira que pode ser afetada por limitações para indivíduos acometidos por distúrbios neuromotores severos como a esclerose lateral amiotrófica, por vezes agravada à síndrome do encarceramento. Tais indivíduos podem necessitar de ferramentas de comunicação aumentativa e alternativa, visto que em alguns casos dispõem unicamente dos movimentos dos olhos como forma de comunicação e interação com o mundo exterior.

Esta tese investiga métodos de interação baseados nos movimentos oculares, discorrendo sobre suas características, especificidades e limitações da interação humano-computador através da direção do olhar. Este trabalho propõe o desenvolvimento de um teclado virtual ocular que utiliza a detecção do traçado do olhar como entrada de texto, descrevendo o desenvolvimento da interface do teclado virtual, do método de rastreamento do olhar e do algoritmo de detecção de forma do traçado e votação de palavras potencialmente digitadas a partir de um léxico em português brasileiro.

A metodologia deste trabalho, exploratória e experimental, apresenta natureza quanti-qualitativa e engloba testes empíricos. Compreende extenso levantamento bibliográfico acerca de métodos de Comunicação Aumentativa e Alternativa, entrada de informação através da direção do olhar e desenvolvimento de teclados virtuais. A seção prática desta tese inclui o desenvolvimento de uma interface assistiva baseada na entrada de texto através de gestos do olhar e a avaliação de usabilidade a partir de experimentos com usuários reais sem treinamento.

Os resultados alcançados indicam precisão de reconhecimento de gestos acima de 90% utilizando teclados virtuais em três diferentes configurações de teclas, e velocidade de digitação variando de 3,56 a 7,41 palavras por minuto para usuários em condições de utilização sem treinamento e taxa média de reconhecimento correto de palavras de 83,75%.

Palavras-chave: Teclado ocular. Teclado virtual. Comunicação alternativa e aumentativa. Esclerose lateral amiotrófica. Interação humano-computador. Digitação ocular.

Abstract

Text production is one of the most frequent activities when using a computer, a common task that can be affected by limitations for individuals with severe neuromotor disorders such as amyotrophic lateral sclerosis, sometimes aggravated into the Locked-in syndrome. Such individuals may need tools for augmentative and alternative communication, since they may in some cases have only the eye movements as a form of communication and interaction with the outside world.

This thesis investigates methods of interaction based on eye movements, discussing their characteristics, possibilities and limitations of human-computer interaction through gaze direction. This work proposes the development of an eye-controlled virtual keyboard that uses gaze traces as text input, describing the development of the virtual keyboard interface, the gaze tracking method and the shape detection and decoding algorithm for words typed from a Brazilian Portuguese lexicon.

The methodology of this work, exploratory and experimental, has a quantitative and qualitative nature and involves empirical tests. It comprises an extensive bibliographic survey about Augmentative and Alternative Communication methods, information input through eye gaze direction and design of virtual keyboards. The practical section of this thesis includes the development of an assistive interface based on text input through gaze gestures and the usability evaluation based on experiments with real users without training.

The results achieved indicate gesture recognition accuracy above 90% with virtual keyboards in three different key layouts, and typing speed ranging from 3.56 to 7.41 words per minute for novice users and average recognition rate of 83.75%.

Keywords: Ocular keyboard. Virtual keyboard. Augmentative and alternative communication. Amyotrophic lateral sclerosis. Human computer interaction. Eye typing.

Lista de ilustrações

Figura 1 – (a) Teclado expandido e programável <i>IntelliKeys</i> , (b) diferentes modelos de mouse e (c) sistema <i>EyeMax</i> para controle do computador com movimento ocular, (d) linha Braille.	44
Figura 2 – Próteses de membros superiores e órtese de membro inferior.	45
Figura 3 – Desenho representativo da adequação postural, módulo postural em cadeira de rodas e crianças sentadas em posição adequada.	45
Figura 4 – Componentes de Comunicação Aumentativa e Alternativa (CAA) considerados por equipes da área durante a correspondência indivíduo-tecnologia.	50
Figura 5 – Exemplo de digitação da palavra <i>cor</i> através de gesto.	57
Figura 6 – Diagrama do sistema de CAA proposto.	57
Figura 7 – Trajeto dos movimentos oculares durante a leitura.	62
Figura 8 – Anatomia do olho humano.	64
Figura 9 – Células fotossensíveis: cones e bastonetes.	65
Figura 10 – (a) Distribuição de densidade de cones e bastonetes; (b) Diferença de acuidade entre visão central e periférica.	65
Figura 11 – Músculos do olho humano.	66
Figura 12 – Diagrama de um filtro linear simples de modelagem dos movimentos sacádicos.	68
Figura 13 – Movimentos oculares correspondentes a uma sacada.	70
Figura 14 – Movimentos oculares durante uma fixação em um ponto estacionário.	71
Figura 15 – Diagrama de um sistema linear com realimentação de modelagem dos movimentos de perseguição.	73
Figura 16 – Mapa de calor representando áreas de interesse na leitura de uma página web.	74
Figura 17 – Teclado virtual e dispositivo de rastreamento ocular.	82
Figura 18 – Interface do teclado ocular OptiKey.	84

Figura 19 – Localização de alvos fora da tela na interface proposta por Isokoski (2000).	86
Figura 20 – Gestos definidos para os caracteres na interface <i>EdgeWrite</i>	87
Figura 21 – Decodificação do traçado do olhar ou de um mouse em uma <i>string</i>	88
Figura 22 – Filteryedping - Interface de teclado virtual controlado por gestos oculares.	89
Figura 23 – Exemplo de bobina de rastreamento embutida em lente de contato e armações de campo eletromagnético para medição do movimento ocular.	93
Figura 24 – Lente de rastreamento ocular instalada no olho com o fino condutor saindo ao lado do nariz.	93
Figura 25 – (a) Unidade de gravação eletrofisiológica; (b) usuária de um rastreador ocular baseado em eletro-oculografia; (c) diversos tipos de eletrodo usados na eletro-oculografia (EOG).	94
Figura 26 – Exemplo de imagem registrada por vídeo-oculografia (VOG) exibindo o centro da pupila e as reflexões da fonte de luz IR adicional aplicada.	96
Figura 27 – Exemplo de vetor de posição existente entre o <i>glint</i> e centro da pupila e posições de nove pontos de calibração na VOG por reflexão corneana.	97
Figura 28 – Visualização das informações de rastreamento ocular em uma página web utilizando <i>heat map</i> e <i>gaze plot</i>	100
Figura 29 – Rastreador ocular Tobii 4c.	102
Figura 30 – Mapa mental das características de um teclado virtual.	106
Figura 31 – Representação de um teclado virtual em formato circular.	107
Figura 32 – Representação de teclados virtuais com posicionamento alternativo, descritos por (a) Sarcar et al. (2010) e (b) Ahsan et al. (2014).	107
Figura 33 – Exemplo de cálculo da velocidade de digitação.	111
Figura 34 – Exemplo de cálculo da quantidade de teclas por caractere.	112
Figura 35 – Layouts de teclado.	114
Figura 36 – Exemplo do processo de decodificação de traçado através da comparação com <i>template</i>	118
Figura 37 – Etapas de um sistema de reconhecimento de padrões típico.	119
Figura 38 – Paradigma do sistema de reconhecimento proposto por Liu et al. (2015).	120
Figura 39 – Exemplo de <i>Finite State Transition</i> (FST) do léxico para as palavras <i>go</i> e <i>forth</i>	122
Figura 40 – Exemplo do método <i>Dynamic Time Warping</i>	124
Figura 41 – Interface do teclado virtual desenvolvido, desenvolvido na forma de uma página HTML.	131
Figura 42 – Layout QWERTY utilizado no teclado assistivo.	132
Figura 43 – Layouts alternativos Fitaly e OPTI utilizados no teclado assistivo.	132
Figura 44 – <i>Feedback</i> visual apresentado ao se efetuar o gesto ocular para a palavra 'olho'.	134

Figura 45 – Exemplo do processo de decodificação de traçado (vermelho) através da comparação com <i>template</i> (azul).	137
Figura 46 – Indicação da matriz de coordenadas correspondente às letras do teclado QWERTY utilizado na interface assistiva.	139
Figura 47 – Indicação da matriz de coordenadas correspondentes às letras do teclado Fitaly utilizado na interface assistiva.	139
Figura 48 – Indicação da matriz de coordenadas correspondentes às letras do teclado OPTI utilizado na interface assistiva.	140
Figura 49 – Interface de configuração do software Gaze Point de controle do cursor através do olhar.	146
Figura 50 – Sistema utilizado nos testes de usabilidade, com rastreador ocular acoplado ao monitor.	150
Figura 51 – Taxa de reconhecimento de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.	153
Figura 52 – Taxa de reconhecimento de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.	153
Figura 53 – Quantidade de palavras não identificadas pelo algoritmo a partir de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.	154
Figura 54 – Entrada de gestos para a palavra 'inconstitucionalidade' no teclado assistivo, no layout (a)OPTI e (b)Fitaly.	155
Figura 55 – Taxa de reconhecimento acumulada na entrada de 100 gestos de palavras, de léxico contendo 3.000 palavras.	156
Figura 56 – Disposição de usuário na bancada de testes.	157
Figura 57 – Usuário durante realização do experimento.	158
Figura 58 – Taxa de reconhecimento de gestos oculares de 20 palavras aleatórias no layout QWERTY.	161
Figura 59 – Número de correções de palavra na digitação por gestos oculares de 20 palavras aleatórias no layout QWERTY.	161
Figura 60 – Análise comparativa da velocidade de digitação em diferentes configurações de tecla.	163
Figura 61 – Análise comparativa da velocidade de digitação em diferentes configurações de tecla.	164
Figura 62 – Estatística comparativa da velocidade de digitação para diferentes layouts de teclado.	165
Figura 63 – Avaliação de utilização da interface por participante do experimento (a) sem óculos e (b) utilizando óculos.	165

Lista de tabelas

Tabela 1 – Classificação Tecnologia Assistiva - ISO 9999:2002.	41
Tabela 2 – Definição dos termos usados na definição das métricas do rastreamento do olhar	76
Tabela 3 – Propriedades dos dispositivos de apontamento	80
Tabela 4 – Sistemas de controle ocular (baseados em VOG) disponíveis comercialmente.	102
Tabela 5 – Exemplo de amostras de coordenadas do olhar obtidas pelo rastreador Tobii 4c.	104
Tabela 6 – 40 palavras mais frequentes do corpus linguístico adotado, ordenadas por frequência e probabilidade de ocorrência.	136
Tabela 7 – Coordenadas das letras para os layouts de teclado adotados, utilizadas no cálculo dos <i>templates</i> das palavras do léxico.	141
Tabela 8 – Taxa de reconhecimento de gestos de palavras feitos com o mouse para tamanho crescente do léxico utilizado.	152
Tabela 9 – Taxa de reconhecimento de gestos de palavras feitos com o mouse para diferentes tamanhos de léxico e layouts de teclado.	152
Tabela 10 – Exemplo de palavras sugeridas na decodificação e tempo de processamento para cinco gestos de palavras efetuados.	154
Tabela 11 – Acurácia na decodificação de gestos oculares para 100 palavras, de léxico contendo 3.000 palavras.	155
Tabela 12 – Teste de digitação por gestos do olhar de 20 palavras selecionadas aleatoriamente, por quatro indivíduos.	160
Tabela 13 – Tempo de digitação (t) e quantidade de erros de palavra na digitação uma frase de 45 caracteres em diferentes teclados oculares.	162
Tabela 14 – Tempo médio de digitação de caractere e velocidade de digitação para uma frase, em diferentes layouts de teclado.	162

Lista de algoritmos

- 1 Interface do teclado virtual com captura e retorno das coordenadas de posição do cursor ao decodificador e desenho do traçado 131
- 2 Criação dos *templates* do léxico a partir das coordenadas centrais das letras no teclado virtual. 140
- 3 Criação dos *templates* do léxico a partir das coordenadas centrais das letras no teclado virtual e re-amostragem de traçado para conter 100 amostras. . 142
- 4 Cálculo do escore de similaridade de forma entre traçado do usuário e *templates* das palavras válidas. 144

Lista de siglas

ALS *Amyotrophic Lateral Sclerosis*

API *application programming interface*

AOI área de interesse

CAA Comunicação Aumentativa e Alternativa

CAT Comitê de Ajudas Técnicas

COGAIN *Communication by Gaze Interaction*

CPM caracteres por Minuto

DTW *Dynamic Time Warping*

ELA Esclerose Lateral Amiotrófica

EOG eletro-oculografia

FST *Finite State Transition*

GPC gestos por caractere

GUI interface gráfica de usuário

HEART *Horizontal European Activities in Rehabilitation Technology*

IHC interação humano-computador

IA inteligência artificial

ISO *International Organization for Standardization*

KSPC *keystrokes per character*

LIS *Locked-in syndrome*

LER lesão por esforço repetitivo

LSTM *Long Short Term Memory*

MSD *Minimum string distance*

PPM palavras por minuto

POR *Point of regard*

POG foto-oculografia

SE Síndrome do Encarceramento

SGD *Speech-generating Device*

TA Tecnologia Assistiva

WPM *words per minute*

VOG vídeo-oculografia

Sumário

1	INTRODUÇÃO	29
1.1	Motivação	30
1.2	Justificativa	31
1.3	Objetivos	31
1.4	Hipótese	32
1.5	Metodologia de Pesquisa	32
1.6	Contribuições	33
1.7	Estrutura deste Trabalho	34
1.8	Considerações Finais	35
2	TECNOLOGIA ASSISTIVA	37
2.1	Introdução	37
2.2	Tecnologia Assistiva	37
2.2.1	Classificação	41
2.3	Comunicação Aumentativa e Alternativa (CAA)	47
2.3.1	Componentes da CAA	49
2.3.2	Áreas de pesquisa	52
2.4	Distúrbios Neuromotores	52
2.4.1	Esclerose Lateral Amiotrófica	53
2.4.2	Síndrome do Encarceramento	54
2.5	Interfaces Oculares	55
2.6	Arquitetura do Sistema Proposto de Digitação por Gestos do Olhar	56
2.7	Considerações Finais	58
3	DIREÇÃO DO OLHAR COMO FONTE DE INFORMAÇÃO	59
3.1	Introdução	59
3.2	Aspectos do controle de interfaces através do olhar	59

3.3	Olho Humano e Visão	63
3.4	Movimentos Oculares	66
3.4.1	Sacada	68
3.4.2	Fixação	71
3.4.3	Perseguição	72
3.4.4	Caminho Percorrido (<i>scanpath</i>)	73
3.4.5	Área de Interesse (área de interesse (AOI))	73
3.4.6	Duração do Olhar	74
3.4.7	Identificação de Fixações e Sacadas	74
3.5	Métricas Relacionadas ao Rastreamento Ocular	75
3.6	Aplicações de Diagnóstico	77
3.7	Mouses Oculares	79
3.8	Teclados Oculares	82
3.8.1	OptiKey	83
3.9	Digitação ocular sem fixação (<i>dwell-free eye typing</i>)	84
3.9.1	Bibliografia Relacionada	85
3.10	Considerações Finais	89
4	RASTREAMENTO DO OLHAR	91
4.1	Introdução	91
4.2	Rastreamento do Olhar	91
4.2.1	Técnicas de Rastreamento do Olhar	92
4.2.2	Critérios de Qualidade dos Rastreadores do Olhar	98
4.2.3	Visualização dos Dados de Rastreamento Ocular	100
4.3	Sistemas de Rastreamento Ocular Disponíveis Comercialmente	101
4.4	Tobii 4C	101
4.5	Considerações Finais	103
5	TECLADOS VIRTUAIS	105
5.1	Introdução	105
5.2	Características de um Teclado Virtual	105
5.2.1	Posicionamento das teclas	106
5.2.2	Tamanho das teclas	107
5.2.3	Distribuição das letras	108
5.2.4	Quantidade de caracteres por tecla	108
5.2.5	<i>Feedback</i> do teclado	108
5.2.6	Navegação	109
5.2.7	Caracteres especiais	110
5.3	Métricas de Avaliação de Teclados Virtuais	110
5.3.1	Velocidade de digitação	110

5.3.2	Teclas por caractere	112
5.3.3	Taxa de erros	112
5.3.4	Avaliação subjetiva	113
5.4	Layout de Teclados	113
5.4.1	QWERTY	113
5.4.2	Dvorak	115
5.4.3	Fitaly	115
5.4.4	OPTI	115
5.4.5	Cirrin	116
5.4.6	Lewis	116
5.4.7	Hooke	116
5.5	Teclados Gestuais	117
5.6	Decodificação de Gestos do Olhar em Texto	117
5.6.1	Comparação entre <i>Strings</i>	119
5.6.2	Aprendizado de Máquina	121
5.6.3	Redes Neurais Recorrentes do tipo LSTM	121
5.6.4	<i>Dynamic Time Warping</i>	122
5.6.5	<i>Proportional Shape Matching</i>	126
5.7	Considerações Finais	126
6	TECLADO VIRTUAL PROPOSTO	129
6.1	Introdução	129
6.2	Interface	130
6.2.1	Disposição das teclas	131
6.2.2	Tamanho das teclas	133
6.2.3	Método de interação	133
6.2.4	<i>Feedback</i> do teclado	133
6.3	Dicionário em português do Brasil	134
6.4	Decodificação de gestos do olhar	135
6.4.1	Criação dos <i>templates</i> a partir do dicionário	138
6.4.2	Re-amostragem do traçado efetuado pelo usuário	140
6.4.3	Cálculo da probabilidade de palavra	143
6.5	Integração com rastreador ocular	145
6.6	Considerações Finais	146
7	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	149
7.1	Introdução	149
7.1.1	Equipamentos	150
7.2	Experimento 1: Avaliação da taxa de reconhecimento de tra-	
	çado de palavras	151

7.2.1	Método	151
7.2.2	Análise dos Resultados	151
7.3	Experimento 2: Avaliação comparativa da digitação ocular com usuários sem experiência prévia	155
7.3.1	Método	155
7.3.2	Participantes	156
7.3.3	Procedimento	157
7.3.4	Análise dos Resultados	159
7.4	Considerações Finais	163
8	CONSIDERAÇÕES FINAIS	167
8.1	Introdução	167
8.2	Contribuições	167
8.3	Trabalhos Futuros	168
8.4	Contribuições em Produção Bibliográfica	169
REFERÊNCIAS	171

Introdução

A entrada de texto é uma das atividades mais frequentes dos usuários ao utilizar um computador ou dispositivo móvel, ao redigir emails, preencher formulários, navegar na web e realizar atividades de comunicação. Pessoas com restrições motoras severas podem experimentar sérias dificuldades na realização da tarefa rotineira de digitar um texto, e podem se utilizar de ferramentas de comunicação assistiva como por exemplo teclados oculares. A tarefa de digitação utilizando os olhos requer atenção, foco visual constante e pode ser exaustiva. Pesquisas recentes se beneficiam de todas as possibilidades que os dispositivos modernos oferecem e procuram diminuir a fadiga visual, aumentar a velocidade de digitação e encontrar formas de entrada de informação mais eficientes através do movimento dos olhos ou direção do olhar (LOJA, 2015; LIU et al., 2016; ZHAI; KRISTENSSON, 2007a; KRISTENSSON, 2007; KURAUCHI et al., 2016).

A área de estudo denominada interação humano-computador (IHC) busca tornar as atividades no computador (incluindo a entrada de texto) mais fáceis, eficientes, intuitivas e inclusivas. O campo de estudo de IHC (em inglês *human-computer interaction*) é bastante amplo, englobando investigações que vão desde etnografia e sua relação com a usabilidade até a modelagem do desempenho humano e desenvolvimento de dispositivos de entrada (DREWES, 2010). O objeto deste trabalho pertence ao ramo de estudo de dispositivos de entrada e entrada de texto, associado à área da Tecnologia Assistiva e sua subárea denominada Comunicação Aumentativa e Alternativa (FILHO, 2009).

A interação com o computador não mais se limita à utilização de teclado, mouse, monitor e impressora. Telas de alta resolução, microfones, alto falantes, telas sensíveis ao toque e diferentes dispositivos de entrada são comuns nos dias de hoje. Novas modalidades de interação como entrada por gestos ou objetos tangíveis (canetas em telas sensíveis ao toque por exemplo), comandos de voz e diversos tipos de sensores usuais. A entrada de informação através rastreamento ocular é uma tecnologia explorada há décadas, mas que tem recentemente encontrado grande aplicação em sistemas de acessibilidade.

1.1 Motivação

A maior parte das interações de um indivíduo com o computador é realizada utilizando-se teclado e mouse, dispositivos que requerem utilização das mãos (HILL, 2010). Algumas pessoas sofrem com tensão excessiva em partes das mãos podendo ocasionar sintomas como a lesão por esforço repetitivo (LER). Existem indivíduos acometidos por distúrbios neuromotores que em estágios avançados tornam a comunicação extremamente restrita e impossibilitam atividades motoras. É o caso de pessoas afetadas pela Esclerose Lateral Amiotrófica (ELA) ou sua condição agravada, a Síndrome do Encarceramento (SE).

O número de interações dos usuários com dispositivos eletrônicos de comunicação apresenta tendência crescente provocada pelo contínuo desenvolvimento tecnológico, e busca-se novas técnicas de interação que minimizem possíveis problemas físicos e que possibilitem a comunicação de indivíduos com mobilidade restrita.

Uma alternativa conveniente pode ser a utilização dos olhos como método de entrada de informação, uma vez que eles irão inevitavelmente se movimentar durante a interação com um computador. Aproveitar a informação implícita no movimento dos olhos pode economizar atividades de interação, especialmente as baseadas nos movimentos das mãos. Desenvolver técnicas e interfaces que se aproveitam da facilidade e velocidade dos movimentos oculares pode trazer benefícios aos usuários.

Inúmeras interfaces assistivas existem até a data de publicação desta tese, destinadas a indivíduos com restrições de comunicação. Um dos elementos fundamentais das interfaces é o teclado virtual assistivo. Diversas opções de teclados virtuais podem ser encontradas, entretanto os teclados assistivos ainda não apresentam eficiência de entrada de dados elevada e geralmente possuem alta complexidade de utilização se comparados a métodos tradicionais de digitação (KRISTENSSON; VERTANEN, 2012).

Os teclados virtuais utilizados em dispositivos com telas sensíveis ao toque obtiveram inúmeros aperfeiçoamentos referentes ao conforto de utilização, velocidade de digitação, taxa de acerto de palavra, predição de palavras entre outros atributos. O contínuo desenvolvimento de redes neurais e elementos de inteligência artificial (IA) para o processamento da linguagem natural (tradução de texto, reconhecimento de voz, análise sintática, entre outros) possibilitam o reconhecimento de gestos efetuados sobre teclados virtuais e sua tradução em texto, elemento este referido extensamente neste trabalho. Destaca-se as produções bibliográficas relacionadas de Kurauchi (2018), Loja (2015), Pedrosa (2015), Alsharif et al. (2015) e Kristensson e Zhai (2004).

Busca-se neste trabalho aperfeiçoar o teclado virtual controlado pelos olhos diminuindo-se a quantidade de interações necessárias na operação de entrada de texto e aumentando-se a velocidade de digitação, mensurada em palavras por minuto (PPM). São encontradas na literatura diversas implementações de teclados que utilizam gestos feitos pelos dedos do usuário como entrada de texto (KRISTENSSON, 2007; ZHAI; KRISTENSSON, 2003; ISOKOSKI, 2000). Há diversas denominações para esses teclados, incluindo teclados ges-

tuais, teclados *swipe*¹, teclados de escrita através de formas, ou ainda teclados por gestos de palavras.

Desta forma, a principal motivação desta tese consiste no desenvolvimento de um teclado virtual assistivo híbrido que combine teclados oculares com teclados virtuais por gestos e utilize como entrada o traçado do olhar, proporcionando a indivíduos com capacidades motoras reduzidas uma opção mais confortável, eficiente e que ofereça menor carga cognitiva (cansaço mental) durante a operação de entrada de texto.

1.2 Justificativa

Investigações referentes à Comunicação Aumentativa e Alternativa (CAA) ainda são exíguas na comunidade acadêmica brasileira. Esta tese propõe uma abordagem distinta de otimização de um teclado virtual controlado por movimentos oculares.

A tarefa de rastreamento ocular para entrada de texto possui diversas particularidades e apresenta grande complexidade; as tentativas de aperfeiçoamento desse método esbarram em soluções não triviais (MAJARANTA; RÄIHÄ, 2002). A utilização de IA aplicada à linguagem natural pode ser uma direção conveniente para se alcançar maiores velocidades de digitação, métodos de predição de palavras mais eficientes e técnicas robustas de correção de erros, visando a diminuição do esforço do usuário ao digitar utilizando os olhos e melhor qualidade de vida a pessoas que dependem da Tecnologia Assistiva (TA).

1.3 Objetivos

O objetivo do presente trabalho consiste no desenvolvimento de uma interface de comunicação assistiva capaz de decodificar gestos do olhar sobre um teclado virtual na tela de um computador em entrada de texto.

O desenvolvimento e otimização de cada módulo da interface - da aquisição do traçado ocular à sugestão de palavras prováveis - deve resultar no bom desempenho da aplicação, destinada à utilização em dispositivos microcomputadores genéricos e de capacidade moderada de processamento. O software desenvolvido será integralmente composto de código aberto e será disponibilizado gratuitamente à comunidade científica, possibilitando seu aperfeiçoamento e adaptação a necessidades específicas.

A interface desenvolvida pode servir no auxílio à comunicação de pessoas com distúrbios neuromotores severos provendo melhor desempenho de digitação - maior taxa de palavras por minuto e menor taxa de erros - sendo uma ferramenta de inclusão e aplicação da tecnologia na promoção da saúde e bem-estar.

Objetivos específicos desta tese:

¹ O gesto de deslizar os dedos em interfaces sensíveis ao toque é comumente denominado *swipe*, em inglês.

- ❑ Discorrer acerca da área de estudo denominada Tecnologia Assistiva (TA), sobretudo voltada à Comunicação Aumentativa e Alternativa (CAA);
- ❑ explicar sobre a anatomia e os movimentos oculares;
- ❑ apresentar o estado da arte dos teclados virtuais;
- ❑ desenvolver uma interface de teclado virtual controlado pelos olhos, multiplataforma e de código aberto;
- ❑ formular um algoritmo de decodificação de gestos do olhar no teclado virtual em palavras;
- ❑ integrar o software ao rastreador ocular, resultando em um sistema funcional de comunicação assistiva;
- ❑ realizar testes de usabilidade da solução desenvolvida.

1.4 Hipótese

Os teclados assistivos são ferramentas importantes a indivíduos acometidos por distúrbios neuromotores, no entanto apresentam maior esforço de digitação e menor desempenho de entrada de texto em comparação aos meios tradicionais de escrita. São encontradas na literatura relacionada diversas abordagens que buscam minimizar as desvantagens de teclados assistivos, através de aperfeiçoamentos específicos como proposto pelos autores [Majaranta et al. \(2006\)](#), [Loja \(2015\)](#), [Filho \(2009\)](#), [Kurauchi \(2018\)](#), [Kristensson e Zhai \(2004\)](#).

A digitação ocular letra-a-letra, método mais comum de entrada de texto por indivíduos com mobilidade severamente reduzida, produz uma carga cognitiva elevada ao indivíduo e resulta em baixas taxas de digitação. Uma alternativa eficaz pode ser a entrada de palavras inteiras por meio do trajeto do olhar considerando o modelo linguístico e modo de escrita do usuário. Poucos trabalhos desenvolvidos até a publicação desta tese consideram as especificidades de escrita do usuário no processo de desambiguidade necessário no reconhecimento de forma para digitação.

A hipótese deste trabalho reside na otimização do método de digitação em um teclado virtual assistivo utilizando o caminho percorrido pelo olhar no teclado ao invés da entrada de caracteres individuais utilizando a fixação ocular em cada tecla.

1.5 Metodologia de Pesquisa

A metodologia empregada neste trabalho é em grande parte exploratória mas também experimental, o que é comum no ramo da engenharia e mais especificamente no estudo

de Interações humano-computador (**IHC**). Esta tese compreende um levantamento bibliográfico acerca de dispositivos de Comunicação Aumentativa e Alternativa baseadas em interações oculares, seguido do desenvolvimento de um teclado virtual, algoritmo de decodificação de traçados do olhar em texto e da integração do teclado virtual com um rastreador ocular comercial acessível.

São apresentadas duas revisões de literatura, conteúdo análise bibliográfica e estado da arte. A primeira visou delinear o campo do conhecimento da Tecnologia Assistiva (**TA**), sobretudo relacionada às ferramentas de auxílio a indivíduos com distúrbios neuromotores. Foram detalhadas especificidades da anatomia e movimentos oculares, assim como técnicas de entrada de texto através do movimento dos olhos.

A segunda investigação bibliográfica atentou para o desenvolvimento de teclados virtuais em suas diversas configurações e fins, métodos de otimização, métricas de avaliação e diversos modelos de disposição de teclas existentes.

A etapa prática da investigação compreende o desenvolvimento de uma interface assistiva, de forma iterativa e progressiva, para reconhecimento de texto através da decodificação de formas efetuadas pelo olhar ao percorrer a interface de um teclado virtual. A interface foi concebida como uma ferramenta gratuita e multiplataforma, na forma de uma aplicação web que possa funcionar com diferentes marcas e modelos de rastreadores oculares.

A avaliação e validação do sistema proposto foi realizada utilizando-se hardware genérico, representando computadores comuns à maior parte da população, e um rastreador ocular de baixo custo do modelo Tobii 4c. Os testes empíricos envolveram indivíduos que nunca utilizaram rastreadores oculares, alguns sem problemas de visão e alguns com miopia e/ou astigmatismo, que utilizaram o sistema assistivo de digitação em diferentes layouts de teclado para aferição da acurácia do algoritmo de decodificação e velocidade de digitação.

1.6 Contribuições

Em uma visão geral do estado da arte atual, este trabalho expõe os conceitos relacionados à Tecnologia Assistiva (**TA**), voltada especificamente à Comunicação Aumentativa e Alternativa (**CAA**). O desenvolvimento de teclados virtuais tem sido objeto de pesquisa e desenvolvimento durante as últimas décadas.

Uma discussão se direciona aos distúrbios neuromotores e à necessidade de explorar o movimento dos olhos como forma de entrada de informação em interações humano-computador (**IHC**), culminando no desenvolvimento de um protocolo de comunicação aumentativa e alternativa baseado no rastreamento do olhar para entrada de texto.

A proposição e desenvolvimento da interface de comunicação depende de levantamento bibliográfico extensivo relacionado à elaboração e aperfeiçoamento de teclados virtuais,

especificamente teclados controlados pelos olhos do usuário.

Ademais, esta tese apresenta uma nova abordagem de controle do teclado virtual por meio dos movimentos oculares, visando eliminar o tempo de permanência - *dwell time* - em cada tecla na interface e utilizando por sua vez o trajeto do olhar na interface do teclado para decodificar o texto desejado pelo usuário.

As contribuições deste trabalho são:

□ Contribuições Diretas

- Desenvolvimento de um algoritmo de decodificação de gestos do olhar em palavras;
- Implementação de interface de um teclado virtual em diferentes layouts para digitação utilizando olhos;
- Integração entre interface assistiva e sistema de rastreamento do olhar, em um teclado operado por gestos visuais;

□ Publicações de Artigos

1.7 Estrutura deste Trabalho

No Capítulo 1 são apresentados os objetivos e motivações deste trabalho, seguido da metodologia de pesquisa. A descrição da estrutura da Tese conclui o capítulo.

O Capítulo 2 trata da teoria relacionada à Tecnologia Assistiva (TA), sobretudo voltada à Comunicação Aumentativa e Alternativa (CAA). Inclui ainda aspectos sobre os distúrbios neuromotores que podem ocasionar restrições de comunicação em indivíduos, e uma visão geral a respeito de interfaces de comunicação controladas pelos olhos.

O Capítulo 3 apresenta aspectos anatômicos e funcionais sobre o olho humano e a visão. Explora as formas de entrada de informação por meio da direção do olhar e movimento dos olhos, sua relevância na medicina e interação homem-computador, vantagens e desafios na sua implementação. São expostas algumas das abordagens existentes para comunicação por meio dos movimentos oculares.

O Capítulo 4 compreende aspectos práticos do rastreamento do olhar, independente da finalidade. São especificadas técnicas de rastreamento da posição do olhar, sistemas de visualização de dados de rastreamento ocular e sistemas de rastreamento ocular existentes e disponíveis comercialmente.

O Capítulo 5 aborda os teclados virtuais, em teoria e aplicações na engenharia e IHC. Os elementos básicos de um teclado virtual são explanados, enfatizando os fundamentos da interface, disposição das teclas, métodos de *feedback*. São descritas ainda métricas relacionadas ao desenvolvimento e avaliação de teclados virtuais.

O Capítulo 6 contém a descrição detalhada do software desenvolvido neste trabalho - fruto da otimização e integração de algoritmos existentes com vistas a obter melhor desempenho em precisão e velocidade de digitação. É apresentada a descrição da IHC completa, compreendendo o teclado virtual desenvolvido e sua integração com o hardware utilizado, evidenciando os desafios relacionados à precisão e adequação geral do rastreador ocular ao sistema e configurações de *drivers* de dispositivo.

No Capítulo 7 são apresentados testes efetuados e resultados obtidos, assim como análises quantitativa e qualitativa da solução desenvolvida. Os testes de usabilidade permitem vislumbrar o potencial da técnica de reconhecimento de gestos de palavra aplicada à tecnologia Assistiva.

No Capítulo 8 estão presentes a conclusão, as contribuições feitas por esta tese e sugestões de possíveis trabalhos futuros relacionados.

1.8 Considerações Finais

Este capítulo apresentou a motivação, a justificativa, os objetivos, o problema, a hipótese, a metodologia de pesquisa, as contribuições desta tese e a estrutura deste trabalho.

O próximo capítulo compreende a introdução e conceitos relacionados à comunicação aumentativa e alternativa por meio do movimento dos olhos. Engloba uma revisão sistemática com o intuito de investigar o estado da arte dessas áreas.

Tecnologia Assistiva

2.1 Introdução

Este capítulo descreve o ramo de estudo da Tecnologia Assistiva e sua subárea denominada Comunicação Aumentativa e Alternativa (CAA) - à qual este trabalho se enquadra. São apresentadas noções a respeito dos distúrbios neuromotores e sua capacidade de resultar em condições de restrição motora e de comunicação - sobretudo na situação da Síndrome do Encarceramento¹.

A seguir são descritas interfaces oculares existentes de comunicação assistiva. Por fim é apresentada a arquitetura do sistema de CAA proposto, incluindo os requisitos, desafios e sequência de desenvolvimento do *software*.

2.2 Tecnologia Assistiva

A Tecnologia Assistiva (TA) denomina o conjunto de recursos e serviços que auxiliam na promoção ou ampliação das habilidades funcionais de pessoas com deficiência, promovendo dessa forma a inclusão e a possibilidade de manter uma vida independente (BERSCH, 2008). O termo Tecnologia Assistiva é ainda recente, referindo-se a um conceito em processo de construção e sistematização.

O emprego de recursos de Tecnologia Assistiva remonta aos primórdios da civilização humana ou à pré-história. Instrumentos de TA aumentam o desempenho humano em suas tarefas e comumente nem são percebidas pelo usuário durante sua utilização cotidiana. Ferramentas desenvolvidas com o intuito de melhorar e simplificar qualquer atividade humana são exemplos desses dispositivos. Pode-se citar como exemplo talheres, canetas, automóveis, dispositivos eletrônicos e uma incontável quantidade de equipamentos e ferramentas. Manzini (2005, p. 82) descreve os recursos de TA da seguinte maneira:

¹ Doença neurológica rara, em que ocorre paralisia de todos os músculos do corpo, com exceção dos músculos que controlam o movimento dos olhos ou das pálpebras.

Os recursos de tecnologia assistiva estão muito próximos do nosso dia-a-dia. Ora eles nos causam impacto devido à tecnologia que apresentam, ora passam quase despercebidos. Para exemplificar, podemos chamar de tecnologia assistiva uma bengala, utilizada por nossos avós para proporcionar conforto e segurança no momento de caminhar, bem como um aparelho de amplificação utilizado por uma pessoa com surdez moderada ou mesmo veículo adaptado para uma pessoa com deficiência. (MANZINI, 2005, p. 82)

A área de estudo e desenvolvimento da TA é multidisciplinar, envolvendo diversas áreas do conhecimento com o objetivo de proporcionar melhor qualidade de vida a pessoas com deficiência física. Uma cadeira de rodas controlada através dos olhos é um exemplo de dispositivo de TA destinado a prover independência e qualidade de vida a portadores de deficiência. O projeto de uma cadeira de rodas envolveria as áreas de engenharia, enfermagem, computação, ergonomia entre outras. Como exemplifica Loja (2015), no desenvolvimento da cadeira de rodas a área de saúde seria responsável por definir as especificações do dispositivo e a área de engenharia por projetar e construir o equipamento (LOJA, 2015).

A expressão tecnologia Assistiva só viria a aparecer pela primeira vez em 1988, como mencionam Bersch e Sartoretto (2019):

O termo *Assistive Technology*, traduzido no Brasil como Tecnologia Assistiva, foi criado oficialmente em 1988 como importante elemento jurídico dentro da legislação norte-americana, conhecida como *Public Law 100-407*, que compõe, com outras leis, o ADA - *American with Disabilities Act*. Este conjunto de leis regula os direitos dos cidadãos com deficiência nos EUA, além de prover a base legal dos fundos públicos para compra dos recursos que estes necessitam. Houve a necessidade de regulamentação legal deste tipo de tecnologia, a TA, e, a partir desta definição e do suporte legal, a população norte-americana, de pessoas com deficiência, passa a ter garantido pelo seu governo o benefício de serviços especializados e o acesso a todo o arsenal de recursos que necessitam e que venham favorecer uma vida mais independente, produtiva e incluída no contexto social geral. (BERSCH; SARTORETTO, 2019, online)

Pode-se dizer então que o objetivo fundamental da TA é proporcionar à pessoa com deficiência maior independência, qualidade de vida e inclusão social, de forma a possibilitar a comunicação, controle do ambiente, mobilidade e habilidades de aprendizado e trabalho.

Filho (2009, p. 3) define a Tecnologia Assistiva como "uma ampla gama de equipamentos, serviços, estratégias e práticas concebidas e aplicadas para minorar os problemas funcionais encontrados pelos indivíduos com deficiência". Dessa forma, a TA engloba também processos, estratégias e metodologias relacionadas. O autor descreve o que a legislação norte-americana² - *Public Law 108-364* - entende por serviços de TA:

² Documentação do Sistema Nacional de Classificação dos Recursos e Serviços de TA, dos Estados Unidos, pode ser acessada em: <www.ed.gov/offices/OSERS/NIDRR/Products/National_Classification_System.doc>

- ❑ A avaliação das necessidades de uma TA do indivíduo com uma deficiência, incluindo uma avaliação funcional do impacto da provisão de uma TA apropriada e de serviços apropriados para o indivíduo no seu contexto comum.
- ❑ Um serviço que consiste na compra, *leasing* ou de outra forma provê a aquisição de recursos de TA para pessoas com deficiências;
- ❑ Um serviço que consiste na seleção, desenvolvimento, experimentação, customização, adaptação, aplicação, manutenção, reparo, substituição ou doação de recursos de TA;
- ❑ Coordenação e uso das terapias necessárias, intervenções e serviços associados com educação e planos e programas de reabilitação;
- ❑ Treinamento ou assistência técnica para um indivíduo com uma deficiência ou, quando apropriado, aos membros da família, cuidadores, responsáveis ou representantes autorizados de tal indivíduo;
- ❑ Treinamento ou assistência técnica para profissionais (incluindo indivíduos que provêm serviços de educação e reabilitação e entidades que fabricam ou vendem recursos de TA), empregadores, serviços provedores de emprego e treinamento, ou outros indivíduos que provêm serviços para empregar, ou estão de outra forma, substancialmente envolvidos nas principais funções de vida de indivíduos com deficiência;
- ❑ Um serviço que consiste na expansão da disponibilidade de acesso à tecnologia, incluindo tecnologia eletrônica e de informação para indivíduos com deficiências.

A comunidade europeia também desenvolveu conceituação e legislação própria relativa à Tecnologia Assistiva, denominando-a também como 'Ajudas Técnicas' ou ainda 'Tecnologia de Apoio'. O consórcio *Empowering Users Through Assistive Technology* (EUSTAT) desenvolveu um estudo entre 1997 e 1999 visando a formação de usuários finais de Tecnologias de Apoio, envolvendo pessoas idosas ou com deficiência, seus familiares ou assistentes pessoais. Os documentos do consórcio³ afirmam que a Tecnologia Assistiva "engloba todos os produtos e serviços capazes de compensar limitações funcionais, facilitando a independência e aumentando a qualidade de vida das pessoas com deficiência e pessoas idosas" (EUSTAT, 1999, online). Fica claro na citação a seguir a conceituação pela agência europeia que engloba produtos e também serviços.

É importante ter, à partida, uma noção clara do termo Tecnologias de Apoio (TA), visto tratar-se de uma expressão chave predominante nas presentes Linhas de Orientação. Em primeiro lugar, o termo tecnologia não indica apenas objectos físicos, como dispositivos ou equipamento,

³ Os documentos do consórcio EUSTAT podem ser encontrados em: <www.siva.it/research/eustat/index.html>.

mas antes se refere mais genericamente a produtos, contextos organizacionais ou "modos de agir" que encerram uma série de princípios e componentes técnicos. Uma "tecnologia de acesso a transportes públicos", por exemplo, não consiste apenas numa frota de veículos acessíveis (ex. autocarros com plataforma elevatória), mas engloba toda a organização dos transportes, incluindo controlo de tráfego, implantação das paragens, informações e procedimentos de emissão/validação de bilhetes, serviço de clientes, formação do pessoal etc. Sem uma organização deste tipo, o simples veículo não ofereceria qualquer "transporte público". Em segundo lugar, o termo de apoio é aplicado a uma tecnologia, quando a mesma é utilizada para compensar uma limitação funcional, facilitar um modo de vida independente e ajudar os idosos e pessoas com deficiência a concretizarem todas as suas potencialidades (EUSTAT, 1999, online).

A *International Organization for Standardization (ISO)*⁴ apresenta uma classificação orientada basicamente aos produtos, organizando os dispositivos de TA em 11 classes, subclasses e ainda seções. A norma ISO 9999:2016 (2016) já está em sua quarta iteração e define o termo Ajudas Técnicas como:

"[...] qualquer produto, instrumento, equipamento ou sistema tecnológico, de produção especializada ou comumente à venda, utilizado por pessoa com deficiência para prevenir, compensar, atenuar ou eliminar uma deficiência, incapacidade ou desvantagem [...]" (ISO 9999:2016, 2016, online)

Em âmbito nacional, a definição e regulamentação de dispositivos e serviços relacionados à TA foram responsabilidade do Comitê de Ajudas Técnicas (CAT) em uma iniciativa mais recente que as anteriormente citadas. O comitê foi instituído através da portaria Nº142 em 16 de novembro de 2006 pela Secretaria Especial dos Direitos Humanos da Presidência da República - SEDH/PR. Seria composto de especialistas e representantes de órgãos governamentais em uma agenda de trabalho para apresentar propostas de políticas referentes ao setor de tecnologia assistiva.

Como descreve Bersch (2008), o CAT estava imbuído com os esforços para estruturar as diretrizes da área de conhecimento; realizar levantamento dos recursos humanos que atualmente trabalham com o tema; detectar os centros regionais de referência, objetivando a formação de rede nacional integrada; estimular nas esferas federal, estadual, municipal, a criação de centros de referência; propor a criação de cursos na área de tecnologia assistiva, bem como o desenvolvimento de outras ações com o objetivo de formar recursos humanos qualificados e propor a elaboração de estudos e pesquisas, relacionados à da tecnologia assistiva (BERSCH, 2008).

O CAT, baseado nos documentos dos outros países - ISO, EUSTAT, *Public Law 108-364* - publicou, em 14 de dezembro de 2007, o conceito brasileiro de Tecnologia Assistiva:

⁴ Organização Internacional de Normalização, com sede em Genebra, na Suíça. Tem como objetivo criar normas que facilitem o comércio e promovam boas práticas de gestão e o avanço tecnológico, além de disseminar conhecimentos.

Tecnologia Assistiva é uma área do conhecimento, de característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação, de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social. (CAT, 2007, p.9)

Pode-se perceber que o conceito de Tecnologia Assistiva difere da tecnologia médica ou de reabilitação. A TA diz respeito a recursos ou procedimentos pessoais, que atendem as necessidades diretas do usuário final, provendo autonomia e independência, enquanto os recursos médicos ou de reabilitação são voltados ao diagnóstico ou tratamento na área da saúde e consistem em ferramentas de trabalho para os profissionais da área.

2.2.1 Classificação

Os recursos de TA são classificados segundo seus objetivos funcionais pelas diferentes organizações de normatização. Os documentos técnicos americano, europeu e brasileiro diferem na organização e nomenclatura das categorias. A norma ISO 9999:2016 apresenta classificação voltada prioritariamente a produtos. As 11 classes propostas na legislação original de 2002 estão dispostas na tabela 1.

Tabela 1 – Classificação Tecnologia Assistiva - ISO 9999:2002.

Classe	Descrição
Classe 03	Ajudas para tratamento clínico individual
Classe 05	Ajuda para treino de capacidades
Classe 06	Órteses e próteses
Classe 09	Ajudas para cuidados pessoais e de proteção
Classe 12	Ajudas para mobilidade pessoal
Classe 15	Ajudas para cuidados domésticos
Classe 18	Mobiliário e adaptações para habitação e outros locais
Classe 21	Ajudas para a comunicação, informação e sinalização
Classe 24	Ajudas para o manejo de produtos e mercadorias
Classe 27	Ajudas e equipamentos para melhorar o ambiente, ferramentas e máquinas
Classe 30	Ajudas para a Recreação

Fonte: ISO 9999:2016 (2016)

A classificação adotada pela ISO é amplamente referenciada nas pesquisas relacionadas de todo o mundo, apesar de não contemplar os serviços de tecnologia de apoio e não ser a mais adequada para a organização de programas de formação (EUSTAT, 1999).

O modelo *Horizontal European Activities in Rehabilitation Technology* (HEART) é uma outra classificação relativa à Tecnologia Assistiva, recomendada pelo próprio consórcio EUSTAT (EUSTAT, 1999) por incluir serviços e objetivos educacionais. Esse modelo surgiu no âmbito do Programa TIDE - *Technology Initiative for Disabled and Elderly*

People - da União Europeia e organiza a **TA** em componentes técnicos, humanos e socio-econômicos. Tal classificação advém da constatação de que a pesquisa e desenvolvimento de Tecnologia Assistiva ou de Apoio, deve levar em consideração essa realidade, e estudar soluções, dispositivos, metodologias etc., que compensem ou reduzam as limitações não só do indivíduo, mas também do seu ambiente físico e social (**FILHO, 2009**). A classificação **HEART** é a seguinte:

- ❑ Componentes Técnicos - que consideram os recursos técnicos para o exercício de diferentes atividades;
 - Comunicação
 - Mobilidade
 - Manipulação
 - Orientação

- ❑ Componentes Humanos – que consideram os impactos causados no ser humano pela deficiência;
 - Tópicos sobre a deficiência
 - Aceitação da Ajuda Técnica
 - Seleção da Ajuda Técnica
 - Aconselhamento sobre as Ajudas Técnicas
 - Assistência Pessoal

- ❑ Componentes Sócio-econômicos - que consideram as relações, interações e impactos que podem ser estabelecidos entre o usuário final da **TA** e realidades do seu contexto.
 - Noções básicas de Ajudas Técnicas
 - Noções básicas do Desenho Universal
 - Emprego
 - Prestação de Serviços
 - Normalização/Qualidade
 - Legislação/Economia
 - Recursos de Informação

Essa classificação, apesar de menos utilizada em relação à norma ISO 9999, apresenta abordagem mais ampla indo além da concepção de **TA** voltada somente a produtos e dispositivos e englobando serviços e processos formativos relacionados.

José Tonolli e Rita Bersch formularam outra classificação para a área de Tecnologia Assistiva (BERSCH, 2008). Os autores se basearam em outras classificações do setor, notadamente a Classificação Nacional de Tecnologia Assistiva do Departamento de Educação dos Estados Unidos, e estabeleceram a presença de recursos e serviços para cada subárea definida (LOJA, 2015). Esta classificação foi desenvolvida com finalidade didática, abarcando produtos, dispositivos e também serviços e será adotada neste trabalho, uma vez que possui delimitação clara da subárea de Comunicação Aumentativa e Alternativa (CAA).

A classificação proposta por Bersch (2008) foi utilizada pelo Ministério da Fazenda; Ciência, Tecnologia e Inovação e pela Secretaria Nacional de Direitos Humanos da Presidência da República na publicação da Portaria Interministerial No362, de 24 de outubro de 2012 que trata sobre a linha de crédito subsidiado para aquisição de bens e serviços de Tecnologia Assistiva destinados às pessoas com deficiência e sobre o rol dos bens e serviços (BERSCH, 2008).

Está descrita a seguir a classificação dos recursos de TA proposta por Bersch (2008):

□ Auxílios para a vida diária e vida prática

Recursos que possibilitam autonomia e independência em tarefas rotineiras ou que facilitam o cuidado de pessoas em situação de dependência de auxílio, nas atividades como se alimentar, cozinhar, vestir-se, tomar banho e executar necessidades pessoais. Incluem-se ainda nesta categoria os equipamentos que promovem a independência das pessoas com deficiência visual na realização de tarefas como: consultar o relógio, usar calculadora etc.

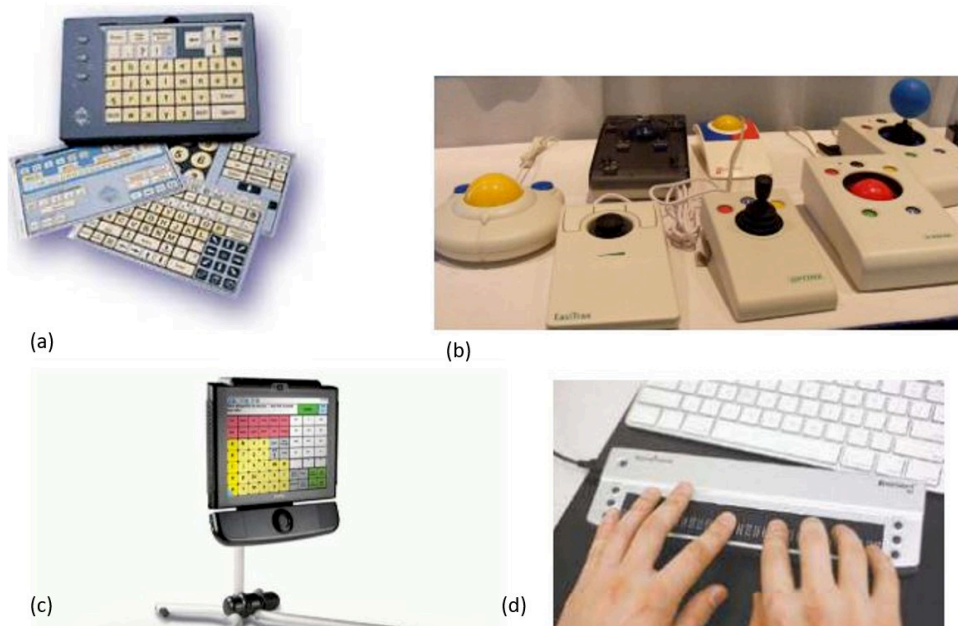
□ CAA - Comunicação Aumentativa e Alternativa

Destinada a atender pessoas sem fala ou escrita funcional ou em defasagem entre sua necessidade comunicativa e sua habilidade em falar, escrever e/ou compreender. Recursos como as pranchas de comunicação, construídas com simbologia gráfica, letras ou palavras escritas, são utilizados pelo usuário da CAA para expressar suas questões, desejos, sentimentos, entendimentos. A alta tecnologia dos vocalizadores (pranchas com produção de voz) ou o computador com softwares específicos e pranchas dinâmicas em computadores tipo tablets, garantem grande eficiência à função comunicativa.

□ Recursos de acessibilidade ao computador

Conjunto de hardware e software especialmente idealizado para tornar o computador acessível a pessoas com privações sensoriais (visuais e auditivas), intelectuais e motoras. Inclui dispositivos de entrada (mouses, teclados e acionadores diferenciados) e dispositivos de saída (sons, imagens, informações táteis), como exemplificado na Figura 1.

Figura 1 – (a) Teclado expandido e programável *IntelliKeys*, (b) diferentes modelos de mouse e (c) sistema *EyeMax* para controle do computador com movimento ocular, (d) linha Braille.



Fonte – Bersch (2008, p. 7)

❑ Sistemas de controle de ambiente

Através de um controle remoto as pessoas com limitações motoras, podem ligar, desligar e ajustar aparelhos eletroeletrônicos como a luz, o som, televisores, ventiladores, executar a abertura e fechamento de portas e janelas, receber e fazer chamadas telefônicas, acionar sistemas de segurança, entre outros, localizados em seu quarto, sala, escritório, casa e arredores. No campo da Tecnologia Assistiva a automação residencial visa a promoção de maior independência no lar e a proteção, a educação e o cuidado de pessoas idosas, dos que sofrem de demência ou que possuem deficiência intelectual.

❑ Projetos arquitetônicos para acessibilidade

Projetos de edificação e urbanismo que garantem acesso, funcionalidade e mobilidade a todas as pessoas, independentemente de sua condição física e sensorial. Adaptações estruturais e reformas na casa e/ou ambiente de trabalho, através de rampas, elevadores, adequações em banheiros, mobiliário entre outras, que retiram ou reduzem as barreiras físicas.

❑ Órteses e próteses

Próteses são peças artificiais que substituem partes ausentes do corpo. Órteses são colocadas junto a um segmento corpo, garantindo-lhe um melhor posicionamento, estabilização e/ou função. São normalmente confeccionadas sob medida e servem no

auxílio de mobilidade, de funções manuais (escrita, digitação, utilização de talheres, manejo de objetos para higiene pessoal), correção postural, entre outros. A Figura 2 ilustra modelos de prótese e órtese de membro superior e inferior, respectivamente.

Figura 2 – Próteses de membros superiores e órtese de membro inferior.



Fonte – Bersch (2008, p. 8)

❑ Adequação Postural

Um projeto de adequação postural diz respeito à seleção de recursos que garantam posturas alinhadas, estáveis, confortáveis e com boa distribuição do peso corporal. Indivíduos que utilizam cadeiras de rodas serão os grandes beneficiados da prescrição de sistemas especiais de assentos e encostos que levem em consideração suas medidas, peso e flexibilidade ou alterações musculoesqueléticas existentes, como pode-se ver na Figura 3.

Figura 3 – Desenho representativo da adequação postural, módulo postural em cadeira de rodas e crianças sentadas em posição adequada.



Fonte – Bersch (2008, p. 9)

❑ Auxílios de mobilidade

A mobilidade pode ser auxiliada por bengalas, muletas, andadores, carrinhos, cadeiras de rodas manuais ou elétricas, *scooters* e qualquer outro veículo, equipamento ou estratégia utilizada na melhoria da mobilidade pessoal.

- ❑ Auxílios para ampliação da função visual e recursos que traduzem conteúdos visuais em áudio ou informação tátil

São exemplos: Auxílios ópticos, lentes, lupas manuais e lupas eletrônicas; os softwares ampliadores de tela. Material gráfico com texturas e relevos, mapas e gráficos táteis, software OCR em celulares para identificação de texto informativo etc.

- ❑ Auxílios para melhorar a função auditiva e recursos utilizados para traduzir os conteúdos de áudio em imagens, texto e língua de sinais

Auxílios que incluem vários equipamentos (infravermelho, FM), aparelhos para surdez, sistemas com alerta tátil-visual, celular com mensagens escritas e chamadas por vibração, software que favorece a comunicação ao telefone celular transformando em voz o texto digitado no celular e em texto a mensagem falada. Livros, textos e dicionários digitais em língua de sinais. Sistema de legendas (*closed-caption / subtitles*).

- ❑ Mobilidade em veículos

Acessórios que possibilitam uma pessoa com deficiência física dirigir um automóvel, facilitadores de embarque e desembarque como elevadores para cadeiras de rodas (utilizados nos carros particulares ou de transporte coletivo), rampas para cadeiras de rodas, serviços de autoescola para pessoas com deficiência.

- ❑ Esporte e Lazer

Recursos que favorecem a prática de esporte e participação em atividades de lazer.

Deve-se discernir a Tecnologia Assistiva e seus recursos de outras tecnologias aplicadas na área médica e de reabilitação. A TA pode ser claramente entendida como o "recurso do usuário" em contraste ao "recurso do profissional", visto que a TA busca promover autonomia e eficiência nas tarefas cotidianas dos usuários.

Bersch (2008, p. 12) diz ainda que a Tecnologia Educacional pode ser facilmente confundida com a Tecnologia Assistiva, e cita o seguinte:

Podemos afirmar então que a tecnologia educacional comum nem sempre será assistiva, mas também poderá exercer a função assistiva quando favorecer de forma significativa a participação do aluno com deficiência no desempenho de uma tarefa escolar proposta a ele. Dizemos que é tecnologia assistiva quando percebemos que retirando o apoio dado pelo recurso, o aluno fica dificuldades de realizar a tarefa e está excluído da participação. (BERSCH, 2008, p. 12)

Ainda com relação à legislação nacional, o Plano Nacional de Educação (PNE) (BRASIL, 2014) estabelece entre as suas metas uma específica para Educação inclusiva. A Meta 4 normatiza o atendimento educacional especializado (AEE) e orienta a comunidade educativa e as políticas públicas voltadas para educandos com deficiência (intelectual, física, auditiva, visual e múltipla), transtorno global do desenvolvimento (TGD) e altas habilidades. O instrumento faz uso do termo Tecnologia Assistiva e orienta a sua aplicação à política pedagógica de inclusão de indivíduos com necessidades especiais no ensino regular. Segue trecho do PNE que menciona a adoção de recursos de TA:

Manter e ampliar programas suplementares que promovam a acessibilidade nas instituições públicas, para garantir o acesso e a permanência dos (as) alunos (as) com deficiência por meio da adequação arquitetônica, da oferta de transporte acessível e da disponibilização de material didático próprio e de recursos de tecnologia assistiva, assegurando, ainda, no contexto escolar, em todas as etapas, níveis e modalidades de ensino, a identificação dos (as) alunos (as) com altas habilidades ou superdotação;

[...]

Fomentar pesquisas voltadas para o desenvolvimento de metodologias, materiais didáticos, equipamentos e recursos de tecnologia assistiva, com vistas à promoção do ensino e da aprendizagem, bem como das condições de acessibilidade dos (as) estudantes com deficiência, transtornos globais do desenvolvimento e altas habilidades ou superdotação. (BRASIL, 2014, online)

2.3 Comunicação Aumentativa e Alternativa (CAA)

A Comunicação Aumentativa e Alternativa (CAA) representa uma subárea da Tecnologia Assistiva (TA) com foco clínico e de pesquisa no campo das patologias da fala. Como descrito por Loja (2015) a CAA reúne os métodos e as tecnologias projetadas para auxiliar ou substituir a comunicação oral de pessoas com limitação de fala (LOJA, 2015).

A Associação Americana de Fala, Linguagem e Audição (ASHA) descreveu a AAC como o "esforço de estudar e quando necessário compensar deficiências temporárias ou permanentes, limitações de atividades e restrições de participação de pessoas com distúrbios graves na produção e/ou compreensão da linguagem falada e escrita" (ASHA, 2019).

O campo de estudo e prática da CAA é reconhecido internacionalmente. Mais de 50 países são membros da Sociedade internacional de Comunicação Aumentativa e Alternativa (ISAAC, 2011). Os serviços e equipamentos de CAA são encontrados na maioria dos países ocidentais, acessíveis para crianças e adultos com necessidades comunicacionais complexas. No entanto, os mesmos recursos são escassos ou de acesso proibitivo em grande parte dos países em desenvolvimento.

A CAA possui, de forma geral, dois propósitos principais com dois públicos-alvo distintos: (a) melhorar a capacidade comunicativa de indivíduos com fala ininteligível ou não-fluente, e (b) fornecer uma forma de comunicação alternativa para pessoas que não

podem falar ou não possuem capacidade de desenvolver uma linguagem suficiente para efetuar uma comunicação eficiente. Não é possível fornecer uma estimativa exata da quantidade de indivíduos que necessitarão, em algum momento de suas vidas, de recursos de CAA, mas uma estatística conservadora aponta que esse número é superior a dezenas de milhões em todo o mundo (BEUKELMAN et al., 2007).

Os recursos de comunicação Aumentativa e Alternativa são geralmente prescritos para indivíduos com pouca ou nenhuma capacidade de fala, situação que pode ser ocasionada por diferentes condições médicas e de incapacidade - podendo ser classificadas como congênitas, adquiridas, progressivas ou temporárias (ZANGARI; KANGAS; LLOYD, 1988). Condições congênitas associadas a necessidades complexas de comunicação incluem o autismo, paralisia cerebral, deficiência intelectual e afasia⁵. Estima-se que de 30 a 50% das crianças diagnosticadas com autismo terão problemas no desenvolvimento da fala e necessitarão, portanto, de algum tipo de recurso de CAA.

Com relação aos distúrbios adquiridos, conforme sua natureza e gravidade os indivíduos também podem necessitar de Comunicação Aumentativa e Alternativa. Esse grupo envolve indivíduos com lesão cerebral adquirida, pessoas com distúrbios neurológicos diversos (por exemplo, doença do neurônio motor, esclerose lateral amiotrófica, esclerose múltipla) e indivíduos que sofreram acidente vascular cerebral ou danos na medula espinhal.

As condições progressivas que podem estar relacionadas ao emprego de recursos de CAA incluem a distrofia muscular, AIDS entre outros. Exemplos de condições temporárias cuja utilização de CAA pode ser benéfica são cirurgias ou situações que resultem em perda da fala temporária.

A CAA é geralmente classificada em dois tipos, a depender se a comunicação é ou não assistida. Os sistemas de CAA assistidos são os que requerem dispositivos auxiliares como placas de comunicação contendo letras, símbolos de desenho de linhas, ou fotos. Outros tipos de sistemas CAA assistidos podem fazer uso de livros ilustrados, cartões, símbolos táteis (como o Braille por exemplo) e telefones de texto (também chamados de teletipo).

Avanços recentes e significativos na área propiciaram o desenvolvimento de dispositivos comerciais que reproduzem a fala, digitalizada (gravada) ou sintetizada. Esses dispositivos de geração de voz (SGD em inglês) já são bastante comuns nas aplicações de CAA. Os símbolos gráficos utilizados para denotar palavras ou sentenças são comumente empregados em conjunto com a CAA assistida.

Sistemas de CAA não-assistidos envolvem o uso de gestos e sinais manuais. Os gestos ou sinais efetuados pelo indivíduo podem ser formais (gestos de cabeça para 'sim' ou 'não', por exemplo) ou informais (podem envolver gestos idiossincrásicos como mover músculos do rosto em padrões específicos para indicar ação ou necessidade). Muitos indivíduos que

⁵ Afasia é uma disfunção de linguagem que pode envolver deficiência na compreensão ou expressão de palavras ou equivalentes não verbais de palavras. (Michaelis, 1999)

necessitam da Comunicação Aumentativa e Alternativa são instruídos na utilização de símbolos manuais para comunicação.

Investigações relacionadas procuram comparar a CAA assistida e não-assistida, com o intento de apontar suas vantagens, desvantagens, e qual a mais adequada. Argumenta-se que a CAA não-assistida é mais conveniente e versátil por não requerer dispositivo auxiliar. Por outro lado, defende-se a CAA assistida afirmando-se que o uso de *Speech-generating Device* (SGD) torna melhor a compreensão do ouvinte.

2.3.1 Componentes da CAA

Beukelman e Mirenda (2012) notam em seu trabalho que a CAA deve ser descrita como um sistema com quatro componentes: símbolos, auxílios, estratégias e técnicas. Isso significa que esse ramo de estudo não se refere somente a modos 'não-falados' nos sistemas de comunicação, mas também a um conjunto de estratégias e técnicas de intervenção que possibilitem a comunicação efetiva através dos símbolos e auxílios da CAA.

Hill (2010) por sua vez, propõe a segmentação do ambiente de CAA em três componentes: primário, secundário e terciário (HILL, 2010), como pode ser observado na Figura 4. Loja (2015) descreve os componentes, como propostos por Hill:

Os componentes primários indicam a maneira como o ambiente de comunicação representa a linguagem “natural”. A interface com o usuário, os métodos de controle e seleção e a saída do software são componentes secundários. Os componentes terciários mostram a relação entre a adequação e a expectativa do usuário ao utilizar o sistema de comunicação. (LOJA, 2015, p. 32)

□ Componentes primários

Pode-se representar a linguagem nos sistemas CAA utilizando figuras de significado único, métodos baseados no alfabeto ou compactação semântica. São métodos independentes da tecnologia empregada, utilizados para desenvolver os métodos e protocolos de CAA (LOJA, 2015).

Pode-se utilizar símbolos gráficos como fotos, desenhos, gráficos ou animações para representar a linguagem natural. Como descreve Hill (2010) esse tipo de representação da linguagem facilitam a comunicação do usuário, auxiliando inclusive indivíduos com pouca ou nenhuma alfabetização, apesar de requerer um grande banco de dados de imagens.

A representação que utiliza elementos baseados no alfabeto utiliza a ortografia e técnicas de otimização de entrada de informação. Como descreve Loja (2015):

Os sistemas que usam essa representação utilizam teclados virtuais, ou um vetor com alfabeto ou até mesmo uma lista de letras

Figura 4 – Componentes de CAA considerados por equipes da área durante a correspondência indivíduo-tecnologia.



Fonte – Adaptado de Hill (2010)

e palavras para gerar uma mensagem. Para usar esse tipo de sistema é necessário que o usuário saiba ler e escrever. (LOJA, 2015, p. 32)

A representação da linguagem através da compactação semântica ou ícones de multi significado foi proposta e patenteada por Baker (1986). O método desenvolvido utiliza ícones para representar palavras e/ou frases. Dessa forma, não é requerido que os usuários possuam alfabetização.

□ Componentes secundários

Os componentes secundários estão relacionados à tecnologia adotada no desenvolvimento do sistema CAA. Esse componente engloba a interface de usuário, os métodos de seleção e o tipo de saída empregado, como pode-se ver no diagrama da Figura 4. Pode-se dividir este componente em duas partes distintas: dispositivo de entrada e *software* de comunicação.

O dispositivo de entrada é utilizado para capturar qualquer tipo de intenção voluntária do indivíduo. Pode-se efetuar a seleção direta, entrada por teclado, ou apontamento utilizando a cabeça ou a direção do olhar. Sinais fisiológicos também podem ser utilizados como forma de entrada, como nos casos de indivíduos com Síndrome do Encarceramento que podem utilizar cinco tipos de sinais biológicos para efetuar comandos computacionais - movimento dos olhos, das atividades cerebrais, por meio do sopro, da capacidade de fala parcial ou das expressões faciais.

Os sinais biológicos são convertidos pelos dispositivos de entrada em informações e comandos para o computador. A direção do olhar, por exemplo, pode ser adquirida por câmeras de vídeo e a leitura do sinal dos músculos oculares pela Eletro-oculografia (EOG), como descreve Loja (2015).

É possível obter múltiplas formas de entrada simultaneamente, combinando sinais de naturezas distintas de forma a obter maior precisão e efetividade do sistema.

□ Componentes terciários

Os componentes terciários de um sistema de Comunicação Aumentativa e Alternativa incluem os periféricos e recursos integrados do sistema, além de elementos referentes a serviços como treinamento e suporte. Modalidades de utilização do sistema e aspectos relativos à satisfação e/ou conforto do usuário também estão englobados neste componente.

É importante que os usuários do sistema de CAA tenham domínio dele, conhecendo suas funcionalidades de forma a potencializar a utilização da ferramenta e melhorar a qualidade de vida dos mesmos. Hill (2010) salienta que devem ser efetuados treinamentos periódicos de utilização, e que o suporte por telefone também é elemento

importante na otimização do treinamento dos pacientes na utilização do sistema de CAA.

Loja (2015) afirma ainda:

Outro fator importante a ser considerado é a complexidade da interface. Interfaces muito complexas podem gerar insatisfação no usuário diante das dificuldades de se adaptar a tecnologia. Além disso, o que é complexo ou simples para um usuário pode não ser para um outro. Assim, os sistemas de CAA devem possuir interfaces simples, intuitivas e que possam se adaptar a cada usuário. (LOJA, 2015, p. 39)

2.3.2 Áreas de pesquisa

As pesquisas relacionadas à Comunicação Aumentativa e Alternativa buscam desenvolver ferramentas inovadoras e criativas para auxiliar a comunicação de indivíduos com pouca ou nenhuma capacidade de fala. Diversas investigações recentes enfocam o desenvolvimento da linguagem e de habilidades de alfabetização entre crianças com deficiências de desenvolvimento, como a paralisia cerebral (SOTO GLORIA; ZANGARI, 2009).

Há pesquisas que procuram encontrar o melhor uso para a CAA em populações com problemas estruturais de educação e para profissionais de saúde. Lancioni et al. (2008) e também Mirenda (2001) exploram a utilização de recursos de CAA, eletrônicos ou não, para comunicação básica de crianças consideradas dentro do espectro do autismo (LANCIONI et al., 2008; MIRENDA, 2001).

Grande parte dos esforços relacionados à pesquisa e desenvolvimento na área de CAA é realizada pelos AAC-Rehabilitation Engineering Research Centers (AAC-RERC) nos Estados Unidos da América. Esses centros de pesquisa integram pesquisa acadêmica e desenvolvimento industrial, desenvolvendo projetos em áreas como:

- ❑ Projeto de novos sistemas de CAA que integrem informação contextual para auxiliar a comunicação de adultos com deficiências intelectual ou linguística;
- ❑ Desenvolvimento de novas interfaces entre sistemas de CAA e indivíduos com distúrbios neuromotores severos;
- ❑ Investigação de sistemas de Comunicação Alternativa dedicados à comunicação interpessoal ('cara-a-cara'), integrando novas tecnologias e habilidades de interação social (BEUKELMAN; MIRENDA, 2012).

2.4 Distúrbios Neuromotores

Nesta seção são descritos os distúrbios neuromotores que ocasionam restrição de comunicação, sobretudo a comunicação verbal e linguagem corporal. Os distúrbios neurodegenerativos mais severos incluem a Esclerose Lateral Amiotrófica - *Amyotrophic Lateral*

Sclerosis (ALS), paralisia cerebral e tetraplegia, além do caso mais extremo causado pela Síndrome do Encarceramento, onde o indivíduo perde completamente suas funções motoras dos membros superiores e inferiores, mantendo apenas estímulos motores limitados como o movimento dos olhos (LOJA, 2015).

O termo neurodegenerativo provém da composição com o prefixo “neuro”, que significa células nervosas, e “degenerativo” que na presente situação possui o sentido de perda da função ou desestruturação de um tecido ou órgão. Os distúrbios neurodegenerativos correspondem, portanto, a qualquer condição patológica que acomete os neurônios e podem ser denominados distúrbios neuromotores ou ainda distúrbios neuromusculares.

Como salienta Loja (2015, p. 31) os distúrbios neurodegenerativos "representam um grande grupo de doenças neurológicas com expressões heterogêneas clínicas e patológicas que afetam um subconjunto de neurônios que possuem funções específicas do sistema anatômico". A degeneração neuronal surge por razões em muitos casos não identificadas e progride inevitavelmente sendo na maioria dos casos fatal.

Esse tipo de doença causa a morte ou o mau funcionamento das células do sistema nervoso de forma excessiva e progressiva de determinadas áreas do cérebro. Portanto, pacientes portadores de doenças neurodegenerativas podem perder sua capacidade cognitiva e as funções motoras.

A Esclerose Lateral Amiotrófica uma condição neurodegenerativa paralisante e aceleradamente progressiva. A maioria dos indivíduos acometidos pela ELA apresenta sobrevida de 2 a 5 anos, uma vez que ainda não há cura e os tratamentos existentes são limitados (BALL; FAGER; FRIED-OKEN, 2012).

Paralisia Cerebral é uma desordem do movimento e da postura, persistente, porém variável, surgida nos primeiros anos de vida pela interferência no desenvolvimento do sistema nervoso central, causada por uma desordem cerebral não progressiva. Esse tipo de deficiência pode causar danos irreversíveis ao sistema motor, impedindo que o paciente realize ações que necessitam de destreza fina e até mesmo impossibilitando o movimento dos membros superiores e/ou inferiores (LOJA, 2015).

A tetraplegia por sua vez pode ser originada por distúrbios neurodegenerativos como os supracitados ou por algum dano causado na espinha dorsal do indivíduo. As capacidades cognitivas dos pacientes tetraplégicos permanecem intactas. Em casos extremos nos quais a capacidade de fala também é prejudicada, essas pessoas ficam impossibilitadas de estabelecer comunicação com seus semelhantes e com o meio em que vivem (BALL; FAGER; FRIED-OKEN, 2012).

2.4.1 Esclerose Lateral Amiotrófica

A Esclerose Lateral Amiotrófica não tem cura até o momento de publicação desta tese, e é caracterizada pela perda progressiva dos movimentos do corpo. O impacto da ELA na

vida de um indivíduo pode depender de sua idade, tipo de esclerose, equipe de assistência e decisões sobre intervenções para prolongamento da vida.

No caso da ELA, a parte cognitiva em geral permanece intacta enquanto a parte motora é totalmente danificada. Como aponta (BERTAZZI et al., 2017) a ELA é um distúrbio neurodegenerativo de origem desconhecida, é progressivo e está associado à morte do paciente em um tempo médio de três a quatro anos. Porém há casos de sobrevida longa. Sua incidência é estimada em 1 a 2,5 indivíduos portadores a cada 100.000 habitantes por ano, com uma prevalência de 2,5 a 8,5 por 100.000 habitantes (LOJA, 2015).

Devido ao comprometimento dos neurônios motores superiores e inferiores, os distúrbios motores da fala resultam em uma disartria espástica-flácida mista. Os sintomas iniciais da fala entre pessoas com ELA variam, mas um padrão pode ser observado inicialmente com a diminuição da velocidade da fala, seguida por reduções moderadas e mais graves na inteligibilidade que afetam a eficácia comunicação do indivíduo. A progressão do impacto na fala acaba deixando o indivíduo sem capacidade de fala (BALL; FAGER; FRIED-OKEN, 2012).

Devido à perda progressiva da fala, o padrão de cuidado para indivíduos com ELA indica o tratamento focado em estratégias para otimizar a fala natural - falar o máximo possível - focando nas capacidades de linguagem expressiva.

Embora a insuficiência respiratória seja a causa mais comum de morte na ELA, a ventilação mecânica pode prolongar a expectativa de vida, mas não interromperá a progressão implacável da paralisia.

Sistemas CAA devem ser aplicados de acordo com cada caso de ELA, e de forma geral devem incluir meios para comunicar mensagens médicas básicas e informações para diálogo cotidiano, uma maneira de conversar e interagir casualmente com outras pessoas, um meio de acessar e usar o telefone, opções para solicitar assistência ou socorro, e maneiras de expressar afeto, humor e emoções. O sistema de CAA deve ser modificado ou adaptado de acordo com a progressão da disartria, contendo componentes do sistema que vão desde o alfabeto básico ou botões de símbolos até sistemas geradores de fala (SGDs) avançados.

Os avanços na tecnologia possivelmente trarão maior capacidade de interação aos indivíduos portadores da ELA através de ferramentas de CAA, possibilitando a continuidade de atividades sociais, recreativas, educacionais, comerciais, voluntárias e de emprego. Como menciona Ball, Fager e Fried-Oken (2012, p. 691) "os indivíduos com ELA que utilizam ferramentas de comunicação assistiva continuam a usar os dispositivos até o fim da vida ou algumas semanas antes da morte.

2.4.2 Síndrome do Encarceramento

A Síndrome do Encarceramento (SE) consiste no caso mais extremo de restrição de comunicação, no qual a pessoa perde completamente suas funções motoras dos membros superiores e inferiores - tetraplegia - mantendo apenas estímulos motores limitados como,

por exemplo, o movimento dos olhos. Os indivíduos geralmente preservam a consciência e o movimento vertical dos olhos, facilitando a comunicação não verbal.

Observa-se nos registros uma taxa de aproximadamente 80% dos indivíduos acometidos pela SE com sobrevivência de 10 anos. Um trabalho de recuperação física, ainda que limitada, pode melhorar a qualidade de vida dos pacientes e permitir voltem a viver com suas famílias. O encaminhamento precoce a serviços especializados de reabilitação e aplicação de tecnologias de comunicação, portanto, imprescindível (SMITH; DELARGY, 2005).

Existem três tipos de SE: incompleta, clássica e completa. Na SE incompleta o indivíduo possui movimentos voluntários limitados, como o movimento de um dedo ou parte do rosto. Pessoas que possuem a SE clássica conseguem apenas movimentar os olhos e piscá-los e na SE completa o paciente fica impossibilitado de realizar movimentos voluntários em qualquer parte do corpo. Essa síndrome pode ser causada por doenças neurodegenerativas como as citadas anteriormente e até mesmo por tetraplegia (LOJA, 2015).

Engenheiros de reabilitação e fonoaudiólogos têm trabalhado no desenvolvimento de IHC assistivas destinadas ao auxílio de pacientes da condição extrema da SE. Smith e Delargy (2005) defendem a importância do desenvolvimento e emprego de ferramentas de CAA na melhoria da qualidade de vida de indivíduos acometidos pela síndrome do encarceramento:

Os computadores tiveram um efeito libertador na vida de pessoas com síndrome do encarceramento, permitindo-lhes realizar diálogos, se comunicar e acessar a Internet. Com a utilização de dispositivos de comunicação aumentativa em conjunto com computadores, os pacientes que tornados incapazes de falar podem encontrar uma maneira de se comunicar amplamente. (SMITH; DELARGY, 2005, p. 408)

Pacientes que recuperam o movimento distal conseguem apontar para as letras utilizando os olhos, possibilitando a utilização de teclados assistivos, tábuas de comunicação ou dispositivos avançados de produção de fala.

2.5 Interfaces Oculares

Serão descritas nesta seção interfaces oculares de comunicação, única alternativa viável de dispositivo de TA para indivíduos acometidos por distúrbios neuromotores em estágio avançado, como a SE. Serão detalhadas as características e requisitos de hardware e software para este tipo específico de ferramenta de CAA, e serão referenciadas pesquisas relacionadas desenvolvidas nos últimos anos.

Devido à progressão contínua da Esclerose Lateral Amiotrófica e consequente perda da função dos membros superiores, as ferramentas de comunicação assistiva adotadas nestas situações são baseadas em tecnologias de controle ocular. Os movimentos oculares de indivíduos com ELA geralmente não são afetados, apesar de deficiências ocasionais terem sido identificadas. A utilização de sistemas de CAA baseados no rastreamento

ocular podem se mostrar essenciais e bastante eficazes nessas circunstâncias. tornando o uso experimental da tecnologia AAC baseada no olhar fixo essencial para determinar a eficácia.

Estudos anteriores buscaram examinar o movimento ocular em indivíduos com ELA e identificaram diminuição da velocidade sacádica e alguns padrões sacádicos anormais, além da diminuição geral dos movimentos dos olhos (DONAGHY et al., 2011). Não foram encontradas até o momento investigações acerca da fadiga ocular em pacientes com ELA, mas os relatórios indicam que o olhar pode ser um meio eficaz de entrada de informação.

A tecnologia de rastreamento ocular por infravermelho projeta uma luz no espectro infravermelho invisível que não provoca danos à visão, que causa um reflexo nas pupilas do indivíduo. A tecnologia de rastreamento ocular emprega a triangulação dos reflexos de forma a rastrear a posição do olhar em relação ao dispositivo rastreador e a uma tela. Ao fixar o olhar em um ponto por alguns instantes, o sistema seleciona o item e realiza um comando que pode ser uma mensagem em texto, uma seleção em um menu ou mesmo a produção de voz sintetizada. Em comparação com a varredura de itens na tela e a seleção do item desejado utilizando um acionador, o controle ocular produz mensagens mais rápidas com menores taxas de erro para pessoas com ELA (GIBBONS; BENETEAU, 2010).

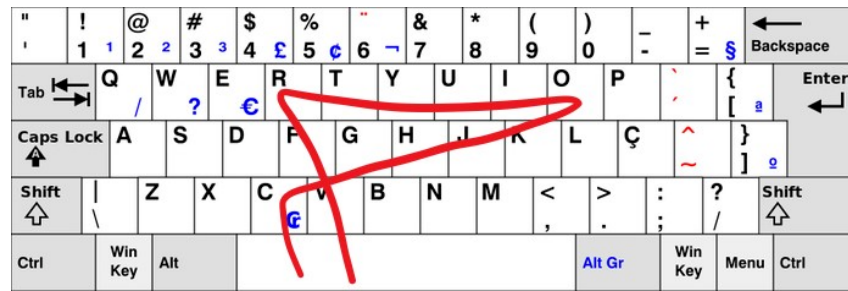
2.6 Arquitetura do Sistema Proposto de Digitação por Gestos do Olhar

Após abordar as particularidades dos dispositivos de comunicação assistida com entrada de informação através dos olhos, propõe-se uma otimização incremental de teclado ocular com entrada de texto através do reconhecimento do traçado do olhar (também chamado gesto ocular).

Este trabalho propõe o desenvolvimento de um teclado virtual para entrada de palavras através de formas realizadas pelo olhar do usuário, utilizando um algoritmo de decodificação do traçado do olhar combinado com um modelo linguístico do português brasileiro para mapear os gestos oculares em palavras de um léxico pré-definido.

O sistema desenvolvido não requer a fixação do olhar do usuário uma letra por vez, funcionando de forma a registrar o caminho percorrido pelos olhos na interface de um teclado virtual em cada palavra e retornar uma lista de prováveis palavras que o usuário deseje digitar. Para digitar a palavra 'cor', por exemplo, o usuário percorre o olhar a partir da tecla de espaço até as teclas 'c', 'o', 'r' e então de volta à tecla de espaço, como observado na Figura 5. O gesto contínuo efetuado pelo olhar corresponde a uma representação espacial da palavra 'cor', a ser reconhecida pelo decodificador do teclado.

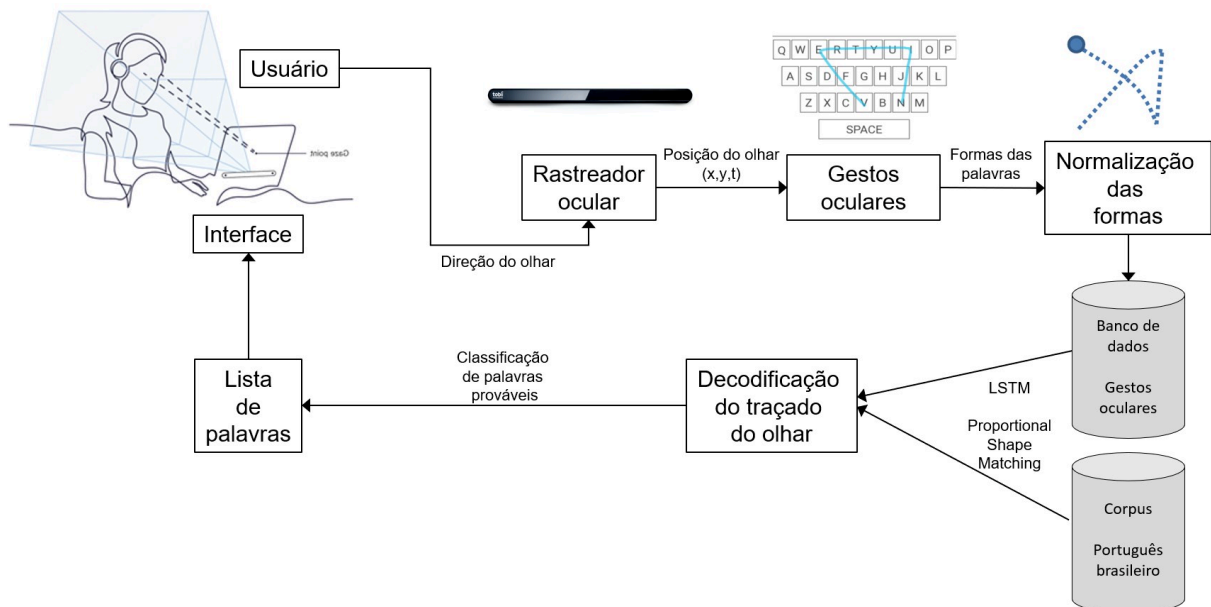
A Figura 6 representa o diagrama de arquitetura do sistema de CAA proposto. O desenvolvimento do sistema abrange a implementação da arquitetura, da interface e do

Figura 5 – Exemplo de digitação da palavra *cor* através de gesto.

Fonte – O autor

algoritmo de reconhecimento dos gestos oculares, além da integração com rastreador ocular comercial.

Figura 6 – Diagrama do sistema de CAA proposto.



Fonte – O autor

Requisitos do sistema proposto:

- ❑ Deve ser capaz de reconhecer todas as palavras do léxico através de traçados oculares unívocos;
- ❑ deve operar em 'tempo real', efetuando o reconhecimento de um gesto em uma fração do tempo gasto para se digitar uma palavra: tempo médio de reconhecimento menor que 500ms;
- ❑ Deve possuir um léxico com quantidade de palavras suficiente para a elaboração de textos genéricos em português brasileiro: léxico mínimo de 2500 palavras;

- ❑ Não deve ocupar espaço excessivo em memória, não possuindo mais que algumas dezenas de *megabytes*⁶.
- ❑ Deve apresentar relativa robustez com relação a entradas ruidosas como erros ortográficos, letras repetidas e tremulações inerentes dos movimentos oculares;

2.7 Considerações Finais

Este capítulo apresentou os conceitos de Tecnologia Assistiva e sua subárea denominada Comunicação Aumentativa e Alternativa. Foi abordado o desenvolvimento dessa área de estudo assim como o panorama nacional relacionado. Foram descritos distúrbios neuromotores que podem ocasionar restrições de comunicação e a necessidade de utilização de ferramentas de CAA. No final deste capítulo foi apresentada a proposta de interface de CAA a ser desenvolvida neste trabalho, detalhando seus princípios de funcionamento e requisitos.

O próximo capítulo contém informações acerca da utilização da direção do olhar e/ou movimentos oculares como fonte de informação, seja para fins de controle ou de comunicação. São apresentadas características anatômicas e funcionais do olho humano, assim como a natureza dos movimentos oculares. O capítulo trata ainda de aplicações que utilizam os movimentos dos olhos para fins de diagnóstico, controle de mouse e teclado (com foco na digitação ocular através de gestos).

⁶ No reconhecimento de fala, modelos linguísticos geralmente não devem exceder o espaço em memória de 5 a 10 Mb, suficiente para permitir a modelagem de até duzentas mil palavras. (OUYANG et al., 2017)

Direção do Olhar como Fonte de Informação

3.1 Introdução

O estudo e desenvolvimento de interfaces baseadas nos movimentos oculares requer a compreensão sobre a anatomia e funcionamento do olho humano. Os conceitos relacionados aos movimentos oculares também são essenciais, e serão apresentados neste capítulo. São trazidas informações acerca de interfaces de entrada de texto através dos olhos - que podem ser chamadas de teclados oculares.

3.2 Aspectos do controle de interfaces através do olhar

O movimento dos olhos possui características singulares, servindo por vezes como única forma de comunicação e interação com o mundo exterior por pessoas com deficiências neuromotoras severas como a esclerose lateral amiotrófica - [ALS](#) e síndrome do encarceramento - *Locked-in syndrome* ([LIS](#)). A utilização dos movimentos oculares como método de entrada e interação apresenta vantagens e desvantagens.

Utilizar a direção do olhar como fonte de informação não é uma tarefa trivial, uma vez que necessita de equipamento ou *software* específico para o rastreamento ocular que possua altas taxas de atualização e capacidade de processamento em tempo real da entrada de dados. A imprecisão dos rastreadores oculares também representa um desafio para pesquisas e desenvolvimentos relacionados, além dos efeitos de escorregamento (*drifting*) e descalibração que também trazem dificuldades na decodificação e utilização da informação implícita no olhar do usuário.

Apesar das inconveniências citadas, existem atualmente diversas soluções e equipamentos modernos (e baratos) o suficiente para que se possa capturar com altíssima precisão e resolução temporal as coordenadas de posição relativas à direção do olhar. Os olhos

humanos se movimentam em alta velocidade, possibilitando uma alta taxa de entrada de informação e velocidade de interação com um computador, e a tecnologia atual possibilita o rastreamento não invasivo de forma relativamente simples. Os autores [Pi e Shi \(2017, p. 1\)](#) apontam o seguinte:

Com o auxílio da tecnologia de rastreamento ocular, o olhar do usuário pode ser localizado em uma tela 2D de um computador e ser utilizado como controle de comandos. Atividades cotidianas podem ser controladas pelo olhar, como interação com computadores, controle de cadeiras de rodas, controle de um braço robótico. ([PI; SHI, 2017, p. 1](#), tradução própria)

Uma interface controlada pela direção do olhar pode oferecer diversos benefícios, segundo [Drewes \(2010\)](#):

❑ Facilidade de uso

O olhar pode ser considerado uma forma natural de entrada de informação. É extremamente fácil focar em um objeto somente olhando para ele. É uma ação tão intuitiva que requer pouco ou nenhum treinamento (o que pode não ser o caso para pessoas com deficiência ([BRASIL, 2015](#))). Pode-se reduzir o estresse nos músculos das mãos e braços com a utilização do controle através dos olhos. Isso não significa que os músculos oculares serão sobrecarregados, uma vez que estes se moveriam de qualquer maneira para outros elementos de interação. Quando utilizamos o mouse por exemplo, ao se clicar em um botão na tela nossos olhos geralmente se direcionam para naturalmente para a posição do botão.

❑ Maior velocidade de interação

Os movimentos oculares permitem uma entrada de dados significativamente mais rápida em comparação a outras formas de entrada. Um usuário normalmente olha para o objeto de destino antes de operar um dispositivo mecânico de apontamento. A direção do olhar pode então servir de indicação para o objetivo do usuário antes mesmo que ele possa atuar qualquer outro mecanismo de entrada ([JACOB; KARN, 2003](#)). Os sistemas de digitação ocular existentes ainda se mostram mais lentos que os teclados tradicionais (mecânicos ou virtuais), mas a combinação da direção do olhar com outro tipo de dispositivo entrada pode ser uma solução para se atingir maiores velocidades de digitação.

❑ Dispensa de manutenção

Os rastreadores oculares passivos operam sem necessitar de contato físico com o usuário, o que resulta na desnecessidade de manutenção do equipamento. Não há necessidade de limpeza frequente do dispositivo, como é o caso para mouses e teclados.

❑ Interface higiênica

Em ambientes de interação pública ou com altas exigências higiênicas (uma sala de cirurgia, por exemplo) uma interface controlada pelo olhar seria extremamente útil, visto que a interação com o dispositivo não requer contato físico.

❑ Controle remoto

O rastreamento ocular passivo por vídeo possibilita também o controle remoto de interfaces. Câmeras de alta resolução e lentes de zoom tornam possível a detecção do olhar a alguns metros de distância. Rastreadores de baixo custo como o *Microsoft Kinect* (Kang Wang; Ji, 2016), por exemplo, já se mostram dispositivos de interação remota extremamente avançados.

❑ Informações sobre a atividade do usuário

Diferente de outros dispositivos de entrada, um rastreador ocular pode detectar o foco de atenção do usuário. O simples ato de apontar utilizando um rastreador ocular se traduz na mudança do ponto de atenção do indivíduo, onde cada mudança pode ser interpretada pela interface não somente como um comando de apontamento, mas também como uma indicação de comportamento sem que haja uma ação de entrada explícita da parte do usuário. Uma análise simples da informação de um rastreador ocular pode detectar atividades como a leitura, enquanto um reconhecimento profundo pode revelar condições físicas ou emocionais, idade e até mesmo grau de instrução do usuário.

Observa-se na Figura 7 o trajeto típico do olhar durante a atividade de leitura. Os pontos marcados representam as letras onde o olho permanece fixo nas palavras durante a leitura, com duração expressa em milissegundos pelos números (JEGERSKI; VANPATTEN, 2013). As sacadas, que efetuam o deslocamento do olhar de uma palavra a outra, possuem duração média de 30 a 50 ms, tão rápidos que o leitor deste texto ainda não notou que já fez algumas centenas somente nesta porção da página.

Como mencionado anteriormente, o rastreamento ocular como fonte de informação também apresenta diversos desafios:

❑ Capacidade de controle

Os olhos efetuam movimentos involuntários e inconscientes o tempo todo, o que pode atrapalhar a decodificação de informações como um dispositivo de entrada. Ainda não é exatamente claro para a ciência o quanto precisamente um indivíduo consegue controlar os movimentos dos olhos voluntariamente por longos períodos de tempo. A

Figura 7 – Trajeto dos movimentos oculares durante a leitura.



Fonte – Jegerski e VanPatten (2013, on-line)

capacidade de controlar os olhos envolveria suprimir os movimentos não proposicionais¹ e efetuar somente os movimentos desejados. É evidente que conseguimos controlar a direção do olhar, mas não se sabe até onde se consegue desenvolver as capacidades motoras oculares como se faz ao aprender a digitar com um teclado, por exemplo.

❑ Conflito entre controle e visão livre

Diferente de um mouse, os movimentos do olho não podem ser desativados. Com o mouse podemos clicar e soltar seus botões, informando à interface quando ativar ou não a entrada, o que não acontece com os olhos e pode resultar em conflito. É um problema bem conhecido na área de estudo do rastreamento ocular chamado de 'Toque de Midas' (*Midas Touch*) - dificuldade de interfaces oculares em decidir se o olhar em uma região só visa inspecionar ou executar um comando (PI; SHI, 2017). Se o usuário fecha os olhos para, por exemplo, não entrar com a informação de ponteiro, o mesmo não consegue enxergar as informações ou interagir com a interface. Esse é um problema comum de novos dispositivos de entrada passivos ou que não requerem comandos - necessitam de um desenvolvimento cuidadoso da interface para separar movimentos naturais da entrada de forma intencional.

❑ Fadiga visual

¹ Os olhos se deslocam rapidamente de um ponto a outro mesmo quando o indivíduo acha que está olhando fixamente para um único objeto, o que poderia ativar automaticamente um comando em uma interface controlada pelo olhar. (JACOB; KARN, 2003)

Sabe-se que a utilização por longos períodos de tempo de certos músculos ou grupos musculares pode causar problemas físicos como a LER. Há preocupações de que as lesões também possam ocorrer nos músculos oculares. Ainda que os olhos se movam o tempo todo, mesmo durante o sono, deve-se considerar cuidadosamente as consequências à saúde de uma utilização exagerada que vá além da fadiga visual.

Um rastreador ocular fornece dados brutos na forma de coordenadas x e y correspondentes à posição do olhar de um indivíduo na tela do computador. Os pontos observados - da expressão em inglês *Point of regard* (POR) - podem ser utilizados em tempo real ou registrados para posterior análise ou reprodução, permitindo a identificação dos pontos de interesse do usuário. Os pontos observados são comumente utilizados na caracterização de movimentos mais complexos como fixações e sacadas (COSTA, 2017).

Os rastreadores oculares existem há décadas (DUCHOWSKI, 2017), mas somente nos últimos anos eles se tornaram mais baratos e capazes. Antes destinados a experimentos de laboratório e aplicações de ponta, pode-se encontrar atualmente rastreadores voltados a consumidores finais em aplicações reais de interfaces humano-computador (IHC) como Comunicação Aumentativa e Alternativa (CAA) ou jogos eletrônicos².

Necessita-se atualmente de pesquisas voltadas ao aperfeiçoamento de interfaces que incorporem o rastreamento ocular de forma natural e inobstrusiva. Existem inúmeras pesquisas e soluções voltadas à implementação de interfaces e métodos de rastreamento ocular, mas o verdadeiro desafio reside na integração com Inteligência Artificial e processamento de linguagens naturais como a reconhecimento de fala e análise semântica.

Em resumo, a análise dos movimentos oculares permite caracterizar um número elevado de alterações visuais, neurológicas ou psicológicas no ser humano, sendo uma fonte eficiente de informação e controle.

3.3 Olho Humano e Visão

De forma simplificada, utilizando termos da engenharia, podemos comparar o olho e seus músculos a uma câmera com estabilização de imagem. O olho funciona de forma a converter luz - que consiste em radiação eletromagnética em diferentes comprimentos de onda³ - em impulsos nervosos e transmiti-los ao cérebro. Essa conversão é efetuada pela retina, localizada na parte posterior do olho (TORTORA, 2003).

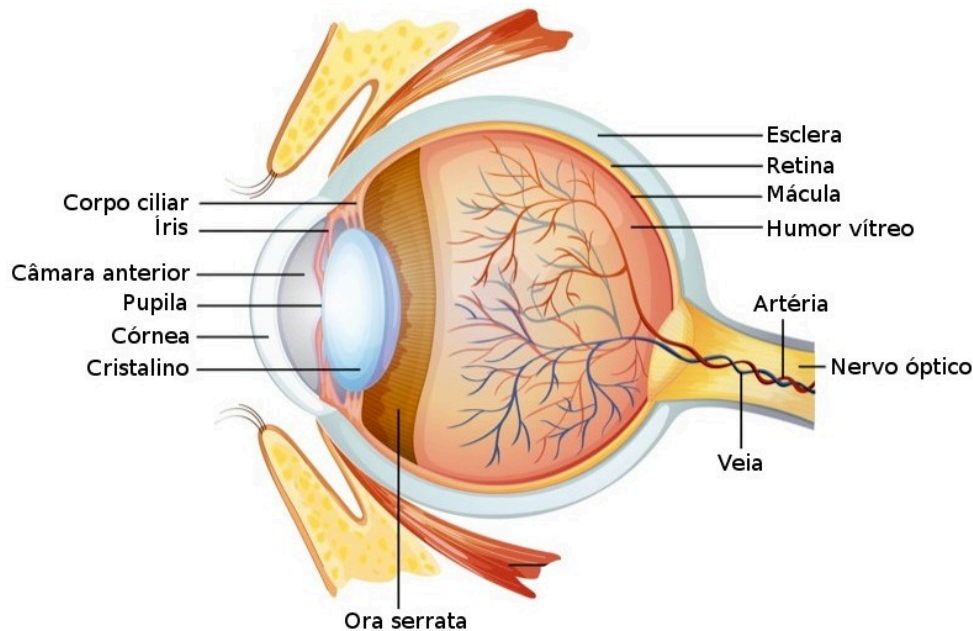
Como dito anteriormente, a anatomia do olho se mostra similar à construção de uma câmera: Há um diafragma, chamado íris, que permite o ajuste da abertura. O orifício na íris é chamado de pupila, como pode-se observar na Figura 8. Diferente de uma câmera,

² O Tobii 4c é um rastreador ocular de baixo custo voltado à experiência com jogos eletrônicos. Os dados fornecidos pelo Tobii 4c incluem as coordenadas do olhar e dos eventos de fixação. (Tobii, 2016)

³ Os comprimentos de onda da luz visível são classificados segundo a cor, do violeta, que apresenta comprimento de onda $\lambda = 4 \times 10^{-7}m$, ao vermelho, cujo comprimento de onda é $\lambda = 7 \times 10^{-7}m$. (SILVA, 2012)

que utiliza lentes de distância focal fixa e ajustam o foco alterando sua distância, a lente ocular pode alterar a distância focal modificando sua forma. A área fotossensível interna traseira do olho é chamada retina, que converte a luz em impulsos nervosos.

Figura 8 – Anatomia do olho humano.



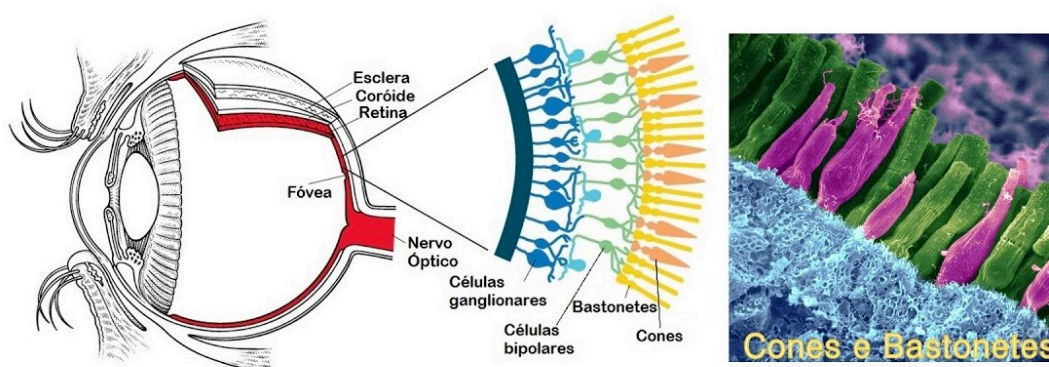
Fonte – [Silva \(2012, on-line\)](#)

A área fotossensível presente na retina apresenta forma elipsoide com um eixo horizontal de 180° e eixo vertical de 130° ([DREWES, 2010](#)). Os neurônios especializados na transdução de luz em impulsos elétricos são denominados cones e bastonetes, como pode-se observar na [Figura 9](#). Os bastonetes são altamente sensíveis ao brilho (atuam na visão noturna, por exemplo), ao passo que não possuem alto reconhecimento de cromaticidade. Os cones, por sua vez, são menos sensíveis mas são destinados ao reconhecimento de cores. Os seres humanos normalmente possuem três tipos de cones, sensíveis a faixas de comprimentos de onda correspondentes às cores vermelho, verde e azul.

A retina humana apresenta três regiões com propriedades distintas, uma vez que os bastonetes e cones não estão uniformemente distribuídos pela retina. A maior parte da retina possui uma baixa concentração de receptores, estando a maioria presentes na sua região circular central com diâmetro aproximado de 1 a 2° . A fóvea contém em torno de 150.000 cones/ mm^2 e um pequeno número de bastonetes, sendo uma região de alta acuidade visual. Em torno da fóvea - região parafoveal, cobrindo cerca de 5° do campo visual - a quantidade de cones diminui para 20.000 cones/ mm^2 . Por essa razão, o campo visual fora da fóvea possui baixa resolução, servindo principalmente à percepção de movimento ambiente (visão periférica). A [Figura 10](#) representa a diferença de acuidade visual nas diferentes regiões da retina ([LAND, 2006](#)).

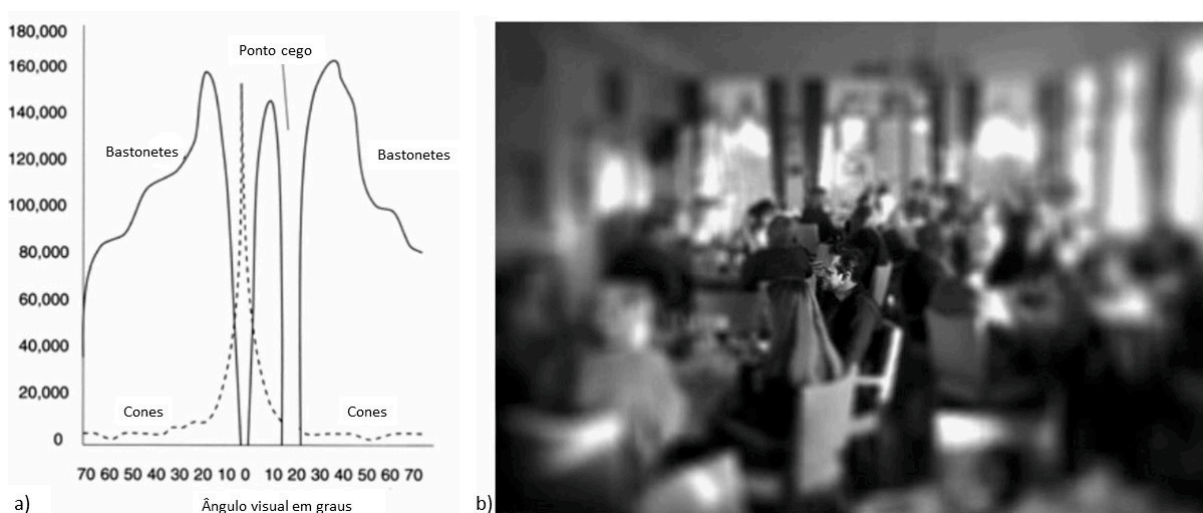
Seis músculos - chamados de músculos extraoculares - conectam o olho à cavidade

Figura 9 – Células fotossensíveis: cones e bastonetes.



Fonte – Silva (2012, on-line)

Figura 10 – (a) Distribuição de densidade de cones e bastonetes; (b) Diferença de acuidade entre visão central e periférica.



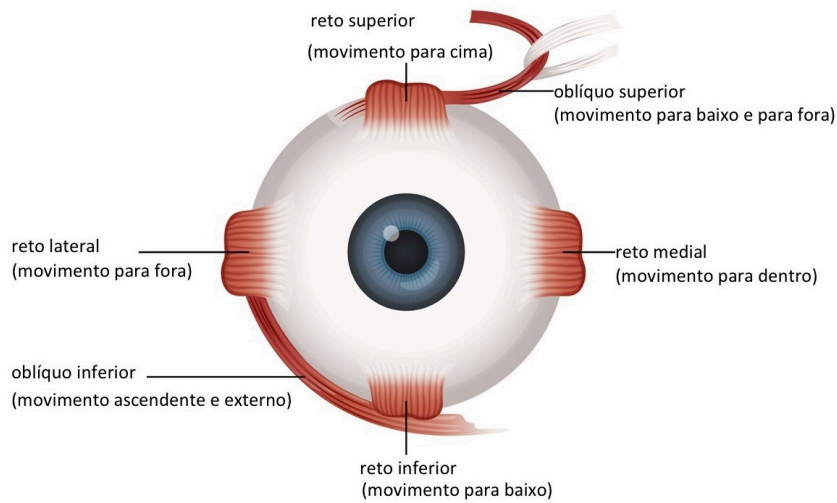
Fonte – Drewes (2010, p. 34)

ocular e são responsáveis pelos movimentos do globo ocular. Esses músculos oferecem seis graus de liberdade - três translações no interior da cavidade ocular e três rotações - e são dispostos de forma antagônica, como ilustrado na Figura 11. Um par de músculos é responsável pelo movimento vertical, outro pelo movimento horizontal e o terceiro par possibilita movimentos rotacionais na direção de visão (CARPENTER, 1988).

Os três pares de músculos são responsáveis pelos rápidos movimentos oculares e asseguram a compensação de todos os movimentos da cabeça. Por isso, os nervos que controlam os músculos oculares estão conectados com o órgão responsável pelo equilíbrio, localizado no ouvido. O sistema neural responsável pelos movimentos oculares é denominado núcleo oculomotor (DUCHOWSKI, 2017).

Quatro músculos são chamados retos, enquanto os outros dois são denominados oblíquos (CARPENTER, 1988).

Figura 11 – Músculos do olho humano.



Fonte – [Silva \(2012, on-line\)](#)

Os músculos retos são os seguintes:

- ❑ reto inferior - permite o movimento do olho para baixo: movimento descendente;
- ❑ reto superior - permite o movimento do olho para cima: movimento ascendente;
- ❑ reto medial - permite o movimento do olho para o nariz: movimento para dentro;
- ❑ reto lateral - permite o movimento do olho em direção à têmpora: movimento para fora.

Os músculos oblíquos, por sua vez, são os seguintes:

- ❑ oblíquo inferior - músculo extraocular mais curto. Permite o movimento do olho em direção à têmpora e uma elevação do campo de visão.
- ❑ oblíquo superior: músculo oculomotor mais longo. Permite o movimento do olho para o nariz e uma redução do campo de visão.

3.4 Movimentos Oculares

Quando se fala em movimentos vitais repetitivos efetuados pelo corpo se pensa nos movimentos cardíacos e pulmonares, mas pode-se pensar também no movimento dos olhos. Estes são provavelmente os movimentos mais frequentes, esteja o indivíduo acordado ou dormindo. Estima-se que o ser humano faça cerca de 170.000 movimentos sacádicos (subseção 3.4.1) por dia.

Os movimentos oculares representam uma infindável fonte de informações úteis para investigações científicas e clínicas. A psicologia e biologia, por exemplo, se interessam

pela relação entre os movimentos oculares e fenômenos cerebrais. Para a clínica médica - neurologia, oftalmologia, otorrinolaringologia etc. - anormalidades na mobilidade ocular podem fornecer pistas diagnósticas (LEIGH; ZEE, 2015).

Nos últimos 60 anos as pesquisas relacionadas à mobilidade ocular aumentaram significativamente graças a avanços na tecnologia de rastreamento ocular e na teoria psicológica relacionando o rastreamento ocular e processos cognitivos (JACOB; KARN, 2003). Os movimentos oculares têm sido usados desde então para se obter melhor compreensão de patologias envolvendo desde a distrofia muscular ao autismo (FAN et al., 2009), passando por distúrbios neuromotores, distúrbios de atenção e marketing digital (a Figura 16 serve como exemplo). Processos cognitivos como memória, tomadas de decisão, atenção, retribuição assim como desordens psicóticas também podem ser associados a padrões de movimentos oculares (COVRE et al., 2005).

Os movimentos do olho são mais fáceis de interpretar em relação a outras partes do corpo. São essencialmente restritos a rotações dos globos oculares⁴ uma vez que as translações (deslocamentos lineares) são insignificantes. Isso facilita a captura e medição precisa dos movimentos, necessárias para análises quantitativas.

Obtemos informação sobre nossos arredores através do olhar, que está relacionado ao nosso foco e atenção. Nós olhamos para alguma coisa fixando o olhar nela, mantendo-o relativamente estático por um curto intervalo de tempo (200 a 600 ms geralmente) denominado fixação. Entre as fixações o olhar realiza saltos de um ponto a outro, saltos que são chamados sacadas ou movimentos sacádicos (MAJARANTA; RÄIHÄ, 2002).

Os seres humanos, assim como a maior parte dos primatas, movimentam os olhos segundo uma combinação de cinco movimentos básicos: sacádicos, de fixação, de perseguição, vergenciais e de reflexo vestibulo-ocular. (DUCHOWSKI, 2017).

Alguns dos movimentos oculares citados só podem ser registrados através de rastreadores oculares extremamente precisos e rápidos⁵, e costumam apresentar utilidade apenas em exames clínicos específicos. Portanto, a maioria das aplicações atuais e publicações relacionadas adotam uma classificação diferente, não utilizando os cinco movimentos acima descritos. Classifica-se neste trabalho, portanto, os movimentos oculares em sacadas, fixações, perseguições, caminhos percorridos, áreas de interesse e duração das *gazes* (chamadas neste trabalho de 'duração do olhar') (LEIGH; ZEE, 2015). A utilização de uma taxonomia diversa é referida por Giannotto:

Normalmente, os movimentos vergenciais e reflexo vestibulo-oculares são ignorados por esta taxonomia, mas ainda precisam ser diferenciados das sacadas. Eles ocorrem, respectivamente, quando os olhos se movem em direções opostas para mudar a distância focal; e quando movem-se os

⁴ Outros movimentos como adaptação e acomodação se referem ao foco da lente ocular ou dilatação pupilar, e não à mobilidade do próprio globo ocular.

⁵ A frequência de atualização da maioria dos rastreadores oculares atuais é de cerca de 50 Hz. Em alguns casos este número pode ultrapassar os 2000 Hz. (WATANABE, 2013)

olhos para compensar movimentos da cabeça ou do corpo, mantendo o foco no mesmo ponto. (GIANNOTTO, 2009, p. 56)

Pode-se, então, fazer observações importantes a respeito da análise dos movimentos oculares. Primeiramente, necessitamos da modelagem de somente três movimentos para analisar a localização do foco visual: fixações, perseguições e sacadas. Em segundo lugar, baseado nas características dos impulsos neurais relacionados, podemos modelar os três tipos de movimento citados por sistemas lineares invariantes no tempo (SLIT) (DUCHOWSKI, 2017).

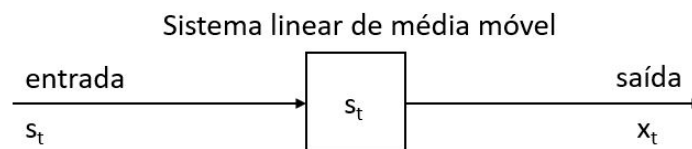
3.4.1 Sacada

Uma sacada (*saccade* ou movimento sacádico) é um movimento balístico com duração típica de 30 a 120 ms. É um movimento rápido - podendo alcançar velocidades de até 500°/s - efetuado para reposicionar a fóvea no centro das áreas de interesse do indivíduo. É dito que as altas velocidades das sacadas reduzem a acuidade visual desses movimentos, que ocorrem entre as fixações. Isso faz com que pouca ou nenhuma informação possa ser registrada durante esse movimento (CLIFTON; RAYNER, 2015).

Modelos matemáticos de geração sacádica procuram descrever o mecanismo e geração dos sinais enviados aos neurônios motores. Há grande consenso da comunidade científica de que o sinal observado se parece com uma função impulso/degrau relacionada ao comando duplo de velocidade e posição aos músculos extraoculares (DUCHOWSKI, 2017). Uma possível representação do degrau sacádico é um filtro derivativo, acoplado a um integrador para produzir as sacadas, como descrito por Carpenter (1988).

O princípio da integração neural proposto para os movimentos sacádicos podem ser aplicados a todos os tipos de movimentos oculares. Segundo Leigh e Khanna (2006) os circuitos neurais que derivam no cérebro e cerebelo efetuam a integração dos sinais conjugados dos movimentos, incluindo sacadas, perseguições, vergenciais e reflexos.

Figura 12 – Diagrama de um filtro linear simples de modelagem dos movimentos sacádicos.



Fonte – O autor

Um filtro derivativo pode ser modelado como um filtro linear como ilustrado na Figura 12. No domínio do tempo o filtro é modelado de acordo com a equação:

$$x_t = g_0 s_t + g_1 s_{t-1} + \dots = \sum_{k=0}^{\infty} g_k s_{t-k} \quad (1)$$

onde s_t é o impulso de entrada, x_t é o degrau de saída e g_k são os coeficientes do filtro. Um filtro que satisfaz as condições de derivação aproximada é o filtro de Haar com coeficientes $\{1,-1\}$ (DUCHOWSKI, 2017). No domínio Z a função de transferência $X(z)/S(z)$ desse filtro linear corresponde a:

$$\begin{aligned}x_t &= g_0 s_t + g_1 s_{t-1} & (2) \\x_t &= (1)s_t + (-1)s_{t-1} \\x_t &= (1)s_t + (-1)z s_t \\x_t &= (1 - z)s_t \\X(z) &= (1 - z)S(z) \\ \frac{X(z)}{S(z)} &= 1 - z\end{aligned}$$

Esse filtro de Haar possui ordem 2, aproximando-se da derivada primeira entre pares sucessivos de entradas (DUCHOWSKI, 2017).

A Figura 13 representa a resposta do movimento ocular a um estímulo sacádico na forma de degrau (por exemplo um ponto luminoso que se desloca horizontalmente). Percebe-se que há um período de latência - o qual pode variar de 100 a 1000 ms - após a qual os olhos se movem rapidamente em direção à nova posição, com um pequeno erro de posição que é eliminado quando os olhos se movem novamente para a nova posição (MALMIVUO; PLONSEY, 2012).

Uma vez iniciado, o movimento sacádico não pode ser interrompido ou ter a direção modificada - por isso é classificado como um movimento balístico. Em geral, as sacadas apresentam amplitude de movimento menor que 15° , visto que movimentos maiores requerem movimentos da cabeça ao mesmo tempo que os olhos são redirecionados. Segundo os vários autores aqui citados, a maioria das sacadas possuem trajetória curva ou oblíqua.

Carpenter (1988) efetuou medições para a amplitude e duração das sacadas e encontrou uma relação linear entre o tempo T de uma sacada e sua amplitude A :

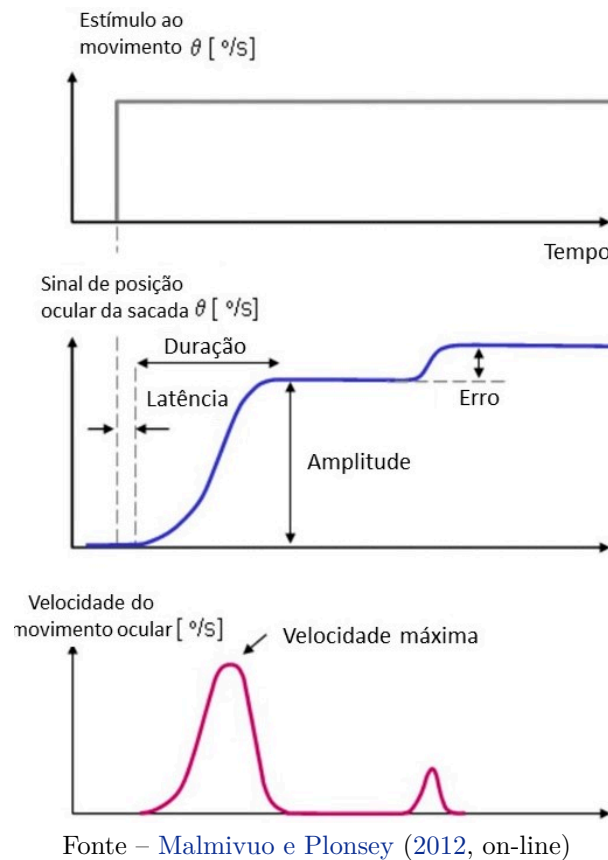
$$T = 2,2ms/^\circ \cdot A + 21ms \quad (3)$$

Em 1989 Abrams, Meyer e Kornblum (1989) sugeriram um modelo onde a força muscular é crescente no tempo, o que significa que a aceleração $a(t)$ aumenta de forma constante com o tempo:

$$a(t) = k \cdot t \quad (4)$$

onde k é a constante de aceleração. De acordo com os autores essa relação é válida para movimentos angulares de até 15° . A distância $s(t)$ (rotação angular) percorrida no tempo t pode ser calculada integrando-se duas vezes no tempo.

Figura 13 – Movimentos oculares correspondentes a uma sacada.



$$s(t) = k/6 \cdot t^3 \quad (5)$$

No modelo proposto uma sacada seria um movimento com aceleração linearmente crescente até a metade do deslocamento e com aceleração linearmente decrescente na segunda metade do deslocamento. A relação entre o tempo t e a distância percorrida s seria (onde c é também um valor constante):

$$t = c \cdot s^{1/3} \quad (6)$$

A modelagem acima descrita, proposta por [Abrams, Meyer e Kornblum \(1989\)](#) difere da chamada Lei de Fitts, que publicou o famoso artigo "*The information capacity of the human motor system in controlling the amplitude of the movement*" em 1954 ([FITTS, 1954](#)) onde afirma que a velocidade e precisão do movimento muscular são dependentes.

Assim como os reflexos pupilares, em condições normais os movimentos sacádicos são binoculares, significando que possuem mesma direção e magnitude em ambos os olhos. Nos humanos, espera-se no máximo quatro sacadas por segundo ([CHALUPA; WERNER, 2004](#)).

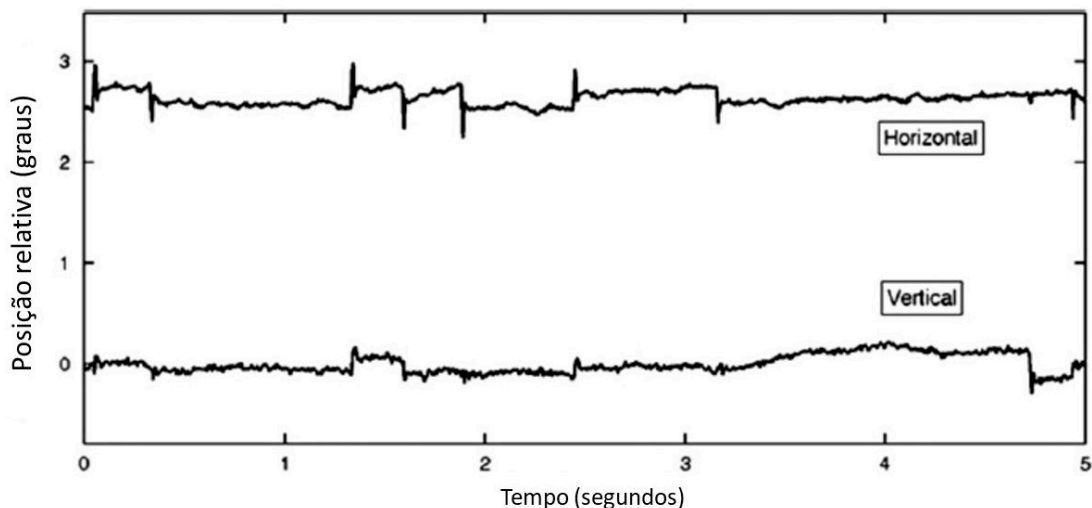
Os movimentos vergenciais⁶ e reflexos vestibulo-oculares⁷ não são considerados nesta taxonomia, apesar de apresentarem similaridades em relação à sacada.

3.4.2 Fixação

A fixação corresponde ao momento em que os olhos permanecem parados, apesar do fato de que os olhos nunca permanecem completamente estáticos. As fixações são caracterizadas, então, pelos períodos onde os movimentos oculares são mínimos, e são denominados movimentos fixionais: os tremores, escorregamentos e microscadadas (CHALUPA; WERNER, 2004). Os movimentos fixionais ocorrem entre as sacadas, e são necessários para se estabilizar a imagem de um objeto estacionário na retina.

O primeiro dos três movimentos fixionais considerados é um tremor constante, também conhecido como nistagmo. Pouco se sabe em relação aos tremores - dada a sua natureza mínima é praticamente impossível mensurá-los de forma não invasiva. Acredita-se que tais microtremores estejam relacionados à percepção da cena, ajudando as células fotorreceptoras a continuarem trabalhando. Ocorrem concomitantemente aos escorregamentos, a uma frequência de 60Hz (GIANNOTTO, 2009). A Figura 14 representa os movimentos oculares durante uma fixação, onde pode-se observar os microtremores presentes mesmo durante a fixação em um ponto estacionário.

Figura 14 – Movimentos oculares durante uma fixação em um ponto estacionário.



Fonte – Kowler (2011, p. 1459)

Os escorregamentos, por sua vez, são movimentos curvos lentos ($< 2^\circ/s$) e de pequena amplitude, provavelmente relacionados a falhas no sistema nervoso de controle.

⁶ Os movimentos de vergência são adaptações do olhar que reposicionam objetos que se encontram a diferentes distâncias ou um mesmo objeto que muda de distância em relação ao observador em relação à fóvea. (MELCHIADES, 2014)

⁷ O reflexo vestibulo-ocular é um movimento ocular de reflexo que estabiliza as imagens na retina durante o movimento da cabeça, ao produzir um movimento ocular na direção oposta ao movimento da cabeça. Assim, a imagem é mantida no centro do campo visual. (GIANNOTTO, 2009)

Um escorregamento gera, como consequência, uma microssacada que corresponde a um movimento muito rápido que reposiciona o olho na posição desejada. Grande parte dos pesquisadores simplesmente desconsidera os micromovimentos fixionais, considerando-os como "ruído" (MAJARANTA et al., 2006).

Embora a classificação de tais micromovimentos como ruído possa ser uma simplificação excessiva do processo, ela permite que o sinal seja modelado por um simples sistema realimentado semelhante ao ilustrado na Figura 15. O ruído aditivo nesse exemplo pode ser representado por $e_t = s_t - x_t$, onde a posição do olho x_t é subtraída da posição de fixação estável s_t na junção de soma. Nesse modelo, o sinal de erro estimula o sistema fixional de forma similar ao sistema de perseguição (*smooth pursuit*) onde e_t é um sinal de erro de posição ao invés de erro de velocidade da Equação 7.

Acredita-se que mais de 80% do tempo do observador seja dedicado à realização de fixações, e que a maior parte da percepção visual seja efetuada nesses instantes. Os movimentos fixionais possuem pelo menos 150 ms, mas a maioria dura de 230 a 300 ms aproximadamente (MELCHIADES, 2014).

Horstmann, Ahlgrim e Glöckner (2009) utilizaram a duração e distribuição das fixações em um estudo sobre tomadas de decisão. Gegenfurtner, Lehtinen e Säljö (2011) usaram a duração e número de fixações e duração de sacadas para analisar a percepção de imagens. Tsai et al. (2012) realizaram investigações sobre a atenção utilizando a duração de fixações. Outros estudos ainda relacionam o número e duração de fixações à estimacão de carga cognitiva⁸ (WANG et al., 2014).

3.4.3 Perseguição

A dinâmica ocular denominada perseguição (do inglês *smooth pursuit*) acontece quando os olhos do observador acompanham um objeto em movimento. É um movimento lento, pequeno e voluntário. Giannotto (2009) cita o movimento de perseguição no seguinte trecho de seu texto:

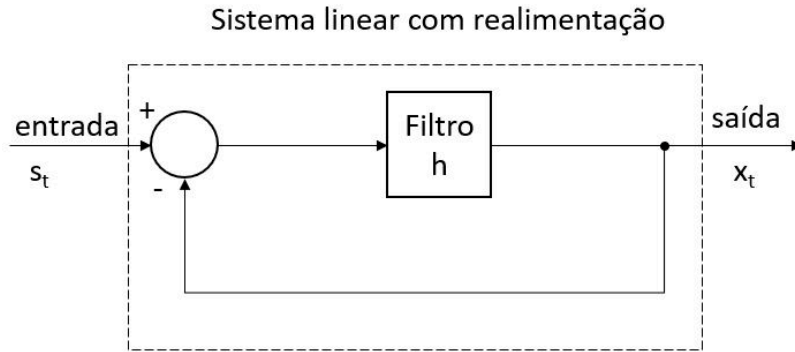
[...] a velocidade da perseguição é destacadamente menor do que a velocidade de uma sacada. Quando um objeto se move muito rapidamente, o observador utiliza sacadas para alcançá-lo. (GIANNOTTO, 2009, p. 57)

O movimento de perseguição é um exemplo de sistema de controle com realimentação negativa (CARPENTER, 1988). Uma simples malha fechada com realimentação é utilizada para modelar o movimento de perseguição, como representado na Figura 15, onde s_t é a posição do alvo, x_t a posição do olho e h é o filtro (linear e invariante no tempo) ou ganho do sistema. A equação da malha no domínio do tempo é a seguinte:

⁸ Carga cognitiva, ou esforço cognitivo, se refere ao nível de utilização de recursos psicológicos como memórias, atenção, percepção, representação de conhecimento, raciocínio e criatividade na resolução de problemas. (Wikipédia, 2016)

$$h(s_t - x_t) = x_{t+1} \quad (7)$$

Figura 15 – Diagrama de um sistema linear com realimentação de modelagem dos movimentos de perseguição.



Fonte – O autor

Ao se aplicar a transformada Z , obtemos a função de transferência $X(z)/S(z)$:

$$H(z)(S(z) - X(z)) = X(z) \quad (8)$$

$$H(z)S(z) = X(z)(1 + H(z))$$

$$\frac{H(z)}{1 + H(z)} = \frac{X(z)}{S(z)}$$

Nesse modelo de malha fechada realimentada os sinais dos receptores visuais constituem o sinal erro indicando a compensação necessária para corresponder ao movimento da imagem do alvo na retina.

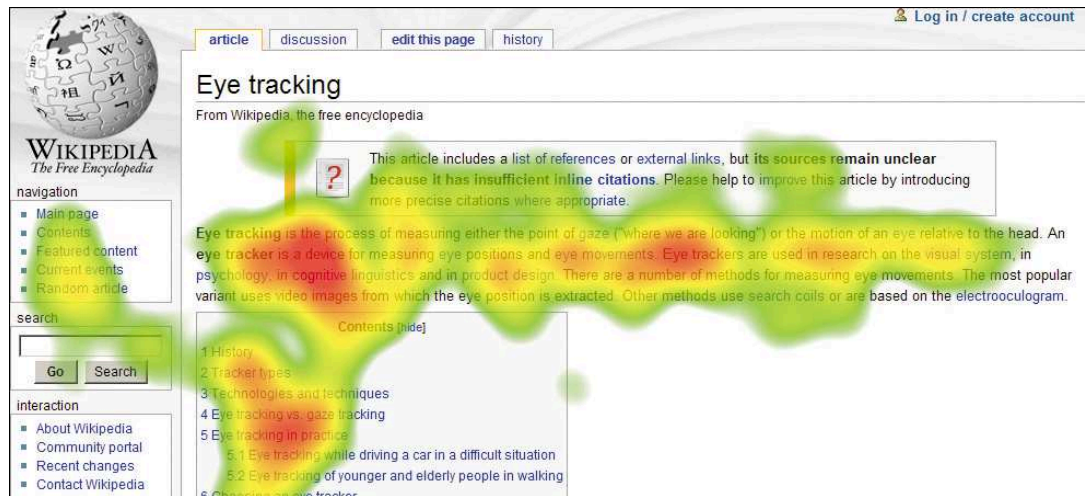
3.4.4 Caminho Percorrido (*scanpath*)

Os caminhos percorridos (ou *scanpaths*) são sequências sucessivas de fixações e sacadas. Pode-se utilizar os caminhos percorridos para deduzir áreas de interesse ou hábitos implícitos no padrão dos movimentos e posições oculares. Um exemplo de caminho percorrido durante a leitura de um texto está ilustrado na Figura 7.

3.4.5 Área de Interesse (AOI)

Segundo [Giannotto \(2009\)](#) as áreas de interesse (AOI) identificam os padrões de fixação de um usuário, podendo traduzir regiões de um sistema ou produto que possuem interesse especial. A Figura 16 representa as áreas de interesse de um usuário ao navegar em uma página da Wikipedia.

Figura 16 – Mapa de calor representando áreas de interesse na leitura de uma página web.



Fonte – Tobii (2016, on-line)

3.4.6 Duração do Olhar

Uma sequência consecutiva de fixações em uma mesma AOI corresponde a uma *gaze* ou duração do olhar, calculada através da soma das fixações individuais que a compõem. A posição é calculada através da média das posições das fixações. A duração da *gaze* é constituída por várias fixações e pequenas sacadas, e é delimitada por fixações que ocorrem fora da AOI (YARBUS, 1967).

Outras denominações para a duração do olhar são *glance*, ciclo de fixações e a mais comum na literatura relacionada, *dwell time*. *Dwell times* variam de 400 a 1000 ms, a depender da natureza do estímulo visual.

3.4.7 Identificação de Fixações e Sacadas

As subseções anteriores descreveram os tipos de movimentos oculares e suas características. A maior parte dos estudos relacionados ao rastreamento ocular considera, no entanto, somente as fixações e sacadas. As principais métricas empregadas nesses trabalhos incluem a posição, duração e velocidade de fixações e sacadas, o número de fixações em uma AOI. Pode-se inferir, portanto, que a identificação de fixações e sacadas é o aspecto mais importante do rastreamento do olhar.

Costuma-se realizar a classificação dos movimentos oculares em fixações e sacadas utilizando-se duas abordagens distintas, que podem apresentar resultados muito diferentes a partir de uma mesma fonte de dados - coordenadas espaciais (GIANNOTTO, 2009). Uma das abordagens, denominada "seletora de sacadas" identifica as sacadas a partir dos dados do rastreador ocular e a partir disso define os períodos entre as posições final e inicial de sacadas consecutivas como sendo fixações. Uma segunda abordagem faz o inverso em um processo de "seleção de fixações", onde se encontram as fixações e se definem como

sacadas os espaços entre duas fixações consecutivas.

A detecção de sacadas requer taxas de atualização muito superiores a 100Hz dos rastreadores atuais, fazendo com que o método de detecção de fixações seja o mais utilizado. Esse processo retorna as posições onde ocorreram fixações e suas durações, estando implícitas as demarcações das sacadas (SALVUCCI; GOLDBERG, 2000).

Giannotto (2009) menciona a taxonomia dos algoritmos de identificação de fixações e sacadas baseada nas características espaciais ou temporais dos movimentos oculares, adotada por Salvucci e Goldberg (2000):

□ Características espaciais

- Baseados na velocidade: são algoritmos que utilizam o fato de que uma fixação é mais lenta que uma sacada;
- Baseados na dispersão: assumem que os pontos observados que compõem uma fixação ocorrem próximos uns dos outros;
- Baseados na área: são algoritmos que agrupam pontos que caiam sobre uma AOI em uma fixação.

□ Características temporais

- Sensíveis à duração: são algoritmos que utilizam o fato de que as fixações costumam durar mais de 100 ms e menos de 400 ms;
- Adaptativo localmente: são algoritmos que tentam se adaptar aos usuários dinamicamente, de forma que a interpretação de um ponto possa ser influenciada por pontos próximos.

Salienta-se que o rastreador ocular utilizado neste trabalho - Tobii 4c (Tobii, 2016) - fornece os dados de posição do olhar baseado nas fixações em relação à tela do computador após o processo de calibração, utilizando algoritmo proprietário não especificado pelo fabricante. Não será necessário, portanto, implementar a porção do algoritmo destinada à identificação das sacadas e fixações.

3.5 Métricas Relacionadas ao Rastreamento Ocular

Em qualquer estudo quantitativo, deve-se atentar meticulosamente à etapa de definição das métricas relacionadas. Ao se investigar o rastreamento do olhar, deve-se considerar métricas adequadas às características e limitações de hardware e software utilizados, além do tipo de tarefa executada pelo usuário e resultado esperado.

Há uma grande diversidade de métricas relacionadas ao rastreamento do olhar, decorrentes da grande produção bibliográfica relativa ao tema nas últimas décadas. É natural, assim sendo, a identificação de diversas métricas diferentes com o mesmo significado ou

métricas com significado diferente, mas que possuem mesmo nome sugerido. Tal inconsistência entre diversos trabalhos é mencionada por [Giannotto \(2009\)](#) em seu trabalho.

Não existe, até o momento, padronização aceita universalmente nos protocolos de investigação e medidas utilizadas na área do rastreamento do olhar. Mesmo utilizando definições semelhantes nos parâmetros e métricas utilizados, pequenas variações nos valores podem resultar em grandes divergências de resultado das análises. Espera-se, portanto, que os pesquisadores exponham em seus trabalhos os protocolos de experimento, incluindo equipamentos utilizados, condições de realização dos testes, métricas e parâmetros empregados.

A maior parte dos estudos de usabilidade adota métricas derivadas da taxonomia empregada na seção 3.4 e resumida na tabela 2, apresentada por [Giannotto \(2009\)](#).

Tabela 2 – Definição dos termos usados na definição das métricas do rastreamento do olhar.

Fixação	São períodos em que os olhos permanecem praticamente parados. Normalmente é definida por dois parâmetros: tempo mínimo necessário para ser considerada uma fixação (ex.: 100ms) e amplitude do movimento máximo aceito (ex.: raio de 50 pixels, quadrado de 40x40 pixels, ou 2°), mas também pode ser definida em relação a sua velocidade máxima (por exemplo, menor que 100°/segundo), dependendo do algoritmo usado para identificar as fixações.
Sacada	São movimentos rápidos usados para reposicionar a fóvea de um ponto de interesse a outro. Quando são usados algoritmos de detecção de fixação é representada implicitamente pela distância de uma fixação a outra.
AOI	Área de interesse. Região que apresenta alguma característica importante para o pesquisador. É definida pelo próprio pesquisador ou sua equipe e não pelos participantes.
Caminho Percorrido	Sequência ordenada de fixações e sacadas executada pelos olhos do usuário.
Duração do olhar	Sequência de fixações consecutivas em uma mesma AOI. A duração da gaze é obtida pela soma da duração das fixações que a compõem, e sua posição é a posição média destas fixações. Uma gaze pode ser composta por diversas fixações e pequenas sacadas, sendo que a primeira fixação que caia fora da AOI marca o fim da gaze.

Fonte: [Giannotto \(2009, p. 74\)](#)

[Jacob e Karn \(2003\)](#) realizaram uma análise extensiva referente ao rastreamento ocular e desenvolvimento de interfaces humano-computador (IHC) apontando mais de 20 trabalhos realizados no período de 1950 a 2003 e as diversas métricas utilizadas. As métricas mais comuns identificadas nos trabalhos são brevemente descritas a seguir:

□ Número de fixações (total):

Número total de fixações executadas durante uma tarefa. Um alto número de fixações indica busca ineficiente de informação (possivelmente causada por um arranjo inadequado dos elementos da interface);

□ Duração média da fixação (total):

Soma da duração de todas as fixações dividida pelo número de fixações total. Fixações longas indicam dificuldade na extração de informações na interface;

□ Número de fixações em cada AOI:

Indicam a importância da AOI para a tarefa ou interface. Outra interpretação é que a AOI é mais perceptível ou mais atrativa que as demais;

□ Duração do olhar (*gaze*) em cada AOI:

Duração da gaze em uma AOI. Usada para verificar quais AOI estão recebendo mais atenção por parte do usuário. Também pode indicar antecipação, quando o usuário observa fixamente uma região da interface aguardando um possível evento;

□ Taxa de fixações (total) (fixações/seg):

Métrica altamente correlacionada com a duração de fixação. Uma vez que o tempo entre fixações (geralmente movimentos sacádicos de curta duração) é relativamente pequeno comparado ao tempo de fixação, a taxa de fixações deve ser aproximadamente o inverso da duração de fixação.

Diversas outras métricas e seus significados são mencionadas por [Giannotto \(2009\)](#). Encontra-se medidas derivadas de variações no diâmetro pupilar, movimentos da cabeça e piscadas do olho, apesar de não serem amplamente adotadas como as anteriormente citadas.

Este trabalho propõe o desenvolvimento de um teclado ocular operado por meio de gestos oculares e possui métricas relevantes que diferem ligeiramente das utilizadas nas investigações relacionadas a carga cognitiva, diagnóstico ou mesmo dispositivos de apontamento através do rastreamento ocular. As métricas mais importantes neste trabalho são o caminho percorrido (*scanpath*), duração média de fixação, taxa de fixações e número de fixações em cada AOI.

As métricas citadas são fundamentais para o desenvolvimento de um teclado virtual ocular, que requer parâmetros meticulosamente definidos como tamanho das teclas, disposição das teclas, método de decodificação do gesto ocular, limiares de tempo de fixação e sacada.

3.6 Aplicações de Diagnóstico

Este trabalho apresenta enfoque na utilização do rastreamento ocular para comunicação assistiva, entretanto os olhos podem fornecer uma enorme quantidade de informações de importância clínica. Muitos sistemas utilizam características oculares como informações diagnósticas. Atributos como a movimentação dos olhos, a dinâmica pupilar, imagem

do fundo do olho, sincronia reflexiva binocular entre outros podem ser importantes na detecção de patologias ou anomalias.

Registros de movimentos oculares são um complemento importante no diagnóstico, compreensão e documentação de uma série de anormalidades neurológicas, incluindo desequilíbrios vestibulares periféricos e centrais, lesões cerebelares e corticais e malformações que resultam em oscilações e interrupções sacádicas, sacadas lentas, perseguições imprecisas, diferentes formas de nistagmo, entre outros.

A pupilometria é umas das aplicações da análise de informações oculares para fins médicos. Consiste em um método simples e não-invasivo capaz de apontar anomalias no funcionamento do sistema nervoso autônomo além de averiguar diversos estados fisiológicos. O teste do reflexo pupilar à luz (PLR) já foi utilizado em trabalhos anteriores na investigação de condições como alcoolismo, uso de drogas, síndrome de Down, depressão, mal de Alzheimer, mal de Parkinson, insuficiência cardíaca, déficit de atenção, diabetes, AIDS, autismo, entre outros. O ciclo pupilar pode ser obtido através de estímulos luminosos aplicados diretamente ao olho e consiste no movimento de contração pupilar seguido de expansão e retorno ao estado inicial (SILVA, 2016).

O teste de registro do movimento dos olhos fornece informações importantes no diagnóstico e tratamento do nistagmo, um problema ocular que faz com que os olhos façam movimentos involuntários. Geralmente afeta os dois olhos e costuma ser mais visível quando os olhos estão voltados para uma direção específica. O nistagmo pode causar problemas de visão e geralmente ocorre com estrabismo - olhos desalinhados - e ambliopia - chamada por vezes de "olho preguiçoso". O teste de registro do movimento ocular faz medições do olho para determinar qual posição faz com que o nistagmo melhore ou piore; quais as distâncias que fazem o nistagmo melhorar ou piorar; e se o nistagmo muda com a direção ou intensidade ao longo do tempo. Todas essas informações são importantes no tratamento do nistagmo, especialmente se uma abordagem cirúrgica para melhorar a visão de uma criança for necessária (BEDELL; STEVENSON, 2013).

A avaliação dos movimento ocular também pode trazer informações clínicas valiosas para pacientes com perda de campo central bilateral. A maioria dos pacientes com perda de campo central bilateral visualiza um alvo através de um ou mais pontos retiniais periféricos em detrimento à fóvea sem visão. O rastreamento das fixações do paciente usando um oftalmoscópio a laser de varredura fornece ao médico informações precisas sobre a localização e extensão dos pontos retiniais periféricos, bem como a relação espacial entre eles e a região de perda de campo visual. A partir disso, os pacientes podem ser incentivados a deslocar o pontos focais para um pontos retinal mais eficaz (BEDELL; STEVENSON, 2013).

Pacientes com dificuldades de leitura também são avaliados através da medição e registro de seus movimentos oculares, teste realizado por muitas clínicas em casos semelhantes. Embora uma relação causal entre as características do movimento ocular e a

leitura ineficiente seja controversa, intervenções como o treinamento visual foram relatadas por estudiosos para melhorar as características dos movimentos oculares de leitura aumentando a velocidade (DUCHOWSKI, 2017).

Embora a tecnologia necessária para avaliar a precisão da sacada, ganho de perseguição ou estabilidade da fixação não seja acessível para grande parte dos médicos, clínicas especializadas geralmente possuem dispositivos dedicados a testes específicos de movimentos oculares, como a vídeo-oculografia para avaliação da função vestibular ou rastreadores de limbo como o *VisaGraph* para registrar os movimentos dos olhos durante a leitura (QUAID; SIMPSON, 2013). A eletro-oculografia e a eletronistagmografia são amplamente utilizadas na avaliação de movimentos sacádicos e vestibulares (BEDELL; STEVENSON, 2013).

O desenvolvimento de Hardware e software de processamento de vídeo desenvolveram-se rapidamente nas últimas duas décadas e estão tornando os sistemas baseados em vídeo mais comuns em aplicações clínicas, à medida que seu desempenho aumenta e o custo diminui continuamente.

3.7 Mouses Oculares

A principal evolução que provocou a adoção em massa dos computadores pela grande população é considerada por muitos como o desenvolvimento da interface gráfica de usuário (GUI). Antes da existência da GUI toda a interação do usuário com um computador era realizada através de texto, digitando-se linhas de comando e referindo-se a objetos e programas através de seus nomes. A GUI permitiu que se passasse a apontar um objeto, uma tarefa certamente mais intuitiva e natural do que digitar seu nome. Quase todas as interações passaram a envolver o apontamento de um item, seja utilizando um mouse, *touchpad*, *trackball*, *joystick* ou mais recentemente as telas sensíveis ao toque (*touchscreens*).

Como menciona Drewes (2010) em seu trabalho, o apontamento é efetuado normalmente em conjunto com alguma outra modalidade de seleção, como o clique do mouse por exemplo. Aplicações voltadas à interação de pessoas com deficiências motoras que utilizam como única fonte de entrada a direção do olhar necessitam de algum mecanismo para evitar o problema do toque de Midas⁹ (*Midas Touch*). A grande maioria dessas interfaces utilizam a permanência do ponteiro em um item para ativar ações extras, método que utiliza o denominado *dwell time*, ou simplesmente duração do olhar (descrito na subseção 3.4.6).

Movimentos oculares são rápidos em comparação ao movimento de um mouse. São movimentos trêmulos, e não formam um traçado suave, devido aos trêmulos sacádicos

⁹ Dificuldade de interfaces oculares em decidir se o olhar em uma região só visa inspecionar ou executar um comando específico.

(WOBBROCK; MYERS; KEMBEL, 2003). O olho não é tão preciso quanto um mouse, haja visto que o tamanho da fóvea limita a acurácia do ponto observado a $0,5^\circ$, ainda sem considerar a precisão do rastreador ocular. Considerando-se todas as vantagens e inconveniências do rastreamento do olhar, os sistemas de apontamento através do olhar (chamados neste trabalho de mouses oculares) se mostram ainda limitados mesmo após décadas de desenvolvimento relacionado.

A tabela 3, adaptada de Drewes (2010), aponta outros atributos dos mecanismos de apontamento, além da velocidade e precisão, que devem ser considerados. Os atributos são espaço requerido, *feedback*, possibilidade de múltiplos apontadores e a forma de clique. Observa-se a comparação destes atributos para os métodos tradicionais de apontamento e o apontamento através do olhar.

Tabela 3 – Propriedades dos dispositivos de apontamento.

	Mouse	Trackball	Track Point	Touchpad	Touch screen	Olhar
Velocidade	Rápido	Rápido	Médio	Rápido	Rápido	Muito rápido
Precisão	Veloc. manuseio	Veloc. manuseio	Veloc. manuseio	Veloc. manuseio	Tamanho do dedo	Tamanho da fóvea
Espaço requerido	Muito	Pouco	Pouco	Pouco	Nenhum	Nenhum
Feedback	Sim	Sim	Sim	Sim	Não	Não (depende)
Tipo de Feedback	Indireto	Indireto	Indireto	Indireto	Direto	Direto
Múltip. Apontamento	2 Mãos	2 Mãos	2 Mãos	10 Dedos	10 Dedos	1 Par de olhos
Clique Associado	Não	Não	Não	Sim (não)	Sim	Não

Fonte: Drewes (2010, p. 74, tradução própria)

A Lei de Fitts (FITTS, 1954) estabelece uma relação entre a precisão e velocidade de dispositivos de apontamento, demonstrando que uma maior precisão de apontamento requer mais tempo na execução da atividade. Para os casos de uma tela sensível ao toque ou apontamento através do olhar, a precisão é determinada por características anatômicas - notadamente a largura da ponta dos dedos ou o tamanho da fóvea - e não pela Lei de Fitts.

O espaço requerido para se operar um dispositivo de apontamento varia consideravelmente, desde os mouses comuns - que necessitam de considerável espaço para deslocamento do dispositivo sobre uma superfície - a telas sensíveis ao toque que não requerem espaço extra além da própria tela. Alternativas como o *trackball*, *touchpad* ou similares utilizam menos espaço em comparação ao mouse convencional. Telas sensíveis ao toque possuem o inconveniente fato de o tamanho da ponta dos dedos diminuir a precisão de entrada e obstruir a visão de certas regiões na tela, problemas contornáveis através do uso de ca-

netas específicas. A situação é similar para o apontamento através do olhar, mas não há alternativas para se aumentar a acurácia deste método de entrada, e nota-se a vantagem de não obstrução de nenhuma parte da tela durante a interação.

Métodos de apontamento indireto, como o mouse, geralmente necessitam de *feedback*. Métodos diretos de entrada como telas sensíveis ao toque e mouses oculares não requerem *feedback*, apesar de ser recomendado para o último caso, onde erros de calibração e o efeito do escorregamento podem prejudicar a utilização do dispositivo.

Não é possível utilizar múltiplos apontamentos oculares simultâneos, uma vez que os olhos se movem sincronicamente e apontam para a mesma direção. Outros dispositivos podem possibilitar a utilização dessa característica, como a utilização de vários dedos em uma *touchscreen* ou *touchpad*.

A utilização do olhar para apontar itens em uma interface é a tarefa mais comum quando se trata de rastreamento ocular. Apesar de natural e intuitiva, a ação de apontar através dos olhos encontra desvantagens já mencionadas, como a baixa precisão dos movimentos oculares e rastreadores, a incapacidade dos olhos de se pressionar botões - salvo através das piscadas que acarretam fadiga visual - além da dupla função do olho para inspeção da tela e ativação de comandos - problema do *Midas Touch* - e das precauções na utilização do *feedback*.

Grande parte dos trabalhos relacionados ao rastreamento ocular compreendem o olhar como entrada de apontamento. Encontra-se, no entanto, um crescente interesse no desenvolvimento de outros tipos de interface e na tentativa de contornar o problema de baixa precisão do rastreamento do olhar.

Zhai e Kristensson (2003) sugerem em seu trabalho uma maneira de contornar o problema de baixa precisão, utilizando a combinação de apontamento através do olhar e uso de um mouse. Sua proposta foi chamada de MAGIC (*Mouse And Gaze Input Cascaded*).

Salvucci e Goldberg (2000) apresentaram uma abordagem diferente, buscando prever probabilisticamente a posição do olhar do usuário em uma interface controlada somente pela direção do olhar. O algoritmo desenvolvido representa o início das investigações que empregam inteligência artificial e linguagem natural para otimizar as interfaces e protocolos de comunicação orientados ao rastreamento ocular.

Várias pesquisas buscaram empregar objetos expansíveis nas interfaces gráficas, como Miniotas, Spakov e Evreinov (2003) que denominaram de expansão estática o aumento da área clicável dos itens na interface sem mostrar visualmente ao usuário e posteriormente desenvolveu a expansão dinâmica, onde os objetos na tela se expandiam com a passagem do cursor (MINIOTAS; ŠPAKOV; MACKENZIE, 2004).

Uma interface chamada *EyePoint*, desenvolvida por Kumar, Paepcke e Winograd (2007) também faz o emprego de itens expansíveis, combinados com uma tecla de ativação como método de entrada. Ao pressionar a tecla a região da tela visualizada se

amplia e o item é selecionado ao se soltar a tecla.

3.8 Teclados Oculares

Muitas aplicações já possuem suporte para entrada de texto utilizando a direção do olhar (KRISTENSSON, 2007; ZHAI; KRISTENSSON, 2012; CECOTTI, 2016; WARD; MACKAY, 2002). Entretanto, a taxa de entrada de texto em teclados visuais ainda é baixa se comparada a métodos tradicionais de escrita.

A escrita através da direção do olhar é realizada olhando-se para os caracteres desejados em uma interface de usuário. Tal processo requer um teclado virtual presente na interface e um dispositivo de rastreamento ocular como ilustrado na Figura 17. A seleção dos caracteres é efetuada fixando o olhar em determinada tecla durante um pequeno intervalo de tempo, chamado de '*dwell time*' ou simplesmente permanência.

Figura 17 – Teclado virtual e dispositivo de rastreamento ocular.



Fonte – Majaranta et al. (2006, p. 200)

Mesmo os sistemas mais avançados de digitação ocular ainda apresentam baixas taxas de entrada de texto, variando de 7 a 20 palavras por minuto - *words per minute* (WPM) (MAJARANTA; AHOLA; ŠPAKOV, 2009).

Para selecionar um item na interface apresentada na tela, o usuário precisa focar o olhar no item desejado, assim como o cursor deve ser deslocado pelo mouse até a área pretendida antes de fazer uma seleção. Muitos rastreadores oculares podem ser utilizados para emular um mouse convencional, substituído pela localização da posição do olhar.

Outro método amplamente utilizado por indivíduos com restrições de comunicação usuários de ferramentas de CAA é a varredura. A digitação por varredura permite ao

usuário utilizar um acionador para mudar o foco de um elemento a outro na interface, no método denominado varredura em etapas. Outro método, a varredura automática, é utilizada se o usuário possuir um único acionador.

No sistema de varredura automática, os itens na tela do computador são destacados sucessivamente através de um sinal visual ou através de um sinal sonoro. A ativação do item escolhido em um grupo de seleção, como letras, ícones ou figuras, é realizada através de um acionador. Por exemplo, se o grupo de seleção for o alfabeto em um teclado virtual e o usuário quiser escrever a palavra “casa”, ele pressiona o acionador para iniciar a varredura esperar até que o destaque passe pela letra “C” para pressionar o acionador novamente e repetir o processo para cada letra na palavra.

Portanto, usuários com deficiência física que apresentem pelo menos um movimento controlado tal como uma rotação do pescoço, uma batida com a mão ou o pé, um piscar de olhos ou qualquer contração muscular podem acessar o computador através do uso de um acionador específico, estrategicamente posicionado e do método de varredura.

3.8.1 OptiKey

O software OptiKey é um teclado assistivo, projetado para ser utilizado com um dispositivo de rastreamento ocular de baixo custo. Ele traz o controle do mouse, do teclado e, também, reproduz a fala, sendo indicado para pessoas com Esclerose Lateral Amiotrófica (ELA) / Doença dos Neurônios Motores (DNM).

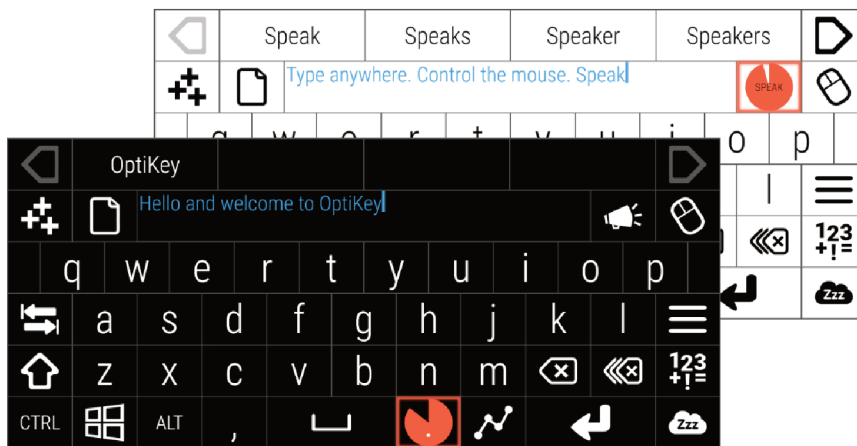
O OptiKey foi desenvolvido por Julius Sweetland como uma alternativa às soluções de CAA extremamente caras e inacessíveis em grande parte do globo terrestre. É um software de código aberto e gratuito (SWEETLAND, 2015).

O teclado assistivo OptiKey requer somente a instalação de um rastreador ocular e possibilita a seleção de letras através de fixações do olhar ou através de botões físicos ou acionadores. Caso o usuário não possua um rastreador ocular, pode utilizar a interface utilizando mouse ou uma câmera. A interface do teclado virtual optiKey está representada na Figura 18.

A interface pode ser utilizada para substituir um teclado físico. É capaz de inserir espaços entre as palavras automaticamente, fazer correções e sugestões de palavras, além da funcionalidade de auto-completar palavras.

O Optikey pode também substituir o mouse, permitindo clique, rolagem e arrasto com precisão em qualquer lugar da tela. Funções avançadas de navegação pela internet, pedido de ajuda e até mesmo fala em voz alta estão presentes nesta aplicação já bastante difundida, principalmente para utilização com o rastreador ocular Tobii 4c que é fácil de ser encontrado e possui custo acessível.

Figura 18 – Interface do teclado ocular OptiKey.



Fonte – Sweetland (2015)

3.9 Digitação ocular sem fixação (*dwell-free eye typing*)

A utilização de gestos é algo comum ao se tratar da área de estudo da IHC. O ser humano está habituado a executar gestos propositadamente ou involuntariamente como auxílio à comunicação, na maioria das vezes utilizando as mãos e cabeça. Exemplos são a escrita manuscrita e a língua de sinais (ou língua gestual).

Grande parte da pesquisa envolvendo gestos se refere a gestos efetuados utilizando-se um mouse ou caneta. Podemos citar as interfaces *Unistroke* (GOLDBERG; RICHARDSON, 1993), *Cirrin* (MANKOFF; ABOWD, 1998), *EdgeWrite* (WOBBROCK et al., 2007) e diversas aplicações para computadores e dispositivos móveis que utilizam gestos do mouse ou dos dedos como entrada. A aplicação mais comum de gestos é na produção textual (digitação) ou controle do computador através de comandos por gesto. Drewes (2010) atesta o seguinte a respeito da utilização de gestos em interfaces:

A utilização de gestos na interação com o computador não é intuitiva, uma vez que requer o aprendizado de um conjunto de gestos e sua semântica. Por essa razão, os gestos na interação homem-máquina têm sido vistos como algo para especialistas ou para propósitos muito específicos. Entretanto, com o surgimento do iPhone e produtos similares, que possuem a tela sensível ao toque como único método de entrada, o emprego de gestos feitos com os dedos se tornou popular. (DREWES, 2010, p. 101, tradução própria)

Um gesto compreende um conjunto de elementos contínuos realizados sequencialmente no tempo, como por exemplo uma sequência de traços ou segmentos de reta. O gesto - seja ele efetuado pela mão, caneta, ponteiro de um mouse ou pelo olhar - deve possuir ponto inicial e final. O gesto de uma caneta, por exemplo, é iniciado assim que a caneta

toca a superfície e termina quando ela é levantada. Ao se utilizar um mouse encontramos uma situação diferente, onde não há como tirar o ponteiro do mouse da tela e se utiliza uma tecla de gesto - geralmente o botão esquerdo do mouse - para indicar quando se está fazendo um gesto. A utilização da tecla garante que o algoritmo não vai detectar gestos a todo o tempo, mesmo quando o usuário está simplesmente navegando pela interface.

O mesmo ocorre com os gestos oculares, que não podem ser delimitados como um gesto efetuado por uma caneta. A utilização de uma tecla de ativação (como em um mouse) resolveria o problema, mas isso eliminaria a grande vantagem e motivação para se utilizar o rastreamento do olhar - uma interface que não requer o uso das mãos ou qualquer contato físico, possibilitando o controle remoto da interface. Dessa maneira, os gestos oculares são monitorados e registrados continuamente, criando a dificuldade de discernimento do algoritmo entre movimentos naturais dos olhos e gestos intencionais. É um desafio similar ao do reconhecimento de comandos de fala específicos em meio ao rastreamento contínuo de diálogos normais que não deveriam ativar comandos.

Há poucos trabalhos relacionados ao desenvolvimento e emprego de gestos realizados pelos olhos em IHC, sobretudo em língua portuguesa. Esta Tese possui foco ainda mais específico, interessando-se na utilização de gestos oculares na digitação através de teclados virtuais, tópico cuja bibliografia correlacionada é exígua.

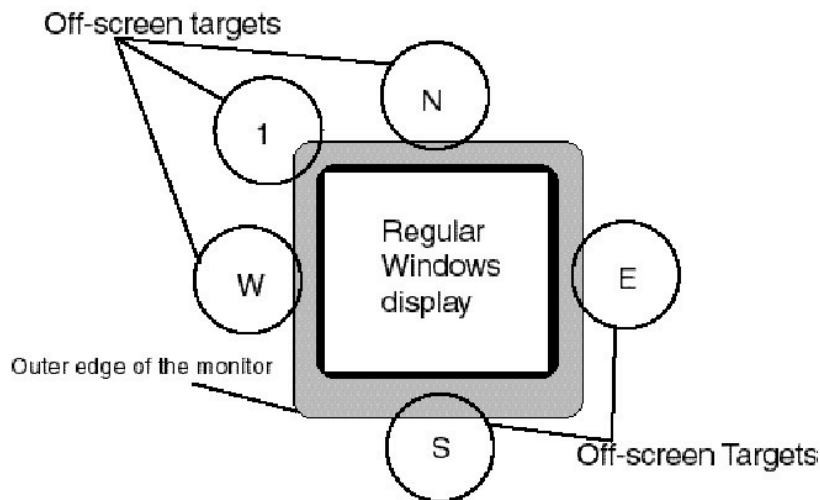
3.9.1 Bibliografia Relacionada

A utilização de gestos oculares é investigada em interfaces para entrada de texto, ou seja, digitação através do olhar. Como mencionado anteriormente, a digitação por pessoas com distúrbios neuromotores pode somente ser realizada pelos olhos em muitos casos, e a entrada de caracteres sucessivos fixando-se o olhar uma letra de cada vez se mostra um processo extenuante do ponto de vista físico e cognitivo para esses indivíduos. Uma alternativa explorada por alguns pesquisadores baseia-se na eliminação do período de fixação ou da piscada como método de seleção de caracteres individuais, em um método chamado de '*dwell-free eye typing*', ou digitação ocular sem permanência, em tradução livre. Adotaremos nesta Tese o termo 'digitação por gestos oculares'.

Uma das primeiras pesquisas a explorar gestos oculares foi apresentada por [Isokoski \(2000\)](#). Ele desenvolveu e efetuou testes em uma interface de entrada de texto utilizando objetos da interface fora da tela, como observado na Figura 19. O usuário deve olhar para alvos fora da tela em padrões específicos para inserir os caracteres de um texto. Os padrões de movimento efetuados pelos olhos são chamados de gestos visuais ([ISOKOSKI, 2000](#)).

O trabalho de [Qvarfordt e Zhai \(2005\)](#) descreve o desenvolvimento de um sistema de diálogo, para interação com o usuário através de padrões do olhar. Eles partiram de uma investigação relacionada ao movimento dos olhos em diálogos entre humanos e utilizaram os resultados para implementar uma interface de diálogo homem-máquina. Os usuários

Figura 19 – Localização de alvos fora da tela na interface proposta por Isokoski (2000).



Fonte – Isokoski (2000, on-line)

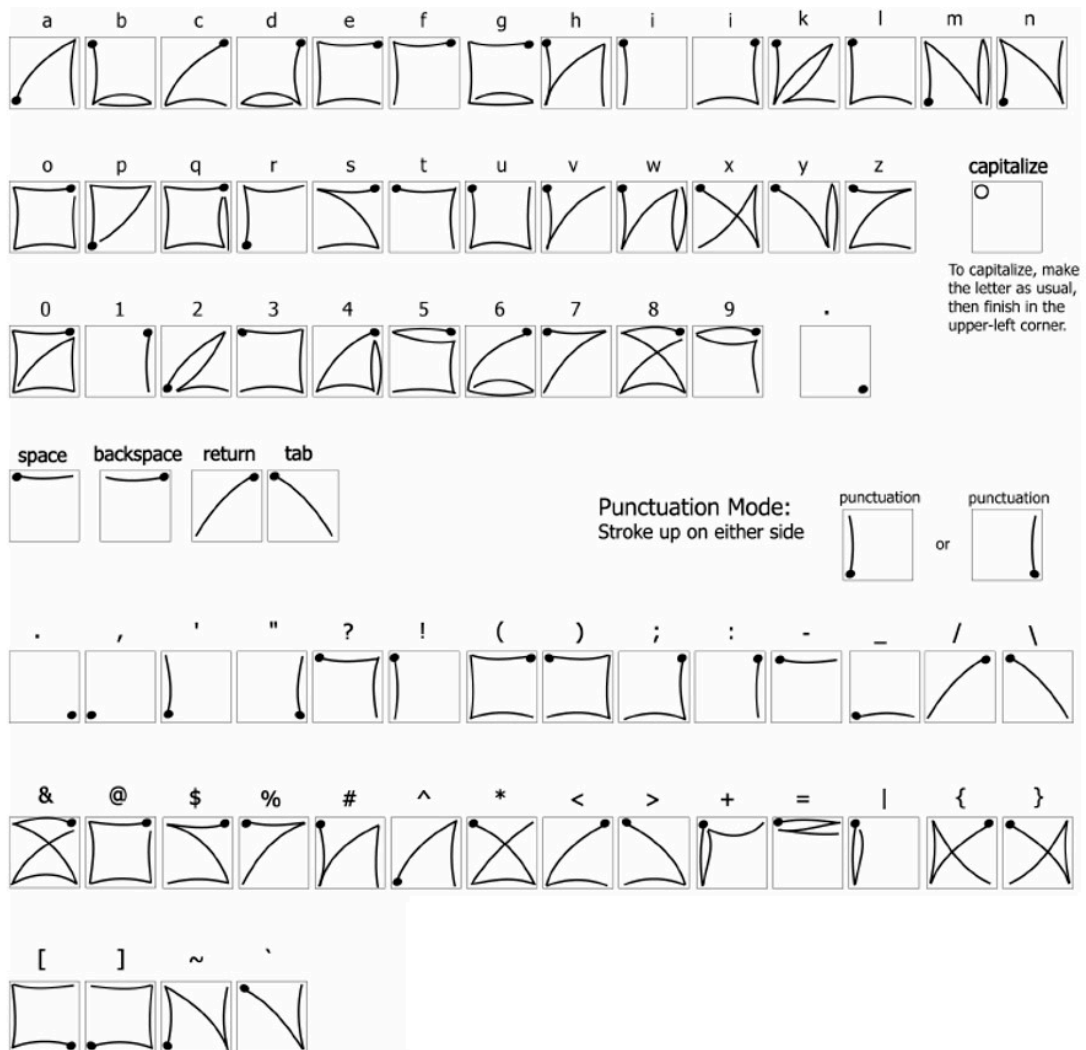
do sistema operavam a interface sem a necessidade de aprender um conjunto de gestos, e nem mesmo se deram conta de que estavam fazendo gestos oculares e os utilizando para controle da interface (QVARFORDT; ZHAI, 2005).

O termo gesto visual (chamado pelo autor de *gaze gesture*) foi apresentado em 2003 por Milekic (2003) em seu trabalho que não era estritamente científico, mas artístico. O autor desenvolveu um *framework* para o desenvolvimento de uma possível interface controlada através de gestos visuais para utilização em um museu (MILEKIC, 2003).

Wobbrock et al. (2007) publicaram em 2007 um trabalho no qual apresentavam a interface apelidada *EdgeWrite* para entrada de texto através de gestos visuais, como ilustrados na Figura 20. Os autores realizaram diversos testes que indicavam que a entrada de texto através de gestos seria mais lenta em relação à entrada de texto tradicional com os olhos, ou seja, utilizando piscadas ou intervalos de espera em cada caractere a ser inserido (método do *dwell time*). Usuários treinados na interface desenvolvida atingiram a velocidade de digitação de 7,99 WPM (palavras por minuto) (WOBBROCK et al., 2007).

No ano de 2008 foram publicados vários trabalhos relacionados a gestos visuais, tanto demonstrando sua aplicação com diferentes métodos de rastreamento ocular (BULLING; ROGGEN; TRÖSTER, 2008), apresentando algoritmos de reconhecimento dos gestos efetuados pelo usuário (BULLING; ROGGEN; TRÖSTER, 2008), e aplicando-os na entrada de texto (PORTA; TURINA, 2008; WOBBROCK; RUBINSTEIN, 2008; BEE; ANDRÉ, 2008).

Drewes (2010) efetua em sua tese uma minuciosa revisão de literatura com respeito ao rastreamento do olhar e sua utilização em IHC. Vários experimentos de usabilidade são realizados, com um algoritmo de rastreamento de gestos do olhar desenvolvido com

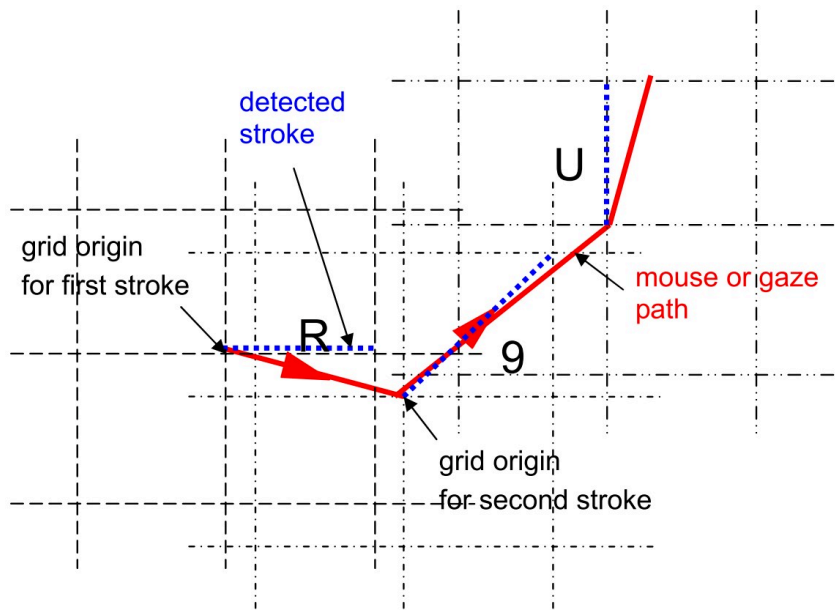
Figura 20 – Gestos definidos para os caracteres na interface *EdgeWrite*.

Fonte – Wobbrock et al. (2007, p. 63)

base nos gestos de mouse dos navegadores Firefox e Opera e na interface *EdgeWrite* (WOBBROCK; MYERS; KEMBEL, 2003).

Para detectar os gestos oculares, Drewes (2010) operou similarmente à detecção de gestos do mouse, onde o algoritmo recebe as coordenadas de posições x e y e calcula as distâncias percorridas entre ponto inicial e final. Se a distância calculada entre dois instantes de tempo for diferente de zero, o ponto final se torna o ponto inicial para o próximo gesto e o caractere a ser retornado é escolhido de acordo com as características do traçado, mas somente se o caractere for diferente do selecionado no último gesto (DREWES, 2010). A Figura 21 ilustra a entrada de caracteres através de gestos oculares pelo algoritmo de Drewes (2010). O discernimento ente movimentos naturais do olho e comandos intencionais em meio aos traçados contínuos gerados pelo olhar requer a introdução de outra variável no algoritmo: o tempo t .

Em uma publicação mais recente, Hoppe, Löchtefeld e Daiber (2013) sugerem em seu

Figura 21 – Decodificação do traçado do olhar ou de um mouse em uma *string*.

Fonte – Drewes (2010, p. 102)

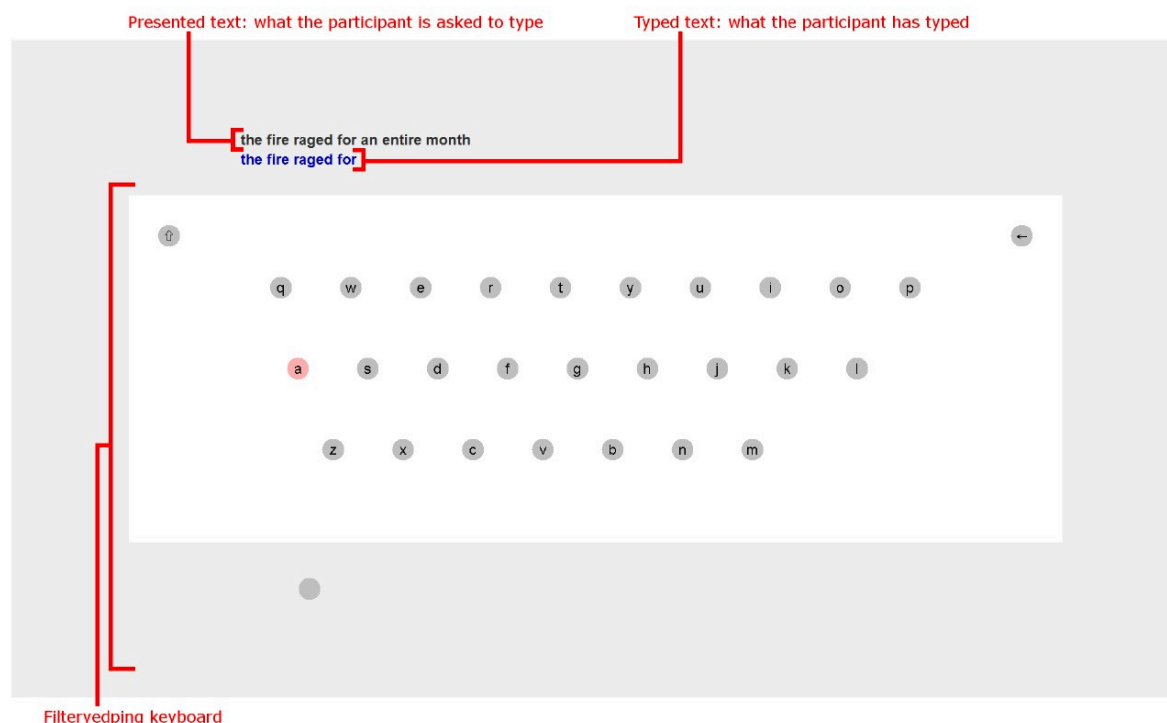
trabalho que uma técnica de digitação sem fixações poderia aumentar significativamente a velocidade de digitação em um sistema. Eles propuseram um método - apelidado de *Eype* - onde o usuário olha para a sequência de caracteres desejada em um teclado QWERTY e o traçado do olhar é capturado e comparado com o traçado ideal pré-calculado para todas as palavras em um dicionário (ou corpus) para identificar a sugestão mais provável (HOPPE; LÖCHTEFELD; DAIBER, 2013). Uma lógica semelhante é desenvolvida por Kristensson e Zhai (2004) na sua interface - denominada SHARK² - de teclado virtual baseado em gestos muito citada na maioria dos trabalhos da área, ainda que utilize gestos feitos por uma caneta *stylus* e não pelos olhos (KRISTENSSON; ZHAI, 2004).

Kristensson e Vertanen (2012) demonstraram que o método de digitação pelo olhar sem utilização de fixações poderia ser muito mais rápido que os métodos tradicionais, obtendo em seus experimentos uma taxa de 46 WPM que representaria o limite teórico de eficiência na digitação ocular. Seu sistema que simulava um decodificador perfeito para digitação através de gestos oculares - significando que seu software sabia de antemão o que o usuário desejava digitar (KRISTENSSON; VERTANEN, 2012).

Outro importante trabalho, desenvolvido no Brasil, foi realizado por Pedrosa, Pimentel e Truong e detalhado em múltiplos trabalhos (PEDROSA; PIMENTEL; TRUONG, 2015), (PEDROSA, 2015). A interface denominada *Filteryedping* e observada na Figura 22 é uma adaptação do SHARK² (KRISTENSSON; ZHAI, 2004) para utilizar rastreadores oculares e permitir a digitação de palavras através de gestos oculares. Este trabalho reconhece a palavra pretendida através de uma busca em uma lista de frequência contendo todas as palavras do dicionário que podem ser formadas, com a etapa adicional de comparação quando descartadas nenhuma ou algumas das letras olhadas pelo usuário (PEDROSA;

PIMENTEL; TRUONG, 2015).

Figura 22 – Filteryedping - Interface de teclado virtual controlado por gestos oculares.



Fonte – Pedrosa (2015, p. 79)

3.10 Considerações Finais

Este capítulo apresentou uma visão geral da utilização da direção do olhar e movimento dos olhos como fonte de informação e método de entrada, para fins de controle ou comunicação. Foram apresentadas características anatômicas e funcionais do olho humano, assim como a caracterização e modelagem matemática dos movimentos oculares, necessários para o desenvolvimento de dispositivos capazes de rastrear e registrar tais movimentos.

Posteriormente foram apontadas diversas aplicações que utilizam a direção do olhar como fonte de informação. Aplicações de diagnóstico, apontamento e entrada de texto foram detalhadas e exemplificadas, sendo o último tipo exatamente o foco de desenvolvimento deste trabalho. O referencial teórico apresentado neste capítulo é essencial para a compreensão e desenvolvimento da aplicação de digitação através dos olhos pretendida.

O próximo capítulo apresenta conceitos e técnicas de rastreamento do olhar para diversos fins, discorrendo sobre o desenvolvimento e estágio atual da tecnologia de rastreamento ocular. São listados sistemas de rastreamento disponíveis comercialmente e o rastreador utilizado no desenvolvimento deste trabalho é apresentado e especificado.

Rastreamento do Olhar

4.1 Introdução

Este capítulo começa com a definição de rastreamento ocular e o desenvolvimento da tecnologia relacionada através do tempo. São descritas as diferentes técnicas de rastreamento, assim como as informações obtidas por esses dispositivos e os campos de aplicação dos rastreadores oculares.

É apresentada uma listagem de alguns rastreadores oculares disponíveis comercialmente, e o modelo escolhido é detalhado e especificado (rastreador ocular Tobii 4c). A integração do rastreador do olhar escolhido com o teclado virtual desenvolvido e descrito nas seções anteriores é detalhada, incluindo-se a arquitetura completa, parâmetros, configurações e *drivers* utilizados.

4.2 Rastreamento do Olhar

O rastreamento ocular é um ramo vastamente explorado pela ciência e indústria, consistindo em uma tecnologia multidisciplinar que envolve a biologia, medicina, psicologia, neurologia e a engenharia. O conhecimento de áreas específicas relacionadas como anatomia e fisiologia do olho, movimentos oculares e percepção humana seria suficiente para preencher vários livros.

O termo rastreamento ocular (também denominado rastreamento do olhar) consiste na estimação da direção do olhar do indivíduo. Na maioria das aplicações, a estimação da direção do olhar significa a identificação do objeto cujo olhar se encontra. No caso de sua utilização em computadores, identifica-se as coordenadas x e y na tela para onde o usuário está direcionando o olhar (DREWES, 2010).

Os rastreadores oculares se diferenciam em relação à quantidade de graus de liberdade que podem rastrear. Rastreadores simples retornam somente a direção do olhar em relação à cabeça (Eletro-oculografia¹ e sistemas fixados na cabeça) ou a partir de uma posição

¹ Oculografia é o nome dado ao método usado para registrar a posição e os movimentos do olho.

fixa do globo ocular (como nos sistemas que requerem a fixação da cabeça). Os sistemas mais sofisticados possuem capacidades de rastreamento da cabeça, efetuando a estimação da posição do olhar mesmo com a cabeça do usuário em movimento.

Grande parte dos rastreadores oculares baseados em vídeo fornecem não só a direção do olhar, mas também o tamanho da pupila. A dilatação e contração pupilares são indicadores da atividade emocional relacionada ao que é observado pelo indivíduo e é objeto de inúmeras pesquisas. Este trabalho, no entanto, não utiliza métricas pupilares e não se interessa aos aspectos emocionais e cognitivos ligados à dinâmica ocular.

4.2.1 Técnicas de Rastreamento do Olhar

Os rastreadores do olhar modernos passaram por notável desenvolvimento, científico e tecnológico. Os indivíduos que fazem uso desses dispositivos são enormemente beneficiados, tendo em vista que os dispositivos aperfeiçoados são menos invasivos e requerem praticamente nenhum conhecimento técnico para sua operação. Apesar dos avanços significativos, são empregados ainda hoje princípios clássicos utilizados na detecção do ponto de observação de um indivíduo, descritos a seguir:

□ Lentes de contato especiais

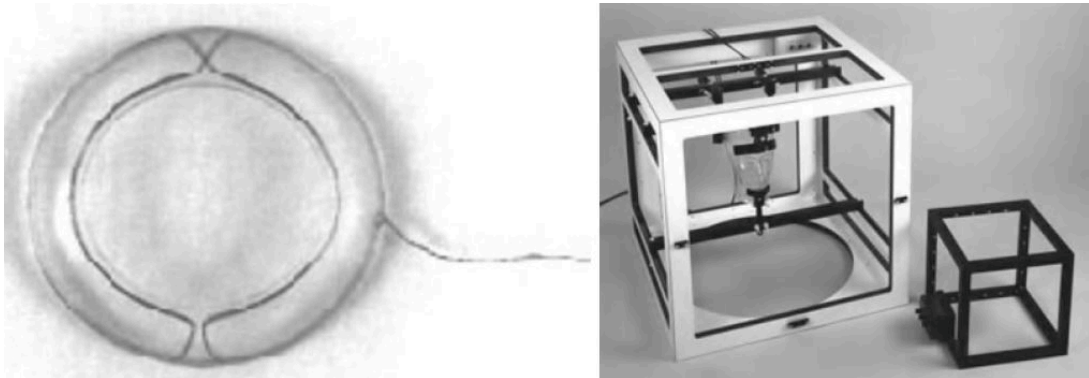
Um dos métodos mais antigos aplicados na detecção da direção do olhar faz uso de lentes de contato especiais contendo dispositivos mecânicos ou óticos. Apesar de bastante intrusivo, o método que utiliza lentes de contato modificadas é ainda hoje o mais preciso na determinação da direção do olhar de um observador.

[Duchowski \(2017\)](#) menciona vários modelos de lentes de contato para detecção do olhar, incluindo lentes com fósforo reflexivo e diagramas formados por linhas usados para detectar o olho em imagens, lentes com espirais de fio e bobinas que detectam variações no campo magnético decorrentes da movimentação ocular, e até lentes conectadas mecanicamente a dispositivos de registo em papel ou em disco, utilizadas nos primórdios do desenvolvimento da tecnologia de rastreamento ocular.

Dentre os rastreadores oculares baseados em lentes de contato, o mais utilizado emprega uma espira de fio anexada à lente. A técnica, proposta por Robinson em 1963 e denominada *Scleral Search Coil* utilizava o princípio da indução magnética de uma pequena bobina metálica instalada em uma lente flexível de silicone e conectada a dispositivos de medida através de um fio fino, como se observa na Figura 23. A lente é posicionada no olho de forma concêntrica com a íris, e um campo magnético alternado é produzido na região ao redor da cabeça do usuário, induzindo uma tensão na bobina que possibilita os cálculos de posição ocular ([GIANNOTTO, 2009](#)).

([GIANNOTTO, 2009](#))

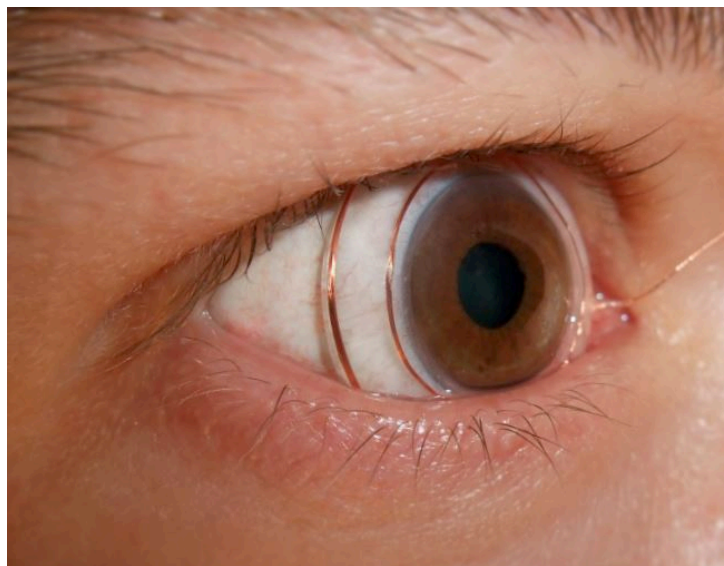
Figura 23 – Exemplo de bobina de rastreamento embutida em lente de contato e armações de campo eletromagnético para medição do movimento ocular.



Fonte – Duchowski (2017, p. 53)

O condutor que conecta a bobina aos instrumentos de medição traz alto desconforto para o usuário, além de requerer um processo complexo de instalação e remoção. A Figura 24 exibe uma lente posicionada no olho do usuário, onde pode-se observar o condutor que é conectado aos instrumentos de medição. A grande vantagem desse método é a alta precisão e a resolução temporal praticamente ilimitada. Isso torna o método adequado para estudo de tremores, escorregamentos, oscilações sacádicas e outros micromovimentos do olho, rápidos e difíceis de se detectar utilizando outras técnicas.

Figura 24 – Lente de rastreamento ocular instalada no olho com o fino condutor saindo ao lado do nariz.



Fonte – Chronos-Vision (2019, on-line)

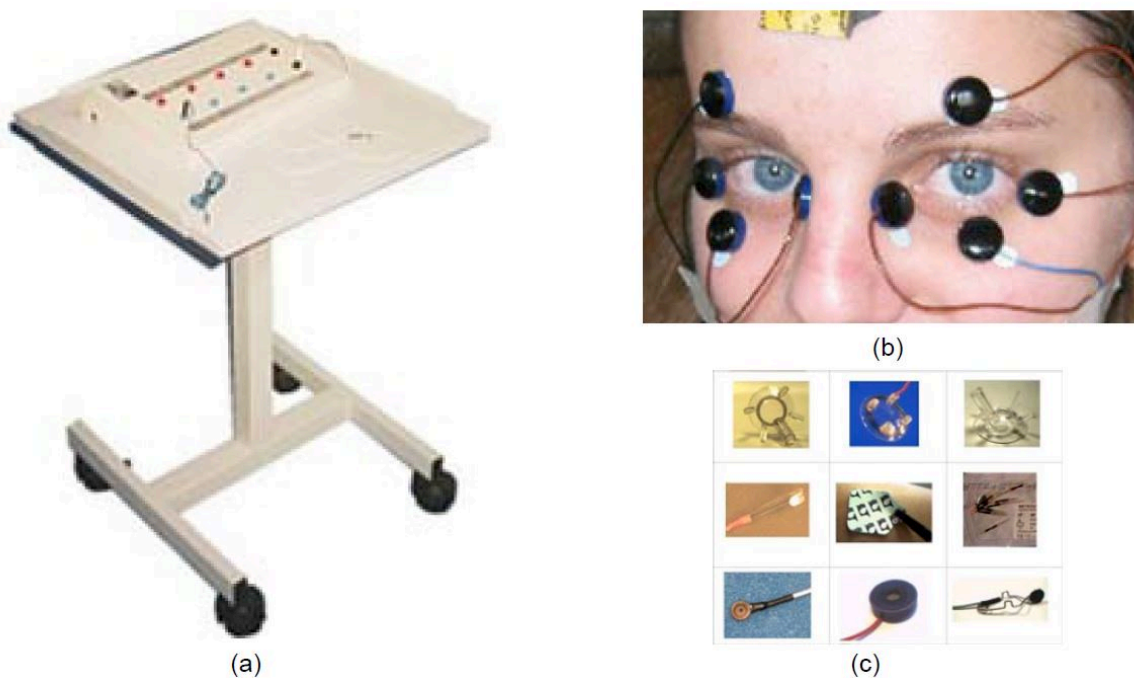
A resolução espacial desse tipo de solução é de cerca de $0,01^\circ$ e a resolução temporal tão rápida quanto se queira (o sinal de captura é analógico). A configuração é

lenta e requer calibração, tornando possível registrar apenas 30 minutos por sessão de gravação. Um sistema de rastreamento do olhar através de lentes com bobinas pode custar em torno de 15.000 dólares, com o custo adicional de 100 dólares por lente (HAIN, 2012). Isso limita a utilização dessa ferramenta para fins de pesquisa, sobretudo com animais.

□ Eletro-oculografia (EOG)

A EOG é uma técnica de rastreamento ocular bastante empregada antes do surgimento dos sistemas baseados em vídeo. Consegue-se, com um dispositivo desse tipo, efetuar o acompanhamento dos movimentos oculares através da diferença de potencial da ordem de 15 - 200 μ V produzida, captada por eletrodos dispostos ao redor dos olhos do indivíduo, como observa-se na Figura 26.

Figura 25 – (a) Unidade de gravação eletrofisiológica; (b) usuária de um rastreador ocular baseado em eletro-oculografia; (c) diversos tipos de eletrodo usados na EOG.



Fonte – Giannotto (2009, p. 37)

A variação da diferença de potencial medida na córnea e retina em relação á cabeça corresponde aos movimentos efetuados pelos olhos - convertidos em graus oculares. Vantagens desta técnica se encontram na simplicidade e disponibilidade dos dispositivos - unidades de gravação eletrofisiológicas possuem baixo custo e são semelhantes às utilizadas em exames comuns como eletrocardiograma (ECG) ou eletroencefalograma (EEG) (GIANNOTTO, 2009).

A EOG é menos invasiva que a utilização de lentes de contato, e não requer conhecimento aplicado de instaladores. Entretanto, continua sendo uma técnica invasiva, se

comparado aos métodos de vídeo-oculografia. Outra desvantagem é a incapacidade do método de efetuar o rastreamento da cabeça do usuário.

A resolução espacial para este método é de aproximadamente 1° (valor aceitável, sensibilidade em torno de $20\mu V/^\circ$), e a resolução temporal máxima 40Hz. O método de instalação e configuração é lento, requerendo cuidado, preparação da pele e calibração. O ruído é o maior problema neste caso, podendo ser da ordem de 1° ou mais, visto que é uma medição elétrica que pode sofrer influência muscular ou de ondas com frequências próximas a 60Hz. Dito isto, testes que considerem sacadas ou movimentos de perseguição não são aconselháveis ao se utilizar os métodos de EOG.

□ VOG

A vídeo-oculografia (VOG) e a foto-oculografia (POG) são ferramentas de rastreamento do olhar não-invasivas, não exigindo contato direto com os olhos ou face do usuário e provendo maior conforto e ergonomia na utilização. Este método utiliza câmeras de vídeo para efetuar o rastreamento ocular, sendo na maioria dos casos a técnica mais adequada para aplicações interativas.

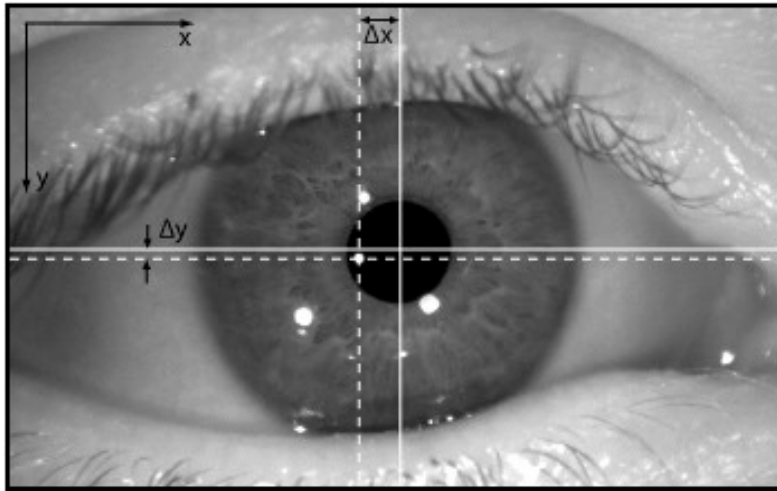
Imagens do usuário são registradas e usadas nos processos de detecção da posição dos olhos e de seus movimentos. Este método era inicialmente baseado em imagens estáticas de filmes fotográficos ou outros meios de captura da luz, e foi aperfeiçoado graças aos avanços na qualidade de imagem das câmeras, velocidade de processamento de vídeo e miniaturização de componentes eletrônicos (GIANNOTTO, 2009).

A VOG utiliza câmeras de vídeo convencionais, que operam na faixa visível do espectro eletromagnético². Características ou propriedades dos movimentos oculares são extraídas através do processamento das imagens capturadas pela câmera. Dentre as características oculares comumente utilizadas encontra-se a cor dos olhos, sua forma, e até sua posição. Como afirma Giannotto (2009), a grande maioria das implementações usando VOG utiliza a detecção do limbo ou a detecção da pupila.

Esta técnica é vantajosa em relação aos métodos anteriormente citados, mas apresenta desvantagens como a susceptibilidade a variações de iluminação, o que pode comprometer sua utilização por períodos prolongados. Esta falta de robustez pode dificultar a estimativa precisa da posição do olhar, tornando esse tipo de sistema menos preciso. Atualmente existem poucas soluções comerciais baseadas em VOG de espectro visível. A vídeo-oculografia foi aperfeiçoada combinando-se a detecção da pupila com a reflexão da córnea para que se possa obter o *point of regard* (POR) com maior precisão através de cálculos, como observa-se no exemplo da Figura 26.

² Os comprimentos de onda na faixa visível de radiação estão compreendidos entre $370\eta\text{m}$ (violeta) e $750\eta\text{m}$ (vermelho).

Figura 26 – Exemplo de imagem registrada por VOG exibindo o centro da pupila e as reflexões da fonte de luz IR adicional aplicada.



Fonte – Platz, Ong e Haslwanter (2009, on-line)

□ VOG baseada em reflexão corneana

A maioria dos sistemas de controle ocular disponíveis atualmente é baseada em vídeo (VOG), empregando a reflexão corneana³. Este método, desenvolvido no final da década de 60, efetua a detecção da pupila e de feixes de luz (geralmente infravermelha) refletidos nos olhos do usuário e a partir dessas informações calcula a posição do olhar estimada (Tobii, 2016).

Este método utiliza câmeras de vídeo convencionais assim como no método VOG convencional, mas inclui a faixa do espectro de luz infravermelha (não visível ao olho humano). São emitidos feixes de luz nessa faixa do espectro para produzir iluminação estável no rosto do usuário e aumentar o contraste entre pupila e íris (independentemente da cor dos olhos), facilitando sua localização. Essa fonte de luz não causa desconforto ou distração ao usuário, visto que não é visível pelo olho humano.

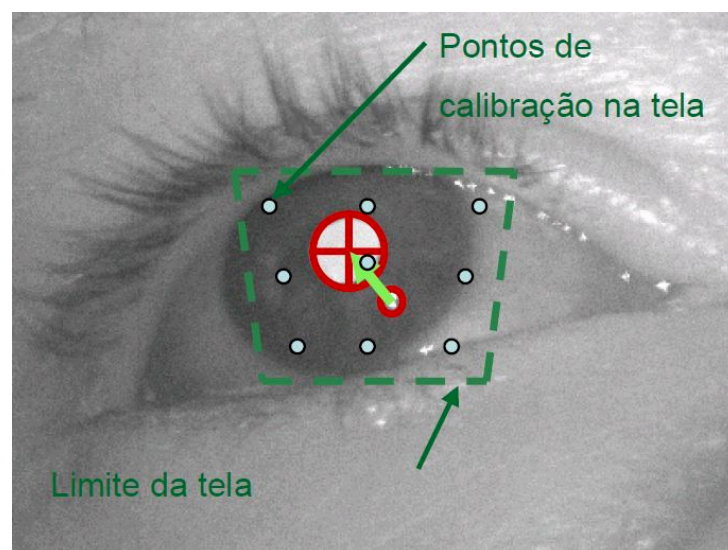
Na prática, pelo menos dois pontos de referência são necessários para o cálculo do ponto do olhar. Medindo as reflexões corneanas de uma fonte de luz artificial infravermelha direcionada no eixo do centro da pupila ou fora dele, o sistema consegue compensar imprecisões e movimentos (limitados) da cabeça. A direção do olhar é calculada neste método medindo-se a distância variável entre a pupila brilhante (se a luz é apontada no eixo) ou escura (luz direcionada para fora do eixo) e a reflexão (praticamente estática) da fonte de luz infravermelha na superfície da córnea (chamada de *glint*).

³ O termo inglês *corneal reflection*, traduzido por Giannotto (2009) como reflexão na córnea, será denominado neste trabalho como reflexão corneana.

O método opera iluminando-se a córnea do olho em ângulos determinados, produzindo reflexos em sua superfície para cada fonte de iluminação. A reflexão corneana apresenta posição praticamente constante durante o movimento dos olhos, permanecendo estática durante a rotação do olho e mudança na direção do olhar, servindo assim como uma referência básica da posição do olho e da cabeça. Essa reflexão também consiste em um ponto de referência que, comparado às várias posições da pupila em movimento, permite o cálculo do vetor de direção do olhar. Uma explicação mais detalhada é apresentada por Duchowski (2017).

Um processo de calibração é geralmente necessário para se estimar os parâmetros de mapeamento dos vetores entre o *glint* e a pupila para várias posições na tela. Quatro pontos de calibração nos cantos da tela resultam nos pontos nos cantos de um trapézio no plano mapeado pelos reflexos corneanos. Quatro pontos são suficientes para uma boa calibração, mas grande parte dos rastreadores oculares modernos utilizam mais pontos para obter melhor precisão. A Figura 27 ilustra o mapeamento resultante de uma calibração com nove pontos.

Figura 27 – Exemplo de vetor de posição existente entre o *glint* e centro da pupila e posições de nove pontos de calibração na VOG por reflexão corneana.



Fonte – Adaptado de Drewes (2010, p. 21)

O método da vídeo-oculografia utilizando reflexão corneana é o mais utilizado atualmente, e sem dúvidas é o mais eficiente levando-se em consideração os fatores custo, facilidade de instalação e operação, obstrutividade. Situações que podem ser problemáticas neste tipo de rastreamento incluem pessoas com deformações no globo ocular e pessoas portando lentes de contato (DREWES, 2010). Óculos são menos problemáticos pois, embora os óculos possam alterar a posição do reflexo, ele permanece na mesma posição sendo compensado pelo processo de calibração anteriormente descrito.

4.2.2 Critérios de Qualidade dos Rastreadores do Olhar

Os rastreadores oculares podem ser categorizados de acordo com suas características, independentemente da tecnologia empregada no seu funcionamento. Essas características são úteis durante o processo de comparação e escolha entre as soluções disponíveis para uma determinada aplicação, permitindo-se estimar o grau de confiança dos resultados esperados em experimentos ou estabelecer pré-requisitos de seleção. Segundo [Gianotto \(2009\)](#) e [Drewes \(2010\)](#) os rastreadores do olhar podem ser classificados de acordo com as características: forma de montagem, tipo de dados, resolução temporal, precisão, intrusividade, robustez, filtragem do sinal e existência de *application programming interface (API)s* ([GIANNOTTO, 2009](#); [DREWES, 2010](#)).

❑ Montagem

Os rastreadores oculares geralmente são montados de duas maneiras: na cabeça do usuário, onde não captura os movimentos e posições da cabeça, ou sobre a mesa ou suporte onde é possível o rastreamento dos olhos e da cabeça do usuário. Essa característica é determinante a depender do tipo de aplicação desejada e das condições de utilização do usuário.

❑ Tipo de Dados

Informações como posição dos olhos na face, posição da cabeça, ponto observado e diâmetro da pupila são geralmente fornecidas pela maioria dos rastreadores do olhar. Alguns sistemas mais avançados podem ser capazes de rastrear também a orientação da cabeça, frequência de piscadas e até mesmo o arqueamento das sobrancelhas. Os tipos de dados entregues pelos sistemas de rastreamento do olhar não são padronizados e podem diferir na unidade utilizada ou até mesmo no significado dos mesmos ([GIANNOTTO, 2009](#)).

❑ Resolução temporal e latência

A direção do olhar é registrada por um rastreador em certos intervalos de tempo, geralmente regulares. Para os rastreadores baseados em vídeo esse intervalo de tempo é determinado pela taxa de quadros da câmera de vídeo, mas essa taxa de quadros não corresponde necessariamente à resolução temporal. O rastreador ERICA descrito por [Drewes \(2010\)](#), por exemplo, retorna as coordenadas de posição do olhar a uma taxa de 60Hz mas não retorna nenhuma informação durante uma sacada. Em muitos casos, a latência - atraso entre a captura e a entrega dos dados ao sistema - é uma característica mais importante que a resolução temporal, especialmente no campo da [IHC](#) onde repostas do sistema em tempo real são cruciais.

❑ Precisão

A precisão indica o erro associado às medições efetuadas por um rastreador. Ela pode ser representada por meio da rotação da pupila em relação à face, sendo expressa em graus do campo visual, ou por meio da distância em pixels entre a posição capturada e a posição real observada pelo usuário na tela do computador⁴ (GIANNOTTO, 2009).

Como foi descrito na seção 3.3, há um limite anatômico para a precisão ocular de aproximadamente $0,5^\circ$. A maioria dos rastreadores do olhar atuais fornecem exatamente esse valor em suas especificações, valor esse que é suficiente para a maior parte das aplicações de interação homem-máquina. Observações que considerem microscacadas e microtremores requerem, no entanto, equipamentos que forneçam maior precisão. Alguns rastreadores apresentam melhores indicadores de precisão rastreando-se os dois olhos do usuário simultaneamente.

❑ Intrusividade

A intrusividade é a característica que busca representar o desconforto do usuário durante a utilização do rastreador. Como descreve Giannotto (2009, p. 53):

As técnicas invasivas, que exigem contato direto com o olho, e o estilo de montagem na cabeça aumentam bastante a intrusividade do equipamento, mas mesmo dispositivos baseados em câmeras e de montagem remota podem ser muito intrusivos quando restringem a movimentação da cabeça do usuário com o uso de *bite bars* e descansos de cabeça. (GIANNOTTO, 2009, p. 53)

❑ Robustez

Um rastreador do olhar deve funcionar de forma confiável em diferentes condições e com todos os indivíduos. Situações que podem prejudicar o rastreamento do olhar incluem condições de iluminação desfavoráveis ou casos de indivíduos portadores de lentes de contato ou óculos. Sistemas pouco robustos podem requerer a recalibração frequente, prejudicando a dinâmica dos experimentos e impossibilitando certas aplicações de interação.

❑ Filtragem do sinal

Os dados brutos de um rastreador ocular são ruidosos - os movimentos dos olhos não são suaves, efetuando microscacadas como descrito na seção 3.4 - e possuem interrupções durante piscadas, quando os olhos estão fechados. Um bom rastreador ocular deve, então, filtrar os ruídos e suavizar o sinal obtido. Conseqüentemente, certos modelos de rastreadores podem não rastrear pequenos movimentos dos olhos, sendo que alguns deles permitem o ajuste dos parâmetros de filtragem possibilitando o controle da suavização do sinal.

⁴ O método de representação da precisão em pixels não é indicado, pois sofre influência de fatores como distância entre o usuário e o rastreador, tamanho e resolução da tela.

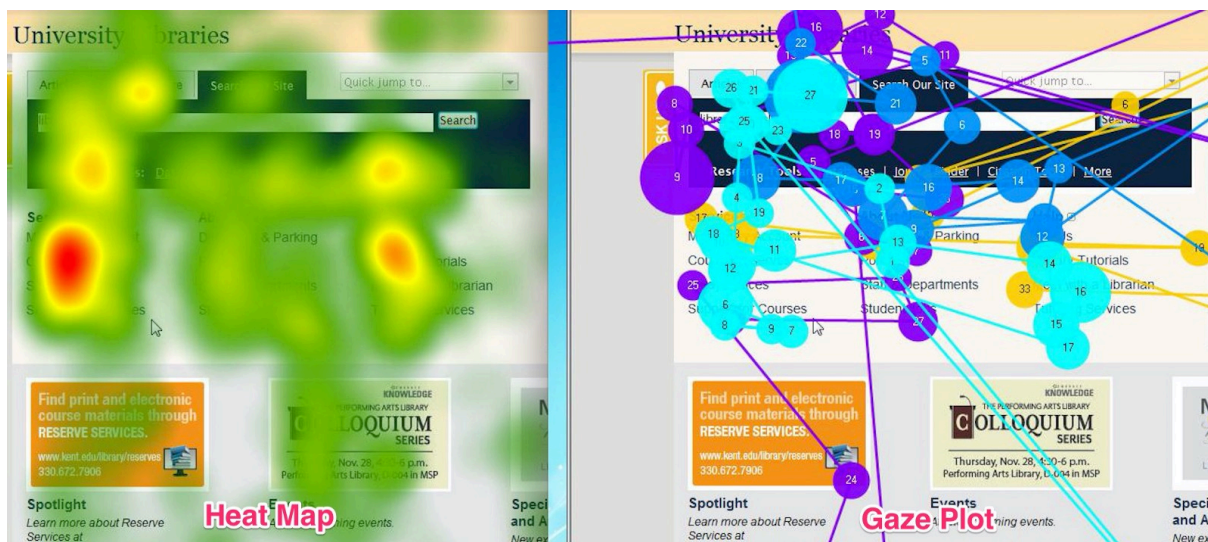
❑ Interface de programação de aplicações e visualização dos dados

A maioria dos rastreadores do olhar possuem aplicações específicas para visualização e avaliação da atividade ocular registrada. Geralmente são fornecidas também ferramentas voltadas ao desenvolvimento de aplicações, as chamadas **APIs**, que auxiliam na criação de *software* que utilize os recursos e dados retornados pelo rastreador.

4.2.3 Visualização dos Dados de Rastreamento Ocular

Os dados do rastreamento ocular são visualizados utilizando-se três métodos mais comuns, que procuram facilitar a análise das informações: *Gaze plots*, *gaze replays* e mapas de calor. A Figura 28 ilustra um *gaze plot*, também chamado de *scan path*, que consiste em uma imagem estática representando uma sequência de pontos indicando as fixações e linhas indicando sacadas. O tamanho dos pontos representa a duração de cada fixação e os números indicam a ordem em que foram executadas. Pode-se representar em um *gaze plot* a atividade ocular de um único indivíduo ou de vários participantes, situação onde são utilizadas cores diferentes para representar cada um deles (RÖSLER, 2012).

Figura 28 – Visualização das informações de rastreamento ocular em uma página web utilizando *heat map* e *gaze plot*.



Fonte – Tang (2016, on-line)

É possível também registrar um vídeo em tempo real da sessão de rastreamento ocular em um *gaze replay*. Esse tipo de representação pode ser reproduzido a baixas velocidades para possibilitar análises detalhadas dos registros. Isso pode ser necessário, uma vez que os movimentos oculares são muito rápidos e difíceis de acompanhar em tempo real. Um *gaze replay* utiliza a mesma representação de sacadas e fixações que os *gaze plots*.

O método mais usado e mais conhecido de visualização da informação de rastreamento ocular é o mapa de calor, ou *heat map*. Assim como os *gaze plots*, os mapas de calor

também são representações estáticas onde diferentes cores são usadas para representar o número ou a duração de fixações que cada parte da imagem recebe, como pode-se observar na Figura 28. Cores quentes (vermelho, amarelo) representam áreas observadas por mais tempo ou que atraíram mais fixações, e da mesma forma cores frias (verde) ou nenhuma cor indicam regiões menos observadas. Os mapas de calor são utilizados para representar a atividade ocular de vários indivíduos em um intervalo de tempo. Esse tipo de representação permite aos pesquisadores observar e analisar o comportamento de visualização de diversos sujeitos de uma observação, suavizando discrepâncias dos comportamentos individuais (RÖSLER, 2012).

4.3 Sistemas de Rastreamento Ocular Disponíveis Comercialmente

A maioria dos sistemas de controle ocular disponíveis atualmente são baseados em vídeo (VOG) por reflexão da córnea. Poucos dentre esses dispositivos são voltados à utilização por pessoas com deficiência. Apesar dos rastreadores oculares apresentarem características técnicas e princípios de operação similares, características como confiabilidade, robustez, segurança e condições de montagem podem tornar muitos dos dispositivos inadequados para utilização em TA.

O *software* de um rastreador ocular e o suporte técnico fornecido pelo fabricante são também elementos determinantes na utilidade e efetividade de um rastreador ocular para pessoas cujo controle ocular pode ser a única forma de se comunicar e interagir com o ambiente, em uma utilização que pode ser extensiva, diária e sujeita a condições diversas (MAJARANTA; AHOLA; ŠPAKOV, 2009). A tabela 4, adaptada de Majaranta, Ahola e Špakov (2009, p. 10), lista dispositivos de comunicação ocular disponíveis comercialmente e voltados a pessoas com diferentes tipos de necessidades especiais. Esses sistemas são utilizados primariamente como dispositivos assistivos - rastreadores oculares de uso geral não estão presentes nesta tabela.

Uma lista extensiva de rastreadores oculares atualmente disponíveis, categorizados em sistemas para tecnologia assistiva, fins de pesquisa, análise e outros está disponível na página web da associação *Communication by Gaze Interaction* (COGAIN) (COGAIN, 2019).

4.4 Tobii 4C

O rastreador ocular Tobii 4c[®] é um dispositivo de baixo custo destinado especificamente à experiência com jogos digitais. Os dados fornecidos pelo dispositivo fornecem coordenadas do olhar e os eventos de fixação. Os dados das fixações (incluindo as sacadas

Tabela 4 – Sistemas de controle ocular (baseados em VOG) disponíveis comercialmente.

Fabricante	Dispositivo	Página web
Alea Technologies	Intelligaze IG-30	< www.alea-technologies.com >
DynaVox Technologies	EyeMax	< www.dynavoxtech.com >
Eye Response Technologies	ERICA	< www.eyeresponse.com >
LC Technologies	Eyegaze	< www.eyegaze.com >
Tobii Technology	MyTobii	< www.tobii.com >
EyeTech Digital Systems	EyeTech TM3	< www.eyetechds.com >
Metrovision	VISIOWBOARD	< www.metrovision.fr >
H.K. EyeCan	VisionKey	< www.eyecan.ca >

Fonte – Adaptada de [Majaranta, Ahola e Špakov \(2009, p. 10\)](#)

que são movimentos entre uma fixação e outra) são largamente utilizados em diversos estudos relacionados aos movimentos oculares ([PILZER; LIU, 2017](#)).

O rastreamento ocular é efetuado iluminando-se o rosto do usuário com feixes de luz na faixa do espectro do infravermelho e calculando-se os padrões de reflexão nos olhos do usuário. O Tobii 4c[®], ilustrado na Figura 29 foi projetado para uso em telas com até 27 polegadas, localizada a uma distância entre 50 e 95cm do rosto do usuário.

Figura 29 – Rastreador ocular Tobii 4c.



Fonte – [Tobii \(2016, on-line\)](#)

O rastreador ocular escolhido para realização desta pesquisa é voltado à utilização em jogos eletrônicos, e não para pesquisas científicas ou investigações do gênero. No entanto, seu baixo custo e especificações adequadas o levaram a ser utilizados em diversos trabalhos acadêmicos.

O modelo escolhido efetua o rastreamento dos olhos e da cabeça do usuário⁵, proporcionando maior flexibilidade e precisão dos dados auferidos. As coordenadas x e y correspondentes à direção do olhar do usuário são retornadas ao sistema (normalizadas em relação ao tamanho da tela, variáveis do tipo *float* no intervalo $[0,1]$ a uma taxa de 90Hz, assim como um marcador de tempo do tipo inteiro fornecido pelo *Tobii Stream Engine*. Deve-se levar em conta, todavia, que a precisão de um sistema de rastreamento ocular é menor que a de um mouse, como salienta o fabricante:

⁵ São registradas as coordenadas de posição (x , y e z) e de rotação (arfagem, guinada e rolagem) da cabeça do usuário.

[...] pode parecer muito como obter uma coordenada de ponteiro de mouse, mas, por favor, não cometa o erro de usar este ponto diretamente como você usaria uma coordenada de ponteiro de mouse. Os dados de rastreamento ocular não são tão precisos quanto os dados do ponteiro do mouse - na verdade, é fisicamente impossível devido a como nossos olhos e visão funcionam. Em vez disso, pense em uma série de *GazePoints* como representando uma área em que o usuário está olhando e leve em conta que a acurácia e precisão dos dados variam de usuário para usuário. (Tobii, 2016)

A Tabela 5 contém um exemplo de amostras de coordenadas do olhar obtidas pelo rastreador Tobii 4c.

4.5 Considerações Finais

Este capítulo apresentou fundamentos do rastreamento do olhar, aspectos práticos e critérios de qualidade das técnicas de detecção da posição do olhar. Foram descritos sistemas de rastreamento ocular disponíveis comercialmente, e especificado detalhadamente o rastreador Tobii 4c.

O próximo capítulo apresenta características de um teclado virtual, como seus elementos, métodos de desenvolvimento, diferentes layouts existentes, métricas utilizadas na sua avaliação. São considerados ainda um tipo especial de teclados virtuais, que utilizam como entrada traçados contínuos efetuados pelos dedos do usuário, por canetas específicas ou mesmo pelo olhar - denominados teclados gestuais.

Tabela 5 – Exemplo de amostras de coordenadas do olhar obtidas pelo rastreador Tobii 4c.

tstamp	mouse_x	mouse_y
1429904386907	0.129612	0.473684
1429904387004	0.151372	0.476132
1429904387102	0.228004	0.494492
1429904387180	0.310312	0.505508
1429904387276	0.430464	0.517748
1429904387341	0.486282	0.52142
1429904387434	0.536424	0.52142
1429904387498	0.571429	0.518972
1429904387582	0.593188	0.517748
1429904387680	0.545885	0.512852
1429904387771	0.419111	0.495716
1429904387868	0.2772	0.47858
1429904387962	0.174078	0.463892
1429904388042	0.137181	0.462668
1429904388129	0.130558	0.471236
1429904388208	0.167455	0.4847
1429904388298	0.251656	0.494492
1429904388414	0.377483	0.492044
1429904388513	0.501419	0.462668
1429904388603	0.592242	0.427173
1429904388668	0.633869	0.407589
1429904388766	0.670766	0.378213
1429904388845	0.687796	0.361077
1429904388941	0.701041	0.351285
1429904389041	0.705771	0.364749
1429904389131	0.689688	0.402693
1429904389245	0.664144	0.450428
1429904389338	0.651845	0.506732
1429904389435	0.644276	0.553244
1429904389532	0.636708	0.581395
1429904389644	0.631031	0.596083
1429904389742	0.61211	0.593635
1429904389840	0.523179	0.571603
1429904389930	0.386944	0.547124
1429904389997	0.307474	0.534884
1429904390079	0.252602	0.523868
1429904390170	0.213813	0.50918
1429904390268	0.204352	0.50306
1429904390366	0.227058	0.501836
1429904390475	0.270577	0.477356
1429904390541	0.293283	0.44186
1429904390636	0.347209	0.390453
1429904390730	0.40492	0.332925
1429904390828	0.46263	0.287638
1429904390912	0.503311	0.277846

Teclados Virtuais

5.1 Introdução

Este capítulo descreve as principais características de um teclado virtual. São apresentados os elementos essenciais em uma interface de entrada de texto e os requisitos de desenvolvimento. Em seguida são descritas métricas de avaliação do desempenho de teclados virtuais e parâmetros de determinação da distribuição das teclas na interface do teclado, incluindo *layouts* clássicos e uma proposta de *layout* otimizado para a digitação por gestos em português brasileiro.

A compreensão acerca dos teclados virtuais e seu desenvolvimento é essencial como referencial teórico antes que se possa partir para a etapa de desenvolvimento propriamente dito. O sistema proposto neste trabalho possui diversos módulos que devem ser devidamente compreendidos para que se possa definir de efetivamente os requisitos de software e de desenvolvimento necessários.

5.2 Características de um Teclado Virtual

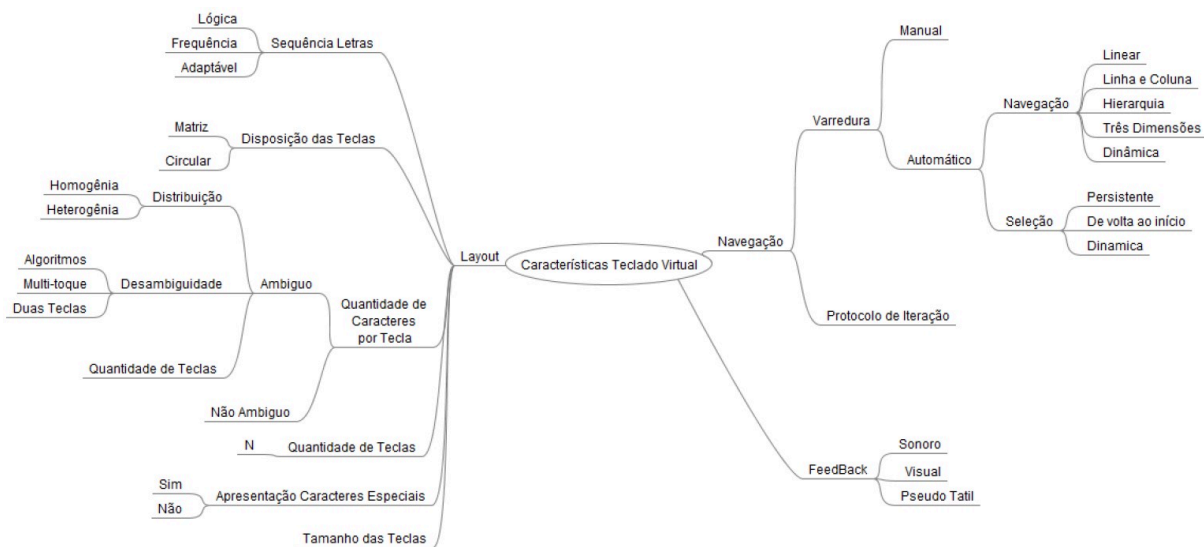
O desenvolvimento de uma aplicação de teclado virtual requer prévia compreensão e planejamento de seus requisitos e de suas características essenciais. Loja (2015) lista uma série de características inerentes a todos os teclados virtuais:

- ❑ Posicionamento das teclas;
- ❑ tamanho das teclas;
- ❑ quantidade de teclas;
- ❑ sequência das letras distribuídas entre as teclas;
- ❑ quantidade de caracteres por tecla;
- ❑ tipo de navegação.

- ❑ *feedback* do teclado;
- ❑ apresentação dos caracteres especiais;

O autor apresenta ainda um diagrama das características de teclados virtuais e valores que cada característica pode assumir, identificadas durante uma revisão sistemática de literatura, como disposto na Figura 30.

Figura 30 – Mapa mental das características de um teclado virtual.



Fonte – Loja (2015, p. 42)

As características dos teclados virtuais serão descritas a seguir.

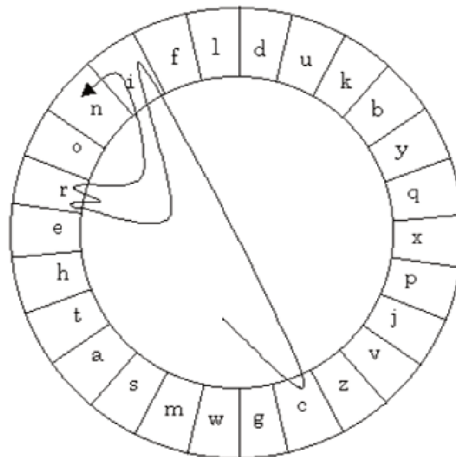
5.2.1 Posicionamento das teclas

Todo teclado virtual apresenta um posicionamento visual das teclas definido, geralmente organizado de duas maneiras principais: em forma matricial ou circular. A maior parte dos teclados virtuais desenvolvidos apresenta layout matricial (JULIATO et al., 2014; LIEGEL; GOGOLA; NOHAMA, 2008; BI et al., 2013), mas os teclados circulares são desenvolvidos e avaliados em investigações que buscam avaliar o posicionamento ótimo das teclas (LOJA, 2015).

Os teclados dispostos matricialmente se assemelham os teclados mecânicos tradicionais, e por isso apresentam uma curva de aprendizado menor. Podem apresentar diversas configurações de teclas diferentes, apesar da distribuição em linhas e colunas.

Os teclados circulares apresentam configuração diferente da usual mas apresentam um posicionamento das teclas mais próximas umas das outras. A Figura 31 apresenta um teclado em formato circular, denominado Cirrin (SARCAR et al., 2010).

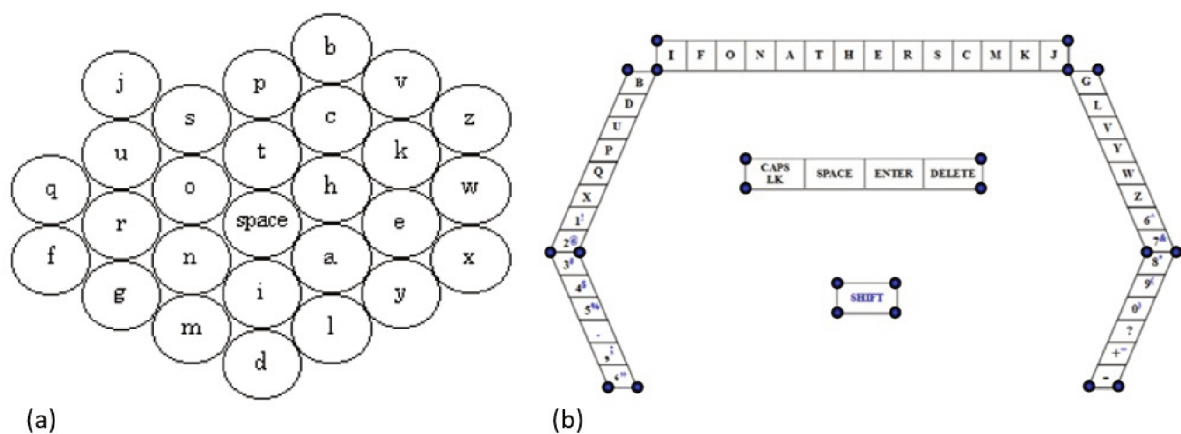
Figura 31 – Representação de um teclado virtual em formato circular.



Fonte – Sarcar et al. (2010, p. 292)

Há ainda posicionamentos híbridos como o teclado Hooke que apresenta formato aproximado de uma colmeia e o teclado apresentado por Ahsan et al. (2014) contendo caracteres alfanuméricos e especiais, como apresentado na Figura 32.

Figura 32 – Representação de teclados virtuais com posicionamento alternativo, descritos por (a) Sarcar et al. (2010) e (b) Ahsan et al. (2014).



5.2.2 Tamanho das teclas

Além da disposição das teclas na interface de um teclado virtual, o tamanho das teclas é um fator determinante em sua usabilidade. Teclados podem apresentar todas as teclas com mesmo tamanho, teclas com tamanhos diferentes ou até mesmo teclas com tamanho dinâmico.

O tamanho das teclas apresenta relevância ainda maior no presente contexto de digitação ocular, onde se procura utilizar teclas o maior possível para facilitar a fixação com os olhos e separação entre cada tecla de forma a minimizar as imprecisões inerentes do

apontamento utilizando o olhar. O tamanho das teclas está diretamente relacionado à Lei de Fitts (FITTS, 1954) como salienta Drewes (2010).

5.2.3 Distribuição das letras

A distribuição das letras na interface de um teclado virtual se refere à sua ordem sobre o posicionamento definido. As principais técnicas de distribuição dos caracteres ocorrem de acordo com sua frequência na linguagem utilizada ou de maneira lógica (LOJA, 2015), além da distribuição poder ser determinada dinamicamente, na distribuição chamada adaptativa (BHATTACHARYA; LAHA, 2013).

A distribuição das letras baseada na frequência utiliza a probabilidade de utilização de cada letra para posicionar os caracteres. Os caracteres mais utilizados geralmente são posicionados ao centro da interface. Os arranjos de letras organizados logicamente se fundamentam no idioma utilizado, e podem variar enormemente segundo uma lógica própria adotada pelo desenvolvedor. A distribuição adaptativa, por sua vez, altera a sequência de letras de acordo com o comportamento do usuário. A ordem das letras é, portanto, dinâmica (BHATTACHARYA; LAHA, 2013).

A seção 5.4 descreve em detalhes os principais padrões de distribuição de teclas utilizados em teclados mecânicos ou virtuais.

5.2.4 Quantidade de caracteres por tecla

Os teclados virtuais possibilitam a utilização de mais de um caractere por tecla, característica por vezes desejada a depender de sua finalidade. Estes são chamados teclados ambíguos, estes classificados de acordo com sua distribuição, homogênea ou heterogênea. Loja (2015) esclarece que "distribuições homogêneas possuem a mesma quantidade de caracteres por tecla, enquanto os arranjos heterogêneos apresentam uma quantidade diferenciada de caracteres por tecla"(LOJA, 2015, p. 45).

5.2.5 *Feedback* do teclado

O *feedback* de um teclado virtual representa a maneira como a interface informa ao usuário que uma tecla ou palavra foi inserida. Em um teclado mecânico o *feedback* é tátil, visto que o usuário pode sentir as teclas pressionadas pelos dedos. Teclados virtuais, no entanto, não são capazes de fornecer retorno tátil, se utilizando de alertas visuais, sonoros ou hápticos.

Retornos visuais podem ocorrer na forma de destaques nas teclas selecionadas ou percorridas, mudanças de cor na interface, ampliação de teclas, realce do traçado no caso de teclados gestuais, entre outros. O *feedback* sonoro pode se dar na forma da emissão de algum som no momento de seleção de letra ou palavra, ou mesmo retorno em fala

para teclados com funcionalidades de *text-to-speech*. O *feedback* háptico, ou pseudo-tátil (LOJA, 2015) apresenta ao usuário sensações sentidas pelo tato, como pequenas vibrações por exemplo.

5.2.6 Navegação

A navegação determina os métodos de interação do indivíduo com o teclado virtual. Um teclado mecânico convencional apresenta uma maneira única de entrada de caracteres - pressionando-se as teclas. Teclados virtuais, por sua vez, podem assumir diversos métodos de navegação, definidos pelo protocolo de interação ou sistema de varredura.

Indivíduos com necessidades especiais podem necessitar de métodos de navegação especiais, dado que em alguns casos eles são capazes de emitir apenas um tipo de sinal - ativo ou inativo. Qualquer diminuição no esforço de digitação possibilitado por métodos alternativos de navegação pode ter grande impacto na capacidade de comunicação de indivíduos com restrições comunicativas.

□ Protocolo de interação

O protocolo de interação define comandos específicos de teclado para determinados estímulos. Um sistema de digitação que utiliza o piscar dos olhos, por exemplo, poderia ser codificado para classificar diferentes padrões de piscada. O teclado desenvolvido nesta tese apresenta protocolo de interação baseado em gestos de palavra efetuados sobre o teclado virtual pela posição do olhar.

Protocolos de interação para sistemas de CAA são definidos levando-se em consideração as capacidades do utilizador, e podem se utilizar de movimentos oculares, impulsos musculares, acionadores mecânicos e até mesmo impulsos nervosos como forma de entrada de informação.

□ Sistema de varredura

Um método de navegação possível é o sistema de varredura, realizado de forma automática letra-a-letra para que o usuário apenas confirme a seleção de elementos no instante certo. Loja (2015) identifica cinco métodos de varredura, com diferentes padrões de movimentação do foco: linear, linha e coluna, três dimensões, hierarquia e dinâmico.

Características configuráveis do método de interação por varredura incluem ainda a ação após a seleção de uma determinada tecla e o tempo de parada em cada elemento.

5.2.7 Caracteres especiais

Os teclados virtuais podem ou não apresentar caracteres especiais como pontuação, navegação e formatação. Teclados assistivos voltados à comunicação básica podem não conter caracteres especiais, assim como teclados que buscam aumento da performance de digitação em detrimento da entrada de caracteres acentuados ou textos mais complexos.

5.3 Métricas de Avaliação de Teclados Virtuais

A pesquisa e desenvolvimento relacionada à entrada de texto requer uma compreensão de seus aspectos objetivos e também subjetivos. Ao se desenvolver novos métodos de entrada de texto, deve-se avaliá-los de forma a mensurar sua eficiência e validade, para que se possa determinar sua aplicabilidade em situações reais. Nesta seção são descritas métricas de avaliação do desempenho de ferramentas de entrada de texto, visto que algumas delas serão aplicadas na avaliação do método proposto neste trabalho.

Uma característica crucial de um dispositivo de digitação é a sua velocidade de entrada de dados. Como apontado por Loja (2015), se um método é mais lento que o outro já existente ele dificilmente conquistará aceitação de usuários. A precisão também representa grande importância na utilidade de uma ferramenta de digitação, onde uma maior precisão resulta em menores taxas de erros de digitação. Wobbrock et al. (2007) apontam que a velocidade de digitação e taxa de erros são inversamente proporcionais. Deve-se buscar, portanto, um equilíbrio entre essas duas métricas.

O esforço de digitação pode ser estimado de forma objetiva através de métricas como teclas por caractere - *keystrokes per character* (KSPC) - e gestos por caractere (GPC), e de forma subjetiva como a escala Nasa TLX que avalia o esforço cognitivo necessário ao processo de digitação.

5.3.1 Velocidade de digitação

A maioria dos trabalhos relacionados ao desenvolvimento de métodos de entrada de texto utiliza a medida de velocidade de digitação de palavras por minuto (PPM) ou WPM, onde cada palavra é considerada como 5 caracteres, incluindo espaços. A medida de WPM é obtida dividindo-se o número de palavras digitadas pela quantidade de minutos gastos desde o início do primeiro gesto até a entrada da última palavra desejada (PEDROSA, 2015).

Loja (2015) apresenta a equação 9 de cálculo da velocidade de digitação, onde T é o número de caracteres transcritos e S é o tempo em segundos decorridos entre a entrada do primeiro caractere até a inserção do último termo do texto. A subtração de uma unidade no denominador se deve à contagem de tempo somente a partir da entrada do primeiro caractere ou gesto.

$$WPM = \frac{|T| - 1}{S} \cdot 60 \cdot \frac{1}{5} \quad (9)$$

A medição da velocidade de digitação pode ser feita de duas maneiras diferentes, onde o usuário reescreve um texto previamente definido ou um texto idealizado por ele próprio. A primeira maneira de aferição da velocidade de digitação é mais confiável, segundo [Wobbrock et al. \(2007\)](#), visto que não exige uma carga cognitiva extra ao indivíduo que deve pensar em qual frase escolher e torna a amostragem mais uniforme.

O processo de medição da velocidade de digitação começa, então, pela definição do texto a ser digitado pelo usuário. É imperativo que o texto utilizado contenha o máximo possível de letras do alfabeto adotado, para que o usuário percorra o maior número de teclas do teclado virtual. Os testes de velocidade presentes na bibliografia relacionada frequentemente utilizam panagramas, frases que contém todas as letras do alfabeto utilizando o mínimo de palavras possível ([LOJA, 2015](#)).

Como descreve [Loja \(2015\)](#), um panagrama bastante utilizado na língua inglesa é "*the quick brown fox jumps over the lazy dog*". O autor cita ainda um panagrama em português e demonstra, como pode ser visto na [Figura 33](#), o cálculo da velocidade de digitação para essa frase. Se o usuário digita os 46 caracteres da frase em 50 segundos por exemplo, a velocidade de digitação calculada será 10,8 palavras por minuto, segundo a equação 9.

Figura 33 – Exemplo de cálculo da velocidade de digitação.

$$\begin{array}{ccc}
 S = 0 & & S = 50 \\
 \text{segundos} & & \text{segundos} \\
 \downarrow & & \downarrow \\
 \text{Um pequeno jabuti xereta viu dez cegonhas felizes} & & \\
 \uparrow & & \uparrow \\
 T = 0 & & T = 46
 \end{array}$$

$$WPM = \frac{|46| - 1}{50} \times 60 \times \frac{1}{5}$$

Fonte – [Loja \(2015, p.73\)](#)

Em algumas situações pode ser preferível utilizar a medida de caracteres por Minuto (CPM), onde simplesmente suprime-se a divisão da quantidade de caracteres por cinco na equação 9. As métricas WPM e CPM não consideram os erros cometidos pelo usuário durante a digitação ([WOBBROCK et al., 2007](#)), não refletindo a precisão do teclado e sua eficiência geral. Sistemas com alta velocidade de digitação e alta taxa de erros, por exemplo, resultam em um teclado com baixo rendimento e poucas chances de utilização em situações reais. Dessa forma, há outras métricas que levam em consideração os erros inseridos pelo usuário durante a entrada de texto, como teclas por caractere e gestos por caractere ([seção 5.3.2](#)).

5.3.2 Teclas por caractere

A medida teclas por caractere, ou **KSPC** representa simplesmente a razão entre a quantidade de caracteres inseridos pelo usuário e a quantidade de caracteres no texto final transcrito. Considera a quantidade de erros cometidos durante a entrada de texto, incluindo a utilização da tecla de apagar e outras teclas utilizadas para que se possa concluir a digitação de uma sequência de caracteres. A fórmula 10 representa o cálculo do **KSPC**, como descrito por Loja (2015).

$$KSPC = \frac{|IS| - 1}{T} \quad (10)$$

onde IS é o total de teclas pressionadas e T a quantidade de caracteres transcritos.

A medida IS inclui, portanto, os caracteres não impressos como backspace e *delete*. Loja (2015) demonstra como exibido na Figura 34 a utilização da métrica teclas por caractere na digitação da mesma frase empregada anteriormente, simulando a digitação do panograma em 50 segundos e efetuando correções (das letras em vermelho) representados pelo caractere " « »".

Figura 34 – Exemplo de cálculo da quantidade de teclas por caractere.

$$KSPC = \frac{|54| - 1}{50}$$

Fonte – Loja (2015, p.75)

Ao se calcular o valor do **KSPC** para esse exemplo, obtém-se o valor de 1,08. Quanto mais próximo de 1 é o valor, significa que a digitação está mais eficiente, resultando em poucos erros na entrada de caracteres. Erros na digitação produzem valores de **KSPC** maiores que um, ao se pressionar teclas de correção e substituição.

5.3.3 Taxa de erros

Grande parte dos teclados virtuais existentes na atualidade apresentam mecanismos de correção de palavras, os quais necessitam de métricas de exatidão como a taxa de erro de palavra. O cálculo de velocidade de digitação é intuitivo, enquanto o cálculo da taxa de erros pode não ser óbvio. Há dois tipos de erros: erros não corrigidos que permanecem na sequência de caracteres transcrita e erros corrigidos com caracteres substituídos durante a digitação.

A taxa de erro de palavra é baseada na métrica de erros não corrigidos denominada distância mínima de palavra, ou *Minimum string distance (MSD)*, que representa o menor

número de modificações necessárias (inserções, omissões e substituições) para transformar uma frase transcrita na frase desejada (sem erros) (MACKENZIE; SOUKOREFF, 2002).

A taxa de erros de palavra não corrigidos pode ser definida, então, pela equação 11 onde S é a frase inserida e P a frase ideal a ser inserida pelo usuário (BI; ZHAI, 2016). Essa medida representa a diferença, portanto, entre a frase pretendida e a frase efetivamente digitada pelo usuário.

$$\text{Taxa de erros} = \frac{MSD(S, P)}{\max(|S|, |P|)} \cdot 100\% \quad (11)$$

Observe que esta fórmula não diz quais caracteres estão errados, apenas o número de erros. Para erros corrigidos utiliza-se a medida KSPC descrita anteriormente.

5.3.4 Avaliação subjetiva

A avaliação subjetiva de métodos de entrada de texto está relacionada à sensação de conforto e satisfação de um usuário ao digitar e pode buscar a mensuração da carga cognitiva (ou *workload*) decorrente da utilização de determinado método de entrada.

Apesar da dificuldade de se mensurar e comparar resultados subjetivos de diferentes métodos de entrada e diferentes condições de teste, o instrumento NASA TLX (índice de carga de tarefa) é utilizado por trabalhos relacionados para aferir de forma multidimensional a carga cognitiva decorrente do processo de digitação.

Pedrosa (2015) emprega em seu trabalho o NASA TLX para comparar a carga cognitiva durante a utilização dos teclados virtuais desenvolvidos sob diferentes condições como, por exemplo, empregando ou não métodos de *feedback* ao usuário durante a digitação.

5.4 Layout de Teclados

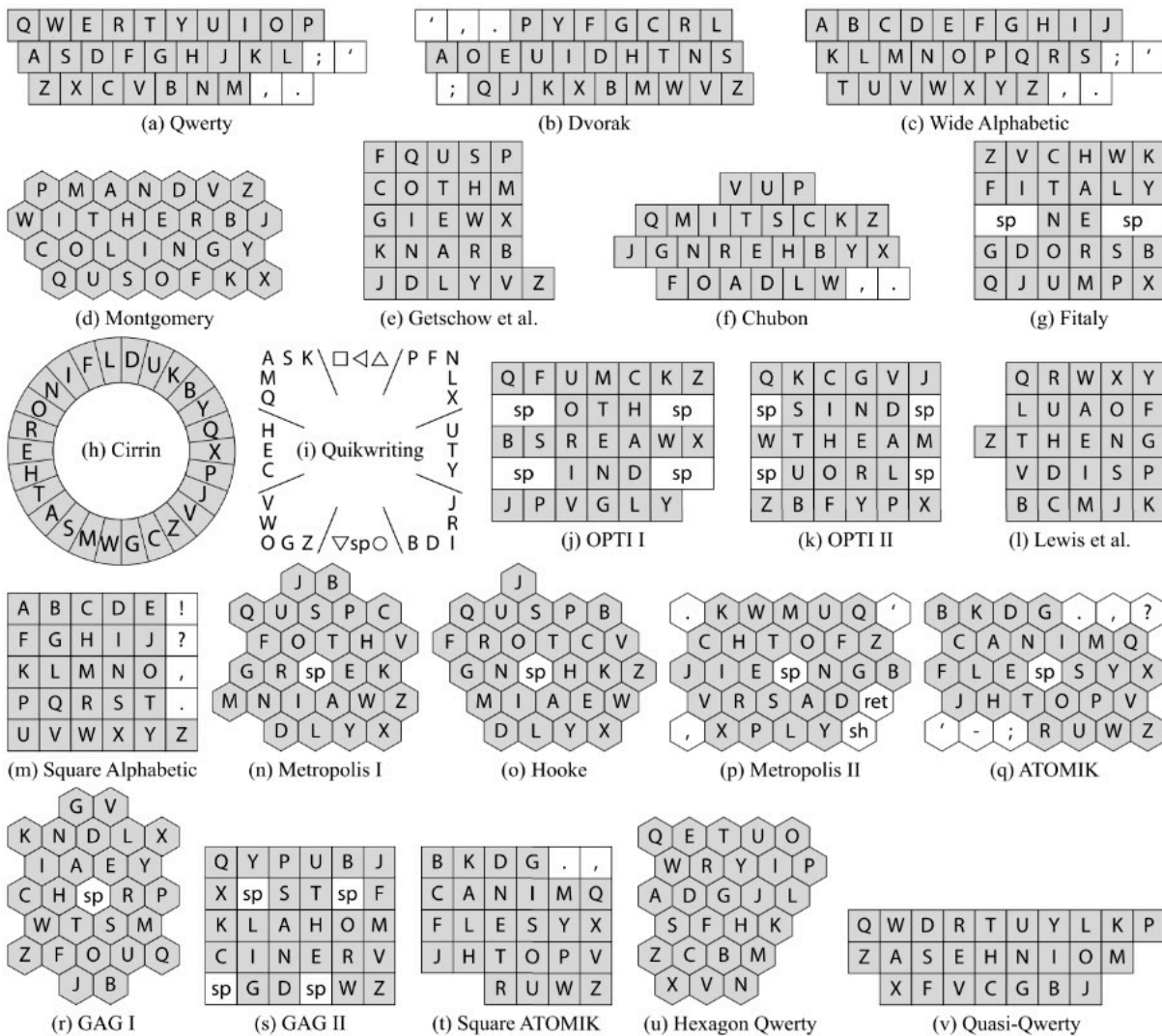
Os teclados surgiram com a invenção da máquina de escrever, dispositivo inicialmente rudimentar cujo mecanismo era complexo. Devido às características mecânicas das máquinas de escrever, não era raro ocorrerem travamentos e colisões entre teclas, sobretudo quando digitadas em alta velocidade. Desde então, diversas configurações de disposição das teclas foram desenvolvidas para tentar corrigir problemas mecânicos ou otimizar a experiência de digitação (RICK, 2010). A Figura 35 representa diversas configurações de teclados existentes.

Alguns dos principais arranjos de teclas, desenvolvidos para aplicação em teclados mecânicos ou virtuais, são descritos a seguir.

5.4.1 QWERTY

O teclado QWERTY (Figura 35a) foi idealizado por Christopher L. Sholes em 1867 para utilização em máquinas de escrever mecânicas, inicialmente adotando uma configu-

Figura 35 – Layouts de teclado.



Fonte – Rick (2010, p. 4)

ração de teclas em ordem alfabética (RICK, 2010). Os problemas de travamento entre as teclas levaram à modificação do layout de forma que as teclas de dígrafos (pares de letras consecutivas) frequentes fossem dispostas em lados opostos do teclado, evitando as colisões (YANG; MALI, 2016). O layout QWERTY tem sido desde então o padrão de utilização mundial, adotado em larga escala tanto em teclados mecânicos quanto virtuais, apesar de suas ineficiências apontadas por diversos autores (YANG; MALI, 2016; WADDINGTON et al., 2017; BI; ZHAI, 2016).

O layout QWERTY buscou sanar problemas de máquinas de escrever da época mas se mostra uma configuração ineficiente para teclados virtuais, sobretudo em dispositivos móveis, apresentando baixas velocidades de digitação. Diversos pesquisadores têm explorado maneiras de se obter melhores configurações de teclas, utilizando diferentes requisitos e estratégias (BI; ZHAI, 2016).

5.4.2 Dvorak

A mais bem sucedida tentativa de encontrar um layout de teclado melhor que o QWERTY foi realizada por Dvorak e Dealey em 1936 (RICK, 2010). A disposição alternativa das teclas buscava atribuir mais letras à mão direita que à esquerda, distribuir os caracteres de acordo a força de cada dedo das mãos, localizar as letras mais frequentes na linha central e colocar vogais e consoantes mais usadas onde sejam digitadas alternadamente, por cada mão.

O teclado Dvorak é a alternativa mais conhecida ao QWERTY, possuindo forma similar a este de forma que para transformar um teclado no outro basta renomear as teclas. A ideia central do layout Dvorak é a obtenção de maiores velocidades de digitação através da entrada de dígrafos frequentes por mão opostas sendo, portanto, um teclado destinado à digitação com duas mãos (SARCAR et al., 2010).

5.4.3 Fitaly

O teclado Fitaly é um produto comercial desenvolvido pela *Textware Solutions* destinado a dispositivos de entrada que utilizam um único ponto de contato, como caneta ou dedo (SMITH; BI; ZHAI, 2015). Procurou minimizar o movimento da mão ao digitar, dispondo duas teclas de espaço ladeadas pelas letras mais utilizadas no idioma inglês.

Em um teste conduzido pela própria desenvolvedora do teclado foram coletadas velocidades de digitação de 34 indivíduos utilizando diferentes disposições de teclado como QWERTY, Graffiti e o próprio Fitaly. A maior velocidade, de 44,4 palavras por minuto, aferida foi obtida utilizando-se o teclado Fitaly. Zhai e Hunter (2002) demonstram que a velocidade média obtida é de aproximadamente 36 palavras por minuto, mais eficiente que o layout QWERTY.

5.4.4 OPTI

Outra configuração de teclado desenvolvida afim de otimizar a digitação na língua inglesa é a configuração OPTI, descrita por MacKenzie e Zhang (1999) e ilustrada na Figura 35(j). O layout otimizado foi desenvolvido utilizando-se o método de tentativa e erro, Lei de Fitts (FITTS, 1954) e frequência de dígrafos no inglês (SARCAR; PANWAR; CHAKRABORTY, 2013).

A Lei de Fitts permite calcular o tempo de digitação de tecla em função do comprimento do trajeto e das dimensões das teclas desejadas, permitindo o cálculo teórico da velocidade de digitação para determinada configuração de teclas. Através do método de tentativa e erro, pode-se gerar e avaliar o desempenho de inúmeras disposições de teclas.

Segundo os pesquisadores, o layout OPTI é em teoria 25% mais rápido que o QWERTY após algumas horas de prática e 5% mais rápido que o teclado Fitaly (MACKENZIE; ZHANG, 1999). Estudos longitudinais efetuados pelos autores sugerem que a melhoria

na velocidade de digitação é constatada em situações reais. As duas versões observadas, OPTI I (Fig. 35(j)) e OPTI II (Fig. 35(k)) foram desenvolvidas pelo mesmo método, baseado em heurística.

5.4.5 Cirrin

Pode-se observar na Figura 35(h) o design radial do teclado Cirrin, concebido para digitação com canetas em superfícies sensíveis ao toque. A digitação é efetuada movendo-se a caneta da partir da região interior da circunferência até cada caractere que se deseje digitar. A entrada é registrada a partir das coordenadas do ponto onde a caneta cruza a circunferência interior.

Testes realizados com o layout Cirrin indicam que a precisão obtida é suficiente para entrada de texto, ainda que a velocidade de digitação possa ser um 'gargalo' dessa configuração de teclas, segundo [Sarcar, Panwar e Chakraborty \(2013\)](#).

5.4.6 Lewis

[Mackenzie, Zhang e Soukoreff \(1999\)](#) MacKenzie, Zhang e Soukoreff projetaram um teclado virtual em que as letras são colocadas alfabeticamente em duas colunas, o que não apresentou vantagem de desempenho, provavelmente devido a sua forma alongada. [Lewis, Kennedy e LaLomia \(1999\)](#) propuseram um outro teclado virtual otimizado com 5 linhas e 6 colunas - observado na Figura 35(l) - denominado pela bibliografia como teclado Lewis, além de uma alternativa com uma sequência estritamente alfabética - ilustrado na Figura 35(m) - que sofre do problema de descontinuidade alfabética causada pelas quebras de linha ([RICK, 2010](#)).

5.4.7 Hooke

O objetivo de um bom design de teclado visa minimizar estatisticamente a distância para percorrer entre os caracteres. Os dígrafos mais frequentes devem estar mais próximos entre si do que os menos frequentes, o que levou ao desenvolvimento de um sistema dinâmico para atingir um design otimizado.

[Sarcar, Panwar e Chakraborty \(2013\)](#) descreve o método de otimização utilizando um exemplo, onde há uma mola elástica conectando cada par das 27 teclas cujas posições iniciais foram dispostas aleatoriamente em um teclado, com espaços entre elas. A elasticidade das molas, quando ativada, era proporcional à probabilidade de transição entre os dois caracteres, de modo que teclas com maior probabilidade de transição fossem atraídas com maior força. O estado estacionário em que todas as teclas são pressionadas forma uma configuração candidata de teclado virtual ([SARCAR; PANWAR; CHAKRABORTY, 2013](#)).

No final de cada ciclo de simulação, o comprimento dos segmentos ajustáveis é reduzido a zero, de modo que todas as teclas são pressionadas uma contra a outra, formando o layout de um teclado virtual. O desempenho do design é então calculado e comparado com resultados conhecidos. Quando não satisfatório, o layout pode ser esticado para servir como outro estado inicial para a próxima iteração do mesmo processo. A iteração é repetida até que um layout satisfatório seja obtido. A Figura 35(o) mostra o melhor layout obtido com essa abordagem. Para capturar a essência da técnica de simulação por molas, a melhor disposição de teclas obtida é denominada teclado de Hooke (segundo a Lei de Hooke) (ZHAI; HUNTER; SMITH, 2000).

5.5 Teclados Gestuais

Os teclados gestuais são ferramentas de entrada de texto através de gestos efetuados pelo usuário utilizando canetas, os dedos, os olhos ou qualquer dispositivo de apontamento. Há diversas denominações para esses teclados, teclados *swipe*, teclados de escrita através de formas, ou ainda teclados por gestos de palavras.

Há diversos modelos de teclados gestuais disponíveis comercialmente, como o ShapeWriter, SlideIT, Swype, Flex T9, TouchPal, além dos teclados nativos dos sistemas operacionais Android e Windows (BI et al., 2013).

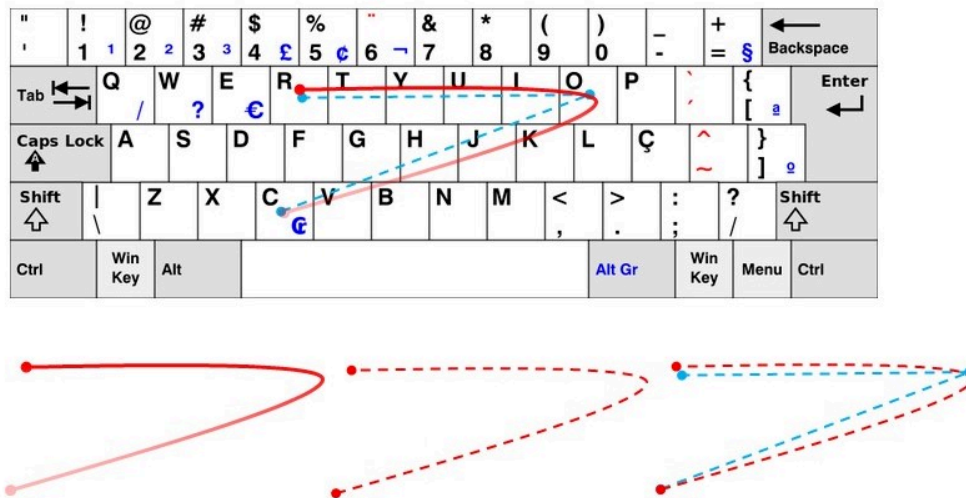
Nenhum dos teclados gestuais disponíveis atualmente apresenta, entretanto, o algoritmo utilizado para a detecção e decodificação de gestos de palavras. Não há, além disso, algoritmos de decodificação de formas disponível de forma livre para utilização em pesquisas ou investigações relacionadas.

A maior parte dos teclados gestuais desenvolvidos apresenta o layout **QWERTY**, amplamente conhecido e de rápido aprendizado. Outros layouts são adotados, como o **Dvorak**, **Fitaly**, **OPTI** e **ATOMIK**, buscando maior eficiência de digitação.

5.6 Decodificação de Gestos do Olhar em Texto

A decodificação do traçado do olhar é semelhante à decodificação de gestos feitos por canetas ou pelos dedos em interfaces sensíveis ao toque. Para esse tipo de entrada por gestos, a trajetória efetuada é amostrada em intervalos regulares de tempo (50 milissegundos, por exemplo), gerando dados que contêm as coordenadas e instantes de tempo de amostragem. A função do decodificador consiste, então, em encontrar a palavra no léxico existente com melhor correspondência à sequência de pontos de entrada (correspondente a um gesto, ou a um traçado do olhar). A Figura 36 exemplifica o processo de reconhecimento de gesto de palavra através da comparação com *template* de palavras do dicionário.

Figura 36 – Exemplo do processo de decodificação de traçado através da comparação com *template*.



Fonte – O autor

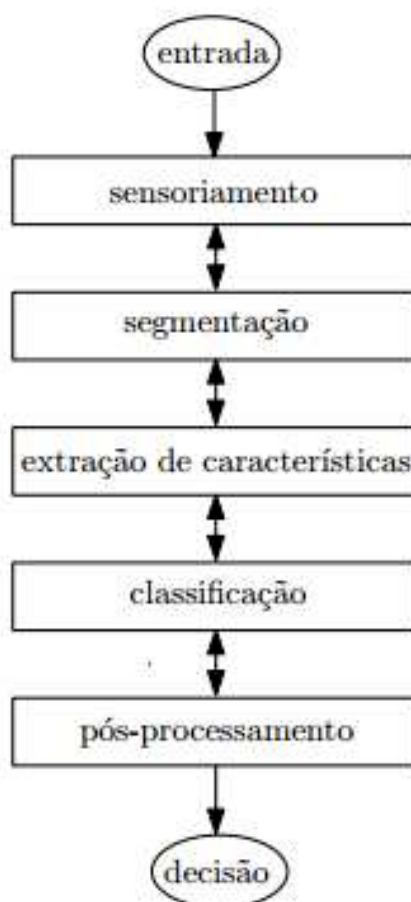
O reconhecimento de texto manuscrito e de gestos efetuados em telas sensíveis ao toque são exemplos de aplicações do campo de estudo denominado reconhecimento de padrões. Seu objetivo consiste na separação de objetos (ou padrões) em categorias (ou classes), uma tarefa não trivial que envolve múltiplas etapas, divididas por [Duda, Hart e Stork \(2001\)](#) em sensoriamento, segmentação, extração de características, classificação e pós-processamento como observa-se na [Figura 37](#).

Sensores ou transdutores convertem estímulos de entrada em conjuntos de dados que podem ser lidos e manipulados por máquinas. O componente de segmentação divide o sinal em porções reconhecíveis (separação de objetos, por exemplo); a extração de características busca características relevantes na entrada; o classificador utiliza as características para encontrar a melhor classe correspondente para a entrada; o pós-processamento executa operações adicionais como integração dos resultados de múltiplos classificadores ou alguma técnica para ponderar as saídas de diferentes classificadores e tomar a decisão final ([DUDA; HART; STORK, 2001](#)).

Na presente situação, a entrada do reconhecedor consiste em uma série de coordenadas bidimensionais quantizadas em uma grade de pixels (correspondentes à tela do dispositivo de interface) e marcadas com os instantes de tempo, capturadas por um rastreador do olhar. A extração de características, classificação e pós-processamento dependem diretamente do método de captura utilizado pelo rastreador.

Não há até o presente momento muitos trabalhos relacionados ao reconhecimento de gestos oculares, mas as abordagens encontradas empregam diversos métodos de conversão do traçado como comparação entre *strings* ([PEDROSA; PIMENTEL; TRUONG, 2015](#)), aprendizado de máquina ([STENDAHL; ARNÖR, 2014](#)) e suas redes neurais recorrentes do tipo *Long Short Term Memory* (LSTM) ([ALSHARIF et al., 2015; JOHNSON; ZHANG,](#)

Figura 37 – Etapas de um sistema de reconhecimento de padrões típico.



Fonte – Duda, Hart e Stork (2001)

2016), *Finite State Transducers* (ANDERSON; BAILEY; SKUBIC, 2004; ELMEZAIN; AL-HAMADI; MICHAELIS, 2009), *Dynamic Time Warping* (KURAUCHI et al., 2016; ZOETEN et al., 2013) ou ainda *Elastic Shape Matching* (KRISTENSSON; ZHAI, 2004; PEDROSA, 2015). As próximas subseções apresentam alguns dos métodos de decodificação do trajeto ocular em texto desenvolvidos e avaliados em pesquisas nos últimos anos.

5.6.1 Comparação entre *Strings*

Pedrosa (2015) desenvolveu dois diferentes métodos de decodificação do traçado do olhar em texto em sua tese de doutorado. O primeiro método, denominado *Filteryedping*, reconhece a palavra desejada efetuando uma busca em uma lista de frequências com todas as palavras possíveis (incluindo variantes das palavras com uma ou mais letras faltando) e ordenando-as, para que o usuário selecione a palavra desejada dentre as cinco mais prováveis.

As palavras da lista de candidatas são organizadas considerando-se seu comprimento e

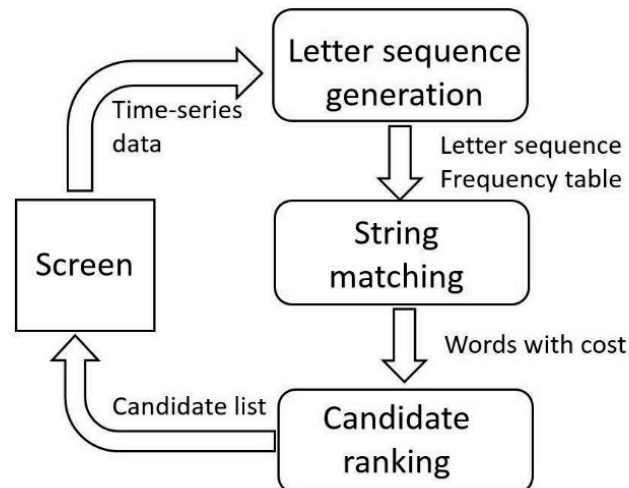
frequência, de acordo com a equação 12, onde $freq(word)$ é o número de vezes que a palavra aparece no corpus, $length(word)$ é o número de caracteres na palavra e w corresponde a um peso definido empiricamente.

$$score(word) = \log_{10}(freq(word)) + w \times length(word) \quad (12)$$

Quanto maior o escore de uma palavra, mais no início da lista de palavras prováveis ela estará. Seu algoritmo de *string matching* apresentou bons resultados, e foi comparado a outro método de reconhecimento do mesmo autor baseado em comparação de forma (PEDROSA, 2015).

Liu et al. (2015) propõem em seu trabalho um sistema de decodificação de traçados oculares em palavras que opera analisando-se a sequência de letras percorrida pelo olhar e a duração de cada letra mirada. O período de tempo gasto em cada tecla permite determinar as letras dominantes (letras em que o usuário passa mais tempo olhando), de forma a atribuir um peso variável a cada letra e tornar o sistema mais robusto com relação erros de letras vizinhas e de letras omitidas. O sistema consiste em três módulos, responsáveis pela geração da sequência de letras, correspondência de palavras e classificação de palavras candidatas respectivamente. A Figura 38 ilustra o método de reconhecimento proposto por Liu et al. (2015).

Figura 38 – Paradigma do sistema de reconhecimento proposto por Liu et al. (2015).



Fonte – Liu et al. (2015, p. 2)

O processo de decodificação desenvolvido por Liu et al. (2015) procura determinar a similaridade entre a sequência de letras inserida pelo usuário e as palavras existentes no léxico, baseado no algoritmo *approximate string matching*. O método desenvolvido pelo autor, apelidado de MoWing (**M**oving **W**indow **S**tring **M**atching), calcula o custo de conversão de uma sequência de letras em uma palavra utilizando duas janelas deslizantes.

O autor afirma ter solucionado alguns dos problemas encontrados no trabalho de [Pedrosa \(2015\)](#).

5.6.2 Aprendizado de Máquina

Outras abordagens buscam efetuar a conversão de gestos em teclados em palavras de um léxico utilizando técnicas de aprendizado de máquina. [Stendahl e Arnör \(2014\)](#) afirmam em seu trabalho que alternativas de aprendizado de máquina como *Self-Organizing Maps* (SOM), *Multilayer Perceptron* (MLP) e *Support Vector Machines* (SVM) são possíveis métodos para reconhecimento de gestos de teclado. O autor desenvolveu um sistema de reconhecimento relativamente simples empregando redes neurais do tipo MLP (*Multilayer Perceptron*) do software *Matlab* sob uma interface desenvolvida em Java. O método apresentado pelo autor registra a sequência de letras percorridas pelo traçado ao se digitar cada palavra do dicionário e alimenta a rede neural.

Seu teclado apresenta layout QWERTY, de forma que para digitar a palavra 'olá', por exemplo, o usuário produziria uma entrada na forma da sequência de caracteres - *o l k j h g f d s a*. Os resultados apresentados por [Stendahl e Arnör \(2014\)](#) constataram altas taxas de erro, devido à baixa robustez com relação a ambiguidades gestuais (palavras diferentes com sequências de caracteres semelhantes). O método de comparação puramente baseado na sequência de letras percorridas não é suficiente para discriminar as entradas possíveis, e apresenta piores resultados à medida que se aumenta o tamanho do léxico.

5.6.3 Redes Neurais Recorrentes do tipo LSTM

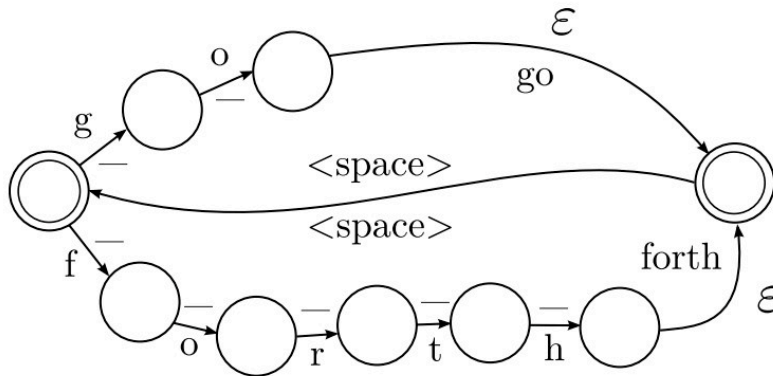
As redes neurais recorrentes pertencem ao ramo de estudo do aprendizado de máquina, mas são apresentadas separadamente nesta subseção devido a suas particularidades e alta eficiência para lidar com linguagem natural em comparação aos métodos apresentados na subseção 5.6.2. As redes neurais recorrentes são muito utilizadas em processos de reconhecimento de fala, modelagem linguística, tradução, reconhecimento manuscrito, legendagem de imagens entre outros. As redes neurais recorrentes do tipo LSTM são altamente eficientes, sendo capazes de aprender a partir de dependências de longo prazo, ou seja, amostras muito antigas.

[Alsharif et al. \(2015\)](#) descrevem em seu artigo uma proposta de decodificação de gestos contínuos de teclado empregando uma abordagem híbrida que combina redes neurais recorrentes (do tipo LSTM) com transdutores de estados finitos (FST), e demonstram ter obtido resultados satisfatórios com léxicos pequenos e também extensos.

A rede neural recorrente seria treinada com a matriz de entrada (coordenadas x e y do traçado, tecla do teclado correspondente, tempo desde o último gesto) utilizando uma função *Connectionist Temporal Classification Loss* (CTC) devido ao desalinhamento e dimensões diferentes das variáveis envolvidas. A rede LSTM retornaria uma matriz

contendo um conjunto limitado de palavras possíveis, a partir da qual é construído um léxico na forma de transdutores de estados finitos (FST) como observado na Figura 39. Os autores afirmam que os resultados são satisfatórios e permitem escalar o método para léxicos com dezenas de milhares de palavras.

Figura 39 – Exemplo de FST do léxico para as palavras *go* e *forth*.



Fonte – Alsharif et al. (2015, p. 3)

5.6.4 *Dynamic Time Warping*

A técnica denominada *Dynamic Time Warping* (DTW) - ou *distorção dinâmica do tempo*, em tradução literal - foi utilizada por vários pesquisadores na conversão de gestos de teclado (gestos *swipe*) efetuados em telas sensíveis ao toque (ZHAI; KRISTENSSON, 2012; ZOETEN et al., 2013) e também efetuados pela direção do olhar (KURAUCHI et al., 2016).

O DTW é definido como um método genérico utilizado para avaliar a similaridade entre dois sinais. Apresentado pela primeira vez na década de 60 (BELLMAN; KALABA, 1959), é empregado comumente no reconhecimento de fala, comparação de sequências de DNA e outros sinais, reconhecimento manuscrito, reconhecimento de símbolos e gestos, *data mining*, visão computacional entre outras aplicações (SENIN, 2008). Os sinais são comparados utilizando-se uma função com pesos para calcular a distância entre eles - pode-se utilizar a distância manhattan, distância manhattan ao quadrado, distância euclidiana ou ainda distância euclidiana ao quadrado. As operações de DTW são computacionalmente intensivas, apresentando complexidade $O(x \cdot y)$ para entradas de dimensões x e y .

O trabalho de Zoeten et al. (2013) emprega um algoritmo de DTW no reconhecimento de gestos *swipe* em telas sensíveis ao toque, de forma a comparar as séries de coordenadas correspondentes ao gesto efetuado e os *templates* das palavras presentes no léxico. A similaridade dos traçados é combinada com um modelo linguístico para que se obtenha a

probabilidade de cada uma das palavras possíveis no contexto. O autor desenvolveu ainda um algoritmo otimizado de DTW de forma a diminuir os requisitos computacionais do processo. Os resultados demonstram desempenho satisfatório com baixas taxas de erro, apesar do dicionário apresentar tamanho reduzido (263 palavras).

O importante trabalho de Zhai e Kristensson (2012) na área do reconhecimento de texto através de gestos *swipe* também adota o *dynamic time warping* para fazer o *template matching* com os traçados de palavras em *touchscreens*. A probabilidade $P(G|W)$ de um gesto G corresponder a uma palavra W é dada pela similaridade entre as duas séries de coordenadas combinada com a probabilidade $P(W)$ de uma palavra aparecer em uma frase - determinada por um modelo linguístico.

Kurauchi et al. (2016) em um trabalho mais recente também faz uso do DTW¹ para reconhecimento de gestos efetuados pelo olhar em um teclado virtual. O autor desenvolveu otimizações no algoritmo para possibilitar a operação em tempo real (requisito de interfaces homem-máquina), visto que o DTW é um processo computacionalmente intensivo.

O autor descreve em um artigo (KURAUCHI et al., 2016) e também em sua tese (KURAUCHI, 2018) o processo de seleção de palavras candidatas que o usuário possa ter digitado, que envolve o cálculo de similaridade entre o traçado do olhar e as formas ideais (*templates*) das palavras do léxico que foram pré-calculadas. Para contornar a elevada complexidade computacional do cálculo da DTW, o autor utilizou uma árvore de prefixos para efetuar os cálculos de similaridades letra-a-letra. Isso possibilitou a execução do algoritmo em tempo real sem necessidade de otimizações de algoritmo que tornariam o código mais complexo.

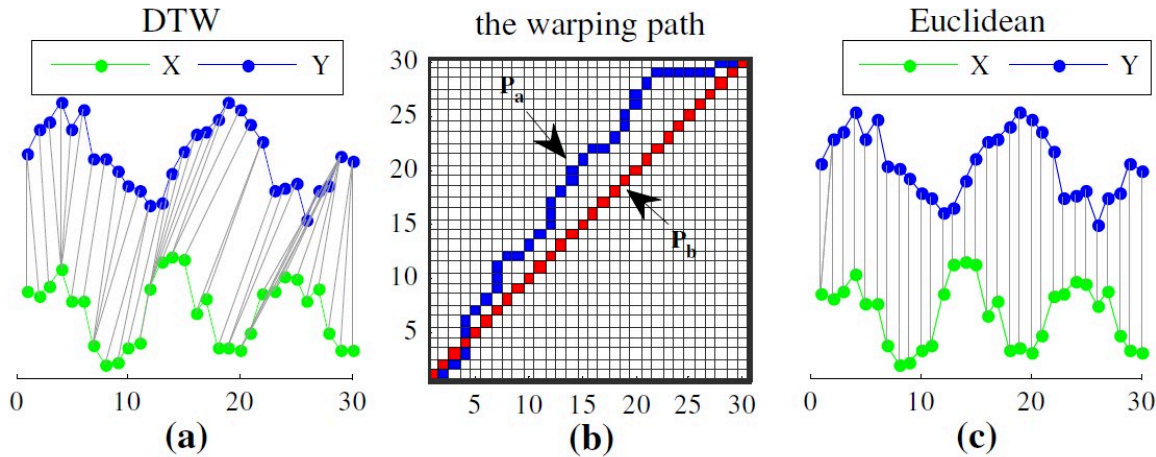
O DTW é um método que encontra o melhor alinhamento entre duas séries temporais, através de uma medida de similaridade que minimiza os efeitos de deslocamentos ou distorções no tempo possibilitando uma transformação 'elástica' nas sequências temporais e detectando formas similares com fases diferentes (SENIN, 2008). A distorção entre as duas séries temporais pode ser utilizada para encontrar regiões correspondentes entre elas ou para determinar sua similaridade. Um exemplo da análise de similaridade entre duas séries assíncronas através do DTW pode ser observado na Figura 40.

A Figura 40(a) representa o alinhamento entre duas séries temporais através do DTW, 40(b) representa o melhor caminho de distorção correspondente, referido como P_a . Em 40(c) pode-se observar a correspondência entre as séries utilizando-se a distância Euclidiana, cujo caminho de distorção é chamado de P_b , também representado em 40(b). Pode-se perceber que o DTW fornece uma melhor correspondência entre pontos que representam forma semelhante, apesar de estarem deslocados no tempo (LI, 2014, p. 2).

Se as duas sequências temporais forem idênticas, todas as linhas conectando as amos-

¹ O autor desenvolveu dois algoritmos de cálculo de similaridade entre formas, um empregando *dynamic time warping* e outro baseado na distância de Fréchet. (KURAUCHI et al., 2016)

Figura 40 – (a) Distância medida pela DTW; (b) *warping paths* produzidos pela DTW e distância Euclidiana; (c) distância Euclidiana.



Fonte – Li (2014, p. 2)

tras como ilustrado na Figura 40 seriam verticais, visto que nenhuma distorção seria necessária para alinhar as séries. A distância do caminho distorcido, ou *warp path*, é a medida de dissimilaridade entre as duas séries, calculada como a soma das distâncias entre cada par de pontos conectados pelas linhas (SALVADOR; CHAN, 2007).

A distância Euclidiana é uma medida de distância comumente utilizada, consistindo simplesmente na soma das distâncias ao quadrado de cada enésimo ponto em uma série temporal ao enésimo ponto em outra série. A desvantagem da utilização da distância Euclidiana é a visualização dos dados pouco intuitiva, além do fato de que se duas séries estão levemente deslocadas no tempo a distância Euclidiana pode considerá-las muito distintas uma da outra. A técnica DTW foi introduzida (BELLMAN; KALABA, 1959) para superar essa limitação e fornecer uma medida de distância intuitiva que ignore deslocamentos locais e globais no eixo temporal (SALVADOR; CHAN, 2007).

Dadas duas séries temporais X e Y , de comprimento $|X|$ e $|Y|$, respectivamente:

$$X = x_1, x_2, x_3, \dots, x_i, \dots, x_{|X|}$$

$$Y = y_1, y_2, y_3, \dots, y_j, \dots, y_{|Y|}$$

Elas constroem um caminho de distorção - *warp path* - W de acordo com a equação abaixo, onde K é o comprimento do caminho de distorção e o k -ésimo elemento desse vetor é w_k ,

$$W = w_1, w_2, \dots, w_k \quad \max(|X|, |Y|) \leq K < |X| + |Y|$$

$$w_k = (i, j)$$

onde i e k são os índices das séries X e Y , respectivamente. O *warp path* deve começar no início de cada série em $w_1 = (1, 1)$ e terminar no fim das duas séries temporais em

$w_k = (|X|, |Y|)$, de forma que cada elemento das duas séries sejam utilizados no caminho de distorção. As linhas verticais observadas na Figura 40 não se cruzam devido a uma restrição que faz com que i e j sejam monotonicamente crescentes (SALVADOR; CHAN, 2007).

O *warp path* ótimo é o que apresenta a menor distância, onde a distância de um warp path W é

$$Dist(W) = \sum_{n=1}^{k=K} Dist(w_{ki}, w_{kj})$$

e $Dist(w_{ki}, w_{kj})$ é a distância entre dois pontos das séries comparadas no k -ésimo elemento do caminho de distorção.

O caminho de distorção de mínima distância é buscado através de métodos de programação dinâmica, onde soluções parciais são utilizadas ao invés de se tentar resolver o problema inteiro de uma só vez. Uma matriz D de custos bidimensional $|X|$ por $|Y|$ é criada, onde o valor em $D(i,j)$ é o caminho de distorção de menor distância encontrado entre as duas séries $X' = x_1, \dots, x_i$ e $Y' = y_1, \dots, y_j$. O valor em $D(|X|, |Y|)$ conterá o caminho de menor distância entre as séries X e Y . Ambos os eixos da matriz D , como ilustrado na Figura 40(b) representam o tempo, sendo a série X no eixo x e a série Y no eixo y . Pode-se ver pela Figura 40 o caminho de menor distância traçado na matriz de custos D , de $D(1,1)$ a $D(30,30)$.

Se o caminho de menor distância passa pela célula $D(i,j)$ da matriz de custos, significa que o i -ésimo ponto da série X está deslocado para o j -ésimo ponto da série Y . As séries temporais não precisam possuir mesma dimensão, visto que um ponto em uma série pode ser mapeado para vários pontos na outra série. Se as séries X e Y fossem idênticas, o caminho de distorção na matriz de custos seria uma linha diagonal. Para encontrar o caminho de menor distância, todas as células da matriz devem ser preenchidas uma coluna por vez, de baixo para cima e da esquerda para a direita.

Após o preenchimento de toda a matriz, o caminho é calculado na ordem inversa, começando em $D(|X|, |Y|)$. Uma busca extensiva é efetuada, através da varredura das células para a esquerda, para baixo e diagonalmente em direção à extremidade esquerda inferior da tabela. A célula entre as três nas direções citadas que possuir menor valor é adicionada ao início do caminho de distorção, e a busca continua até que se chegue à célula $D(1,1)$.

A complexidade computacional do método é de $O(|X|.|Y|)$, e sua natureza quadrática é particularmente proibitiva devido a requisitos de memória. Por essa razão, foram desenvolvidos diferentes métodos para tornar o DTW mais rápido e eficiente. Os métodos podem ser classificados em três categorias:

- Restrições - Limitação no número de células que são avaliadas na matriz de custos.
- Abstração de dados - Execução do DTW em uma representação reduzida dos dados.

- Indexação - Uso das funções com limite inferior para reduzir o número de vezes que o DTW deve ser executado durante a classificação ou clusterização das séries temporais.

5.6.5 *Proportional Shape Matching*

A técnica denominada *Proportional Shape Matching* é muito utilizada no processamento de imagens, e pode também ser utilizada no reconhecimento de gestos de teclado. É, por sua vez, um algoritmo inspirado no *dynamic time warping* e opera buscando a distância entre dois conjuntos de pontos calculando o mínimo de deformação necessária para transformar um conjunto de pontos no outro. É um método invariante em escala, translação e rotação.

Zhai e Kristensson são as referências mais notáveis no ramo do reconhecimento de gestos de teclado. Os autores são pesquisadores da Google e IBM e possuem diversas pesquisas publicadas em importantes periódicos. Eles exploraram várias abordagens na decodificação do traçado de texto, tendo experimentado o *elastic shape matching* - uma modificação do *proportional shape matching* - e o *dynamic time warping*, esta última abordada na seção 5.6.4. O sistema desenvolvido pelos autores foi inicialmente nomeado de *SHARK*, tendo sido posteriormente comercializado com o nome *ShapeWriter* e vendido para outra companhia que integrou as funcionalidades do sistema e descontinuou o aplicativo.

Os importantes trabalhos de Zhai e Kristensson se baseiam no método *Elastic Shape Matching* para decodificação de gestos de teclado efetuados em telas sensíveis ao toque. O autor desenvolveu o sistema apelidado de SHARK² que utiliza as informações de forma juntamente com modelos linguísticos para reconhecer os traçados e decodificá-los em texto² (KRISTENSSON; ZHAI, 2004). O mesmo autor descreve em outro trabalho (KRISTENSSON; VERTANEN, 2012) o potencial do método de reconhecimento de gestos de teclado na digitação através do olhar.

Diversos trabalhos posteriores são baseados no método desenvolvido por Kristensson, como os trabalhos de Zhai e Kristensson (2012), Markussen (2014), Pedrosa (2015), Bi e Zhai (2016), assim como o algoritmo de decodificação de gestos visuais desenvolvido nesta tese.

5.7 Considerações Finais

Este capítulo apresentou uma visão geral acerca dos teclados virtuais, suas características e seu desenvolvimento. Os principais elementos da interface de um teclado virtual

² O sistema foi patenteado (KRISTENSSON; WANG; ZHAI, 2010), desenvolvido como aplicativo de *smartphones* e posteriormente vendido a outra empresa.

foram descritos, e foram referidos os layouts de teclado mais utilizados e abordados na literatura.

Métricas de avaliação de teclados virtuais foram explanadas, e em seguida discorreu-se sobre os teclados gestuais e os métodos de decodificação de gestos de teclado em palavras.

O próximo capítulo descreve o sistema de comunicação assistiva proposto neste trabalho, pormenorizando as etapas de seu desenvolvimento desde a definição da arquitetura à codificação dos vários elementos que compõem o sistema de aquisição e decodificação de gestos oculares de palavra.

Teclado Virtual Proposto

6.1 Introdução

Este capítulo descreve o desenvolvimento do software de um teclado virtual que opera através de gestos oculares. São descritas as etapas de implementação do software que envolvem a criação da interface gráfica do teclado virtual, construção do léxico (ou dicionário) da língua portuguesa a ser utilizado, desenvolvimento do algoritmo de decodificação do traçado ocular em palavras do léxico, e implementação do algoritmo de busca, classificação e desambiguação das palavras de acordo com a forma do caminho efetuado pelo olhar do usuário.

A implementação do presente sistema foi desenvolvida em etapas, desde o desenvolvimento da interface gráfica até a integração com hardware específico de rastreamento ocular. Abaixo estão listadas as etapas sequenciais de desenvolvimento do software:

1. Desenvolvimento da interface gráfica de teclado com *layout* QWERTY contendo barra de palavras prováveis sugeridas;
2. Análise e seleção de *layouts* de teclado alternativos;
3. Organização do léxico em português Brasileiro a ser utilizado pelo teclado;
4. Implementação do mecanismo de captura e normalização das coordenadas do cursor/olhar e armazenamento na forma de gestos correspondentes a palavras;
5. Criação dos *templates* (gestos ideais) a partir do léxico adotado para comparação com gestos reais;
6. Desenvolvimento do sistema de decodificação dos traçados (séries temporais contendo coordenadas de posição do olhar) em palavras presentes no léxico;
7. Elaboração do algoritmo de classificação dos resultados (palavras possíveis) e resolução de ambiguidades.

O desenvolvimento progressivo do algoritmo de reconhecimento de gestos do olhar e sua integração sob a interface de um teclado virtual será descrito nas seções seguintes.

6.2 Interface

A interface de comunicação assistiva pretendida neste trabalho consiste basicamente em um teclado virtual controlado por gestos do olhar. Objetivou-se a estruturação de uma interface intuitiva, responsiva, customizável e de fácil utilização.

Considerando-se requisitos de sistema, buscou-se implementar uma interface que não consuma muitos recursos de memória e processamento (podendo ser utilizada em sistemas com configurações básicas), que seja multiplataforma e que possa ser ajustada e modificada se grandes dificuldades.

O desenvolvimento inicial da interface foi realizado utilizando-se a linguagem QT, destinada à construção de interfaces gráficas de usuário, baseadas em C++. Devido a dificuldades de integração da interface com o algoritmo - desenvolvido em Python para efetiva manipulação de strings do dicionário - substituiu-se a linguagem de programação utilizada, apresentando a interface em uma simples página HTML contendo elementos Javascript que conferem interatividade à interface.

A interface do teclado virtual resultante consiste basicamente em um teclado virtual inserido na forma de um elemento de imagem em uma página HTML, sobre o qual é construído o elemento Javascript responsável pela captura da posição do olhar passada ao cursor, desenho do gesto efetuado na interface e pela interação direta com o algoritmo de reconhecimento de forma desenvolvido em Python.

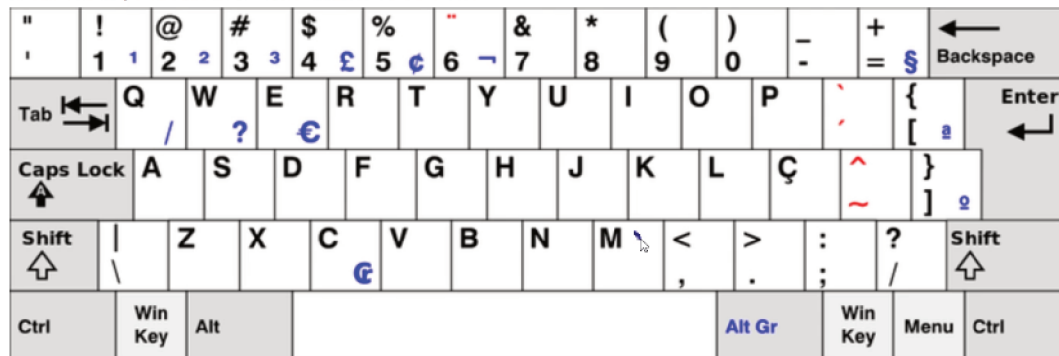
A Figura 41 apresenta a interface básica do teclado virtual embutido em uma página web, contendo o título na parte superior, seguido de uma linha onde são exibidas as palavras decodificadas sugeridas e o tempo de processamento correspondente ao último gesto. Logo abaixo encontra-se o teclado virtual sobre o qual são efetuados os gestos e é retornado *feedback* visual ao usuário na forma da posição do cursor correspondente ao olhar e do traçado efetuado.

O elemento <canvas> presente no ambiente HTML5 possibilita a criação de gráficos utilizando Javascript, assim como a captura de eventos do cursor como cliques e movimentações. O algoritmo 1 apresenta o pseudocódigo do módulo de interface do teclado virtual, que recebe os eventos do cursor e retorna ao algoritmo de decodificação de forma a sequência de coordenadas percorridas entre um clique e uma liberação do mouse.

Figura 41 – Interface do teclado virtual desenvolvido, desenvolvido na forma de uma página HTML.

Decodificador gesto-palavra

Nenhuma palavra encontrada 10.99038ms



Fonte – O autor

Algoritmo 1 Interface do teclado virtual com captura e retorno das coordenadas de posição do cursor ao decodificador e desenho do traçado

Entrada: Posição e eventos do cursor

Saída: Vetor de coordenadas do traçado do cursor

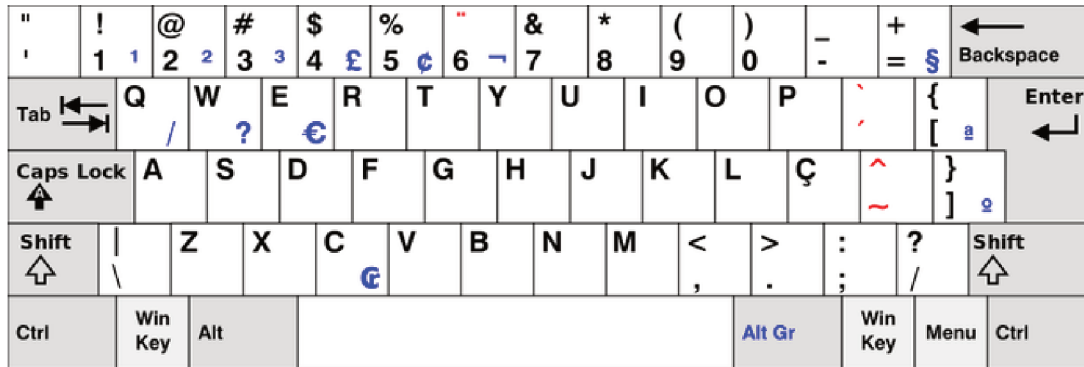
- 1 Definição do canvas com mesmas dimensões do teclado utilizado
 - 2 Definição do traçado efetuado na interface: formato, cor, espessura
 - 3 **Função** *inicioGesto()*:
 - 4 | coordenadas = []
 - 5 | Limpa o canvas
 - 6 | Inicia traçado
 - 7 | coordenadas \leftarrow *Cursor*(x, y)
 - 8 **fim**
 - 9 **Função** *movGesto()*:
 - 10 | Traça linha até posição atual
 - 11 | coordenadas \leftarrow *cursor*(x, y)
 - 12 **fim**
 - 13 **Função** *fimGesto()*:
 - 14 | Interrompe traçado
 - 15 | coordenadas \leftarrow *cursor*(x, y)
 - 16 | **retorna** coordenadas para Decodificador
 - 17 | Recebe lista de palavras prováveis e tempo de processamento de Decodificador
 - 18 | Exibe lista de palavras e tempo de processamento
 - 19 **fim**
-

6.2.1 Disposição das teclas

Diversas opções de disposição das teclas de um teclado existem, desde o layout QWERTY até modelos otimizados para a língua portuguesa ou otimizados para teclados virtuais,

como descrito na seção 5.4. A interface assistiva desenvolvida neste trabalho adotou a disposição de teclas padrão no país - QWERTY - no teclado virtual. Foi utilizada a imagem da Figura 42 para construção da interface interativa baseada em um ambiente web.

Figura 42 – Layout QWERTY utilizado no teclado assistivo.

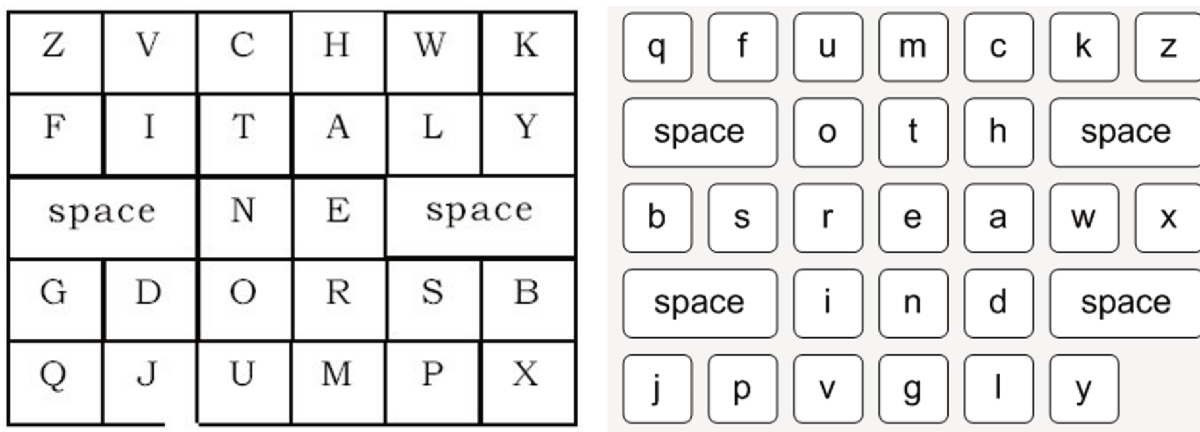


Fonte – O autor

A modularização do algoritmo desenvolvido possibilita a utilização de diferentes configurações de teclado fazendo-se poucos ajustes nos parâmetros do algoritmo. Diversos trabalhos mencionados anteriormente se dedicaram ao desenvolvimento de layouts otimizados de teclados virtuais (LOJA, 2015; SARCAR; PANWAR; CHAKRABORTY, 2013; ANTONIO et al., 2014) ou à avaliação de layouts alternativos já existentes na digitação por gestos (MACKENZIE; READ, 2007; WADDINGTON et al., 2017).

Foram adotados como alternativa ao teclado principal (QWERTY) dois layouts alternativos amplamente validados e apontados como otimizações efetivas da configuração de teclas para avaliação de sua usabilidade na digitação por gestos do olhar: Fitaly e OPTI. A Figura 43 apresenta os teclados otimizados adotados.

Figura 43 – Layouts alternativos Fitaly e OPTI utilizados no teclado assistivo.



(a) Layout Fitaly

(b) Layout OPTI

Fonte – O autor

6.2.2 Tamanho das teclas

O teclado virtual foi planejado para ocupar o máximo possível de área na tela (como observado na Figura 41), levando-se em conta que o controle da interface com os olhos é mais imprecisa que o controle utilizando-se o mouse. A lei de Fitts estabelece que o tamanho das teclas e a distância entre elas influencia diretamente na performance de digitação (FITTS, 1954). Quanto maior a área do teclado virtual, mais confiável e confortável se torna a utilização a interface.

Devido ao formato quadricular dos teclados Fitaly e OPTI, estes ocupam maior área útil da tela e se mostram mais fáceis de utilizar, além de produzirem gestos de palavra com maior distinção entre si. Os experimentos descritos no Capítulo 7 poderão validar essas impressões preliminares.

Ademais, a interface na forma de uma página que pode ser aberta em um navegador comum permite o ajuste do tamanho do teclado livremente, bastando que se ajuste o nível de zoom da página.

6.2.3 Método de interação

Para realizar a inserção de palavras foi adotada uma metodologia híbrida, combinando a execução de gestos do olhar com a ativação de um botão físico. Optou-se por utilizar os dois métodos combinados com a finalidade de diminuir a complexidade de utilização da interface pelos usuários e possibilitar o desenvolvimento e ajuste progressivo do algoritmo.

A presente versão do teclado assistivo requer que o usuário pressione um botão físico durante a execução de cada gesto de palavra, de forma a delimitar o início e fim de cada gesto. Para inserir uma palavra o usuário deve então mirar sua primeira letra, pressionar o botão, percorrer com os olhos todas as letras seguintes e após chegar na última letra da palavra, liberar o botão físico.

Após esse processo a interface retorna automaticamente e em tempo real até cinco palavras cujo gesto mais se aproxima do traçado efetuado pelo usuário, seguido do tempo gasto para encontrar a palavra mais provável.

O método de interação é o mesmo, independente da configuração de teclas utilizada. Objetiva-se na próxima versão da interface assistiva eliminar a necessidade de utilização ativação por botão físico, tornando-o mais adequado e útil para pessoas com distúrbios neuromotores em estágio avançado que podem ter seus movimentos severamente comprometidos.

6.2.4 *Feedback* do teclado

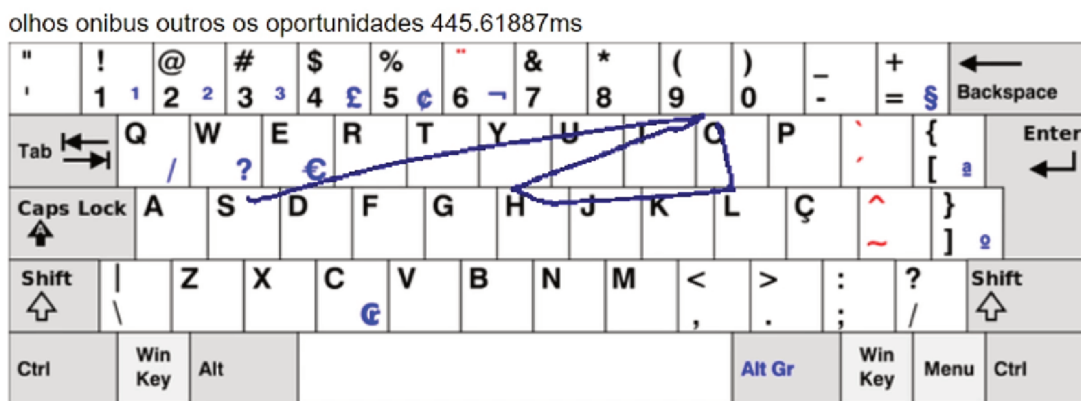
O *feedback* ao usuário é um elemento importante na digitação ocular, que segundo Majaranta et al. (2006) pode afetar significativamente a velocidade de digitação, preci-

são, comportamento do olhar e até mesmo a experiência subjetiva do usuário. Existem implementações de *feedback* do tipo visual, auditivo, háptico entre outros.

O método de *feedback* empregado na interface desenvolvida nesta tese é visual, e envolve a exibição da posição atual do olhar e do desenho do traçado efetuado sobre o teclado virtual, com cor e espessura ajustáveis. O *feedback* visual pode ser desativado a gosto do usuário, anulando a apresentação da posição do olhar ou do traçado sobre a interface.

A Figura 44 representa o *feedback* na forma de cursor visível e traçado da palavra 'olho' inserida pelo olhar.

Figura 44 – *Feedback* visual apresentado ao se efetuar o gesto ocular para a palavra 'olho'.



Fonte – O autor

Versões futuras do teclado assistivo podem incluir *feedback* mais complexo como alteração na cor e tamanho das teclas percorridas, e até mesmo *feedback* auditivo como som de cliques ou resposta em voz da palavra digitada.

6.3 Dicionário em português do Brasil

O desenvolvimento de um teclado virtual que efetue predição de palavras, que utilize digitação através de gestos de teclado ou que implemente otimizações baseadas no idioma requer a integração de um léxico, ou dicionário da língua em utilização. O léxico pode ser extraído do corpus linguístico, que consiste em uma quantidade representativa de texto em uma determinada língua (Michaelis, 1999).

A decodificação de gestos em palavras do português brasileiro desenvolvida neste trabalho é baseada no cálculo da probabilidade de digitação de determinada palavra, que possui um fator de similaridade geométrica e outro fator de modelagem linguística. O modelo linguístico da língua pode ser extraído do corpus do português. O corpus da língua pode ser representado como uma lista de palavras do léxico e sua frequência de ocorrência em uma quantidade massiva de textos. Quanto maior o corpus, ou maior a

quantidade de texto utilizada em sua construção, mais representativo da linguagem em questão será.

Foi utilizado neste trabalho um corpus do português brasileiro desenvolvido por Finatto (2012) a partir de textos de notícias publicados no jornal online 'Massa!' entre 03 de janeiro e 06 de fevereiro de 2013. A amostra total foi extraída de 70 textos processados, onde pode-se identificar 6.101 palavras que não se repetem. O modelo linguístico baseia-se em uma versão do corpus fornecido pela autora contendo as 3.000 palavras mais frequentes da amostra¹. Como salienta a autora, "para um linguista, pessoa que investiga os fenômenos da língua, é sempre útil olhar quais seriam as 3.000 palavras mais frequentes em um texto ou conjunto de textos"(FINATTO, 2012).

Dado o tamanho da amostra do corpus adotado, assume-se que a amostra é representativa, e a probabilidade de utilização de determinada palavra pode ser calculada através da razão entre a frequência de ocorrência da palavra e a quantidade total de palavras do corpus. O modelo linguístico adotado corresponde a um modelo do tipo unigrama, onde a utilização de determinada palavra será considerada um evento independente, não dependendo de palavras anteriores. A probabilidade de palavra pode, portanto, ser calculada de acordo com a equação 13,

$$P(w = W) = \frac{f(W)}{\sum_{i=1}^K f(w_i)} \quad (13)$$

onde k é o número de palavras diferentes presentes no corpus (tamanho do léxico) e $f(w)$ corresponde à frequência de ocorrência da palavra no corpus.

A tabela contendo as palavras do léxico e suas respectivas probabilidades de utilização são utilizadas pelo algoritmo no processo de cálculo das palavras mais prováveis que o usuário tenha desejado digitar. Pode-se observar na Tabela 6 uma amostra contendo as 40 palavras mais frequentes do corpus, frequência e probabilidade de ocorrência calculada para utilização no teclado virtual.

Salienta-se que para a primeira versão do teclado foram suprimidos os elementos de acentuação, caracteres numéricos e de hifenização, de forma a simplificar as condições de avaliação do algoritmo de detecção.

6.4 Decodificação de gestos do olhar

A decodificação do traçado do olhar é semelhante à decodificação de gestos efetuados por canetas ou pelos dedos em interfaces sensíveis ao toque. Para esse tipo de entrada por gestos, a trajetória efetuada é amostrada em intervalos regulares de tempo, retornando dados que contêm as coordenadas de posição correspondentes e instantes de tempo de amostragem. A função do decodificador consiste, portanto, em encontrar a palavra no

¹ Mais informações acerca do corpus utilizado podem ser obtidas no endereço <<http://www.ufrgs.br/textecc/porlexbras/porpopular/indexdm.php>>.

Tabela 6 – 40 palavras mais frequentes do corpus linguístico adotado, ordenadas por frequência e probabilidade de ocorrência.

Palavra	Freq.	Prob.	Palavra	Freq.	Prob.
de	1283	0.055678514	ao	132	0.005728421
o	992	0.04304995	dos	131	0.005685024
a	950	0.041227271	se	122	0.00529445
e	627	0.027209999	mais	111	0.004817081
que	601	0.026081673	ele	91	0.003949139
do	498	0.021611769	segundo	81	0.003515167
da	428	0.01857397	como	77	0.003341579
em	360	0.015622966	anos	73	0.00316799
no	321	0.013930478	mas	71	0.003081196
com	293	0.012715358	r	68	0.002951005
na	286	0.012411578	pelo	68	0.002951005
para	273	0.011847416	também	68	0.002951005
um	225	0.009764354	disse	67	0.002907608
os	219	0.009503971	à	66	0.00286421
não	201	0.008722823	está	66	0.00286421
é	197	0.008549234	das	65	0.002820813
uma	183	0.007941674	pela	62	0.002690622
foi	155	0.006726555	ainda	60	0.002603828
por	153	0.00663976	já	59	0.00256043
as	145	0.006292583	foram	58	0.002517033

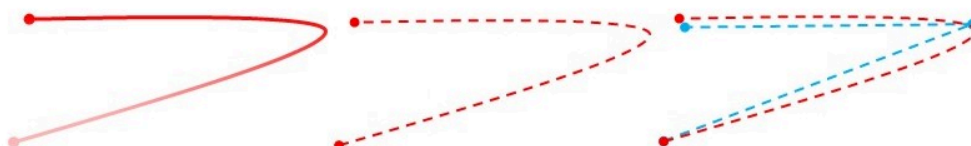
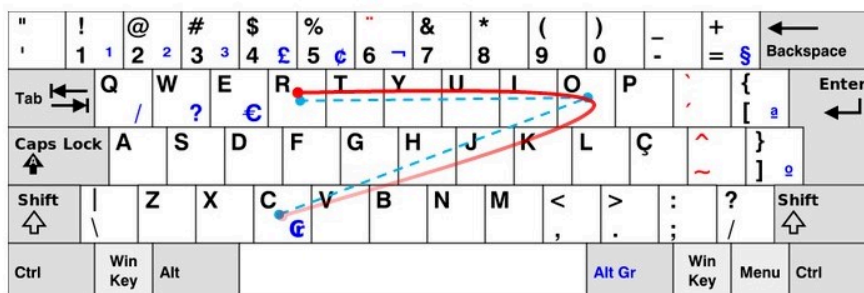
léxico (dicionário) existente com melhor correspondência à sequência de pontos de entrada (correspondente a um gesto ou a um traçado do olhar) (ZOETEN et al., 2013; HOPPE; LÖCHTEFELD; DAIBER, 2013; WOBROCK et al., 2007; LIU et al., 2015).

Não há até momento da publicação desta tese muitos trabalhos relacionados à comunicação assistiva através de gestos oculares, entretanto as abordagens existentes empregam diferentes métodos de reconhecimento. Métodos utilizados englobam a comparação entre strings (PEDROSA; PIMENTEL; TRUONG, 2015), aprendizado de máquina (STENDAHL; ARNÖR, 2014) e redes neurais recorrentes do tipo *Long Short Term Memory* (LSTM) (ALSHARIF et al., 2015; JOHNSON; ZHANG, 2016), *Finite State Transducers* (ANDERSON; BAILEY; SKUBIC, 2004; ELMEZAIN; AL-HAMADI; MICHAELIS, 2009) e técnicas espaciais como *Dynamic Time Warping* (KURAUCHI et al., 2016; ZOETEN et al., 2013) ou ainda *Elastic Shape Matching* (KRISTENSSON; ZHAI, 2004; PEDROSA, 2015) como descrito na seção 5.6.

O algoritmo de reconhecimento de forma empregado no teclado assistivo é baseado no cálculo da distância euclidiana e sua utilização no método *Proportional Shape Matching*, muito utilizado no processamento de imagens e que pode também ser utilizado no reconhecimento de gestos de teclado (KRISTENSSON, 2007; PEDROSA, 2015). É um algoritmo inspirado no *dynamic time warping* e opera buscando a distância entre dois conjuntos de pontos calculando o mínimo de deformação necessária para transformar um conjunto de pontos no outro (TAPPERT, 1982). É um método invariante em escala, translação e rotação (KRISTENSSON; ZHAI, 2004).

A Figura 45 ilustra o processo de comparação entre um gesto efetuado pelo usuário para digitar a palavra 'cor' (em vermelho) e o *template* correspondente (em azul), de forma a mensurar a componente de similaridade de forma.

Figura 45 – Exemplo do processo de decodificação de traçado (vermelho) através da comparação com *template* (azul).



Fonte – O autor

O processo de decodificação dos gestos pode ser descrito pelas seguintes etapas:

- ❑ cálculo das formas ideais (*templates*) para todas as palavras do léxico, contendo uma quantidade arbitrária de 100 amostras por traçado para comparação com gestos efetuados;
- ❑ captura do traçado correspondente a uma palavra (posição do cursor controlado pelo olhar);
- ❑ re-amostragem da série capturada para que tenha mesma dimensão dos *templates* (100 amostras);
- ❑ cálculo da similaridade de forma entre *template* e gesto efetuado, retornando um escore;
- ❑ correção da probabilidade para o gesto efetuado considerando-se a probabilidade de utilização de palavras do léxico;
- ❑ retorno do vetor contendo as palavras mais prováveis (classificadas em ordem decrescente de escore).

O algoritmo de decodificação foi desenvolvido em Python, uma das linguagens mais populares no período de desenvolvimento deste sistema com vasto suporte da comunidade de desenvolvedores e com adoção crescente. A linguagem Python apresenta desempenho

que tem se aproximado dos níveis de linguagens de baixo nível, e foi escolhida neste trabalho principalmente por sua capacidade e flexibilidade de manipulação de strings, listas e dicionários.

A linguagem Python se mostrou, portanto, perfeita para construção e manipulação do léxico - que requer alocação dinâmica de dimensão variável - e comparação entre vetores de posições que acaba envolvendo a utilização de laços aninhados.

O algoritmo de decodificação de forma foi baseado nos princípios apresentados por [Zhai e Kristensson \(2003\)](#) para gestos de teclado em interfaces sensíveis ao toque, e a forma encontrada de interligar a interface Web do teclado ao algoritmo decodificador em Python foi utilizar o *framework* Flask. O algoritmo final de decodificação integrou também adaptações de componentes desenvolvidos pela comunidade, encontrados na plataforma GitHub.

O *framework* Flask permite a utilização da linguagem Python para criar aplicativos Web de forma simples, eficiente e que permite executar o processamento em um módulo separado da interface do usuário².

6.4.1 Criação dos *templates* a partir do dicionário

O primeiro estágio do algoritmo de decodificação de gestos de palavras compreende a criação dos *templates* - ou formas ideais - para todas as 3.000 palavras do léxico no layout de teclado utilizado.

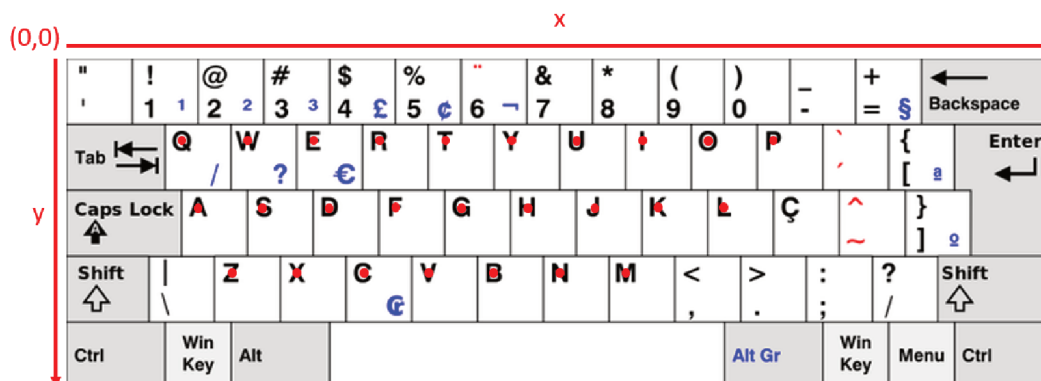
Para que o algoritmo possa operar em tempo real, o gesto efetuado só é comparado com os *templates* cujos pontos inicial e final estejam distantes até um limiar máximo ajustável. Dessa forma não é necessária a comparação entre o gesto efetuado e todas as palavras do léxico, o que tornaria o algoritmo desnecessariamente complexo.

Os *templates* - ou gestos ideais de palavra - são gerados sinteticamente produzindo-se um vetor de unindo as coordenadas centrais das teclas correspondentes às letras de cada uma das palavras do léxico. Os vetores dos *templates* são então re-amostrados para conterem 100 coordenadas de posição. O gesto de palavra efetuado pelo usuário é capturado e também re-amostrado para que possua 100 amostras, de forma a possibilitar o cálculo da similaridade entre gesto e *template* - ambos devem ter a mesma dimensão para serem comparados pelo algoritmo.

A Figura 46 representa o teclado QWERTY utilizado, de dimensão 1.196 x 446 pixels, com as marcações das coordenadas correspondentes a cada letra do alfabeto. Deve-se notar que a origem dos eixos se encontra no vértice superior esquerdo. Os gestos ideais calculados nesta etapa do algoritmo e os gestos efetuados pelo olhar são localizados na área do teclado com relação ao eixo de coordenadas.

² A documentação do microframework Flask pode ser acessada no endereço <https://flask.palletsprojects.com/en/1.1.x>.

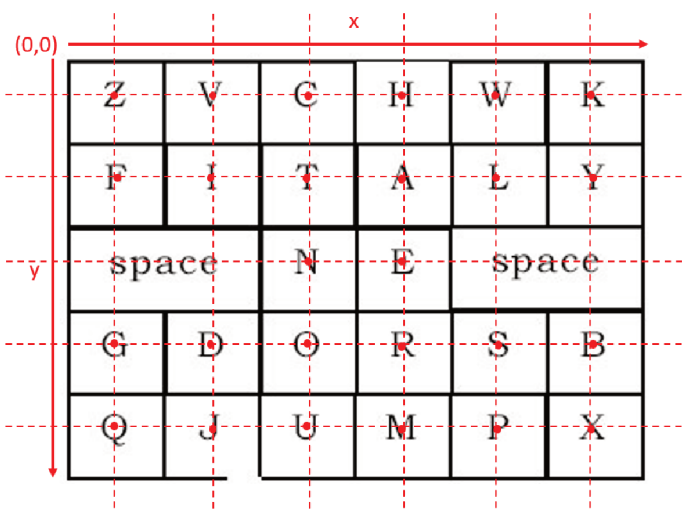
Figura 46 – Indicação da matriz de coordenadas correspondente às letras do teclado QWERTY utilizado na interface assistiva.



Fonte – O autor

O teclado Fitaly, de dimensão 932 x 679 pixels, é representado sobre o eixo de coordenadas e matriz de coordenadas dos caracteres como observado na Figura 47.

Figura 47 – Indicação da matriz de coordenadas correspondentes às letras do teclado Fitaly utilizado na interface assistiva.



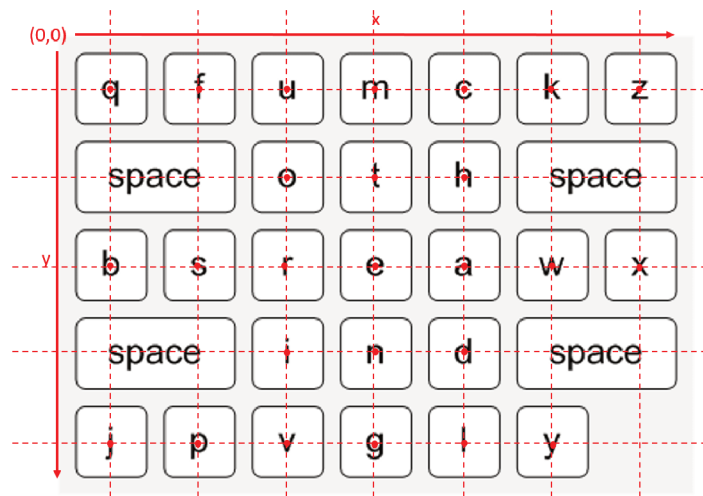
Fonte – O autor

O teclado OPTI é representado na Figura 48, com matriz de coordenadas das letras utilizada na criação dos *templates*. A dimensão do teclado OPTI é 1.076 x 759 pixels.

A tabela 7 representa as coordenadas das letras para os três layouts de teclado da interface desenvolvida. Ao unir os pontos de coordenadas de todas as letras de uma palavra, cria-se o *template* - ou gesto ideal - desta para ser comparada com os gestos efetuados pelo usuário.

O algoritmo de criação dos *templates* está organizado de acordo com o pseudocódigo do Algoritmo 2.

Figura 48 – Indicação da matriz de coordenadas correspondentes às letras do teclado OPTI utilizado na interface assistiva.



Fonte – O autor

Algoritmo 2 Criação dos *templates* do léxico a partir das coordenadas centrais das letras no teclado virtual.

Entrada: Lista de palavras do Corpus do português

Saída: Vetor de coordenadas do traçado do cursor

```

20 LER corpus.txt
21 Define vetores de posições das letras centroides_X, centroides_Y
22 Declaração de vetores pontos_template_X, pontos_template_X
23 para linhas de corpus.txt faça
24     | palavras = corpus[0]
25     | probabilidades = corpus[1]
26     | para letras de palavras faça
27         | pontos_template_X ← centroides_X
28         | pontos_template_Y ← centroides_Y
29     | fim
30 fim
31 retorna pontos_template_X, pontos_template_Y

```

6.4.2 Re-amostragem do traçado efetuado pelo usuário

Os gestos de palavras efetuados pelo usuário e os *templates* do dicionário são representados na forma simples de vetores. Para compará-los eles devem possuir mesmas dimensões, e neste trabalho definiu-se o tamanho padrão dos vetores como 100. É necessário então amostrar 100 pontos equidistantes no gesto do usuário e no gesto ideal.

- Cálculo do comprimento total do gesto calculando-se o comprimento e inclinação de linhas entre pares de pontos sucessivos;

Tabela 7 – Coordenadas das letras para os layouts de teclado adotados, utilizadas no cálculo dos *templates* das palavras do léxico.

Caractere	QWERTY (x, y)	Fitaly (x, y)	OPTI (x, y)
A	(93, 104)	(190, 75)	(275, 155)
B	(300, 149)	(295, 165)	(35, 155)
C	(208, 149)	(135, 30)	(275, 35)
D	(185, 100)	(85, 165)	(275, 215)
E	(173, 53)	(190, 120)	(215, 155)
F	(230, 100)	(30, 75)	(95, 35)
G	(280, 102)	(30, 165)	(215, 275)
H	(323, 102)	(190, 30)	(275, 95)
I	(407, 56)	(85, 75)	(155, 215)
J	(372, 103)	(85, 215)	(35, 275)
K	(415, 101)	(295, 30)	(335, 35)
L	(462, 99)	(240, 75)	(275, 275)
M	(393, 147)	(190, 215)	(215, 35)
N	(346, 145)	(135, 120)	(215, 215)
O	(450, 54)	(135, 165)	(155, 95)
P	(496, 55)	(240, 215)	(95, 275)
Q	(81, 53)	(30, 215)	(35, 35)
R	(220, 55)	(190, 165)	(155, 155)
S	(140, 104)	(240, 165)	(95, 155)
T	(275, 55)	(135, 75)	(215, 95)
U	(359, 58)	(135, 215)	(155, 35)
V	(254, 150)	(85, 30)	(155, 275)
W	(128, 56)	(240, 30)	(335, 155)
X	(163, 152)	(295, 215)	(395, 155)
Y	(313, 55)	(295, 75)	(335, 275)
Z	(15, 150)	(30, 30)	(395, 35)

- Para selecionar 100 pontos amostrais é escolhida uma quantidade de pontos para cada linha proporcionalmente a sua contribuição no traçado total;
- O processo é realizado para criação dos *templates* e re-amostragem dos gestos de palavra;

As fórmulas 14 e 15 são utilizadas para se calcular os pontos amostrais a partir do sinal (+ ou -) obtido pela comparação ponto-a-ponto.

$$x = x_0 \pm (requiredDistance \times \sqrt{\frac{1}{1 + slope^2}}) \quad (14)$$

$$y = y_0 \pm (slope \times requiredDistance \times \sqrt{\frac{1}{1 + slope^2}}) \quad (15)$$

Pode-se observar no algoritmo 3 o método de re-amostragem utilizado para se obter vetores de tamanho padrão de 100 amostras.

Algoritmo 3 Criação dos *templates* do léxico a partir das coordenadas centrais das letras no teclado virtual e re-amostragem de traçado para conter 100 amostras.

Entrada: Vetor de coordenadas de traçado

Saída: Vetor de posições re-amostrado para tamanho de 100 coordenadas

```

32 Declaração de variáveis pontos_amostra_x, pontos_amostra_Y, inclinacao,
    comprimento_linha
33 Função distancia( $x_0, y_0, x_1, y_1$ ):
34 |   retorna  $\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$ 
35 fim
36 Função reamostragem(pontos_X, pontos_Y):
37 |   para cada  $i \in \text{pontos\_X}$  faça
38 |      $\text{comprimento\_linha} \leftarrow \text{distancia}(x_i, y_i, x_{i+1}, y_{i+1})$ 
39 |      $\text{comprimento\_amostragem} \leftarrow (\sum \text{comprimento\_linha})/99$ 
40 |   fim
41 |   para cada  $i \in \text{pontos\_X}$  faça
42 |     se  $x_0 == x_1$  então
43 |        $\text{inclinacao}[i] = \infty$ 
44 |        $x\_pos = x_0$ 
45 |        $y\_pos = y_0 + \text{comprimento\_amostragem}$ 
46 |     fim
47 |     se  $x_0 == x_1 \wedge y_0 == y_1$  então
48 |        $\text{inclinacao}[i] = 0$ 
49 |        $x\_pos = x_0$ 
50 |        $y\_pos = y_0$ 
51 |     senão
52 |        $\text{inclinacao}[i] = (y_1 - Y_0)/(x_1 - x_0)$ 
53 |        $x\_pos = x_0 + (0,001 * \sqrt{1/(1 + \text{inclinacao}[i]^2)})$ 
54 |        $y\_pos = y_0 + \text{inclinacao}[i] * 0,0001(\sqrt{1/(1 + \text{inclinacao}[i]^2)})$ 
55 |     fim
56 |     se  $\text{distancia}(x_1, y_1, x\_pos, y\_pos) < \text{distancia}(x_0, y_0, x_1, y_1)$  então
57 |        $\text{senal}[i] = +$ 
58 |     fim
59 |     senão
60 |        $\text{senal}[i] = -$ 
61 |     fim
62 |   fim
63 |   fim

```

O algoritmo não está integralmente descrito devido a restrições de espaço para representação de algoritmos. A rotina de re-amostragem não produz pontos perfeitamente equidistantes, mas se mostra eficiente e bastante efetiva para a utilização neste trabalho.

6.4.3 Cálculo da probabilidade de palavra

O cálculo da probabilidade de palavra a partir de um gesto é baseado no algoritmo de reconhecimento de gestos de caneta desenvolvido por (KRISTENSSON; ZHAI, 2004), utilizado na grande maioria dos trabalhos acadêmicos relacionados e softwares de reconhecimento de gestos de texto existentes. A comparação entre gesto e *template* é feita integrando-se um escore de similaridade de forma e um escore de localização. O escore final é então ponderado de acordo com a probabilidade de utilização de palavra do léxico, extraída do modelo linguístico.

O cálculo da probabilidade de palavra é efetuado de acordo com as etapas seguintes:

□ Filtragem por localização

Devido à grande quantidade de palavras presentes no léxico, é necessário filtrar os *templates* com os quais o gesto do usuário será comparado, antes de prosseguir às etapas posteriores de comparação. O mecanismo de restrição aplicado considera as coordenadas iniciais e finais do *template*, normalizadas em escala e translação.

É calculada a distância entre os pontos iniciais e os pontos finais dos *templates* e do gesto desconhecido. Se qualquer uma das distâncias for maior que um limiar definido, o *template* é descartado.

O processo de filtragem reduz significativamente o processo de comparação entre gesto e *template*, armazenando somente as coordenadas iniciais e finais dos *templates* em uma lista separada, utilizada na redução de possíveis *templates* a serem considerados nas comparações subseqüentes (KRISTENSSON; ZHAI, 2004).

O valor usual de limiar de localização para reconhecimento de gestos de caneta é igual a 15. Para comparação com os gestos de palavra realizados com os olhos nesta tese determinou-se o limiar igual a 20, com resultado satisfatório.

□ Cálculo do escore de similaridade de forma

Dentre as diversas técnicas possíveis para reconhecimento de gestos descritas na seção 5.6, este trabalho emprega o algoritmo *Proportional Shape Matching*. Esta técnica mensura a similaridade entre dois padrões através da comparação ponto-a-ponto da sequência amostral.

Primeiro é calculada a média das coordenadas x e y do gesto. A média é subtraída de cada valor do gesto. Em seguida é efetuada a normalização do gesto, colocando o lado maior da área de gesto em escala com um comprimento predeterminado L . O fator de normalização é calculado de acordo com a equação 16, onde W e H são as dimensões do teclado: $W = \max(x) - \min(x)$ e $H = \max(y) - \min(y)$.

$$s = \frac{L}{\max(W, H)} \quad (16)$$

Em seguida todas as coordenadas do gesto são normalizadas através da multiplicação pelo fator s . O mesmo é realizado para todos os *templates* candidatos válidos - que passaram pela etapa de filtragem.

É calculada então a distância ponto-a-ponto entre gesto e cada um dos *templates* candidatos, divididos pelo número total de pontos amostrais N dos traçados (KRISTENSSON; ZHAI, 2004). A comparação, realizada pelo método *Proportional Shape Matching*, é representada na equação 17.

$$x_s = \frac{1}{N} \times \sum_{i=1}^N \|u_i - t_i\|_2 \quad (17)$$

O resultado final corresponde a uma medida aproximada de distância entre os traçados, invariante em escala e deslocamento, baseada na média do somatório entre a distância espacial dos pontos amostrais equidistantes.

O algoritmo 4 representa o processo de cálculo do componente de forma.

Algoritmo 4 Cálculo do escore de similaridade de forma entre traçado do usuário e *templates* das palavras válidas.

Entrada: Coordenadas dos traçados re-amostrados do usuário e dos *templates*

Saída: Escore de forma após comparação entre os traçados

```

64 Definição do canvas com mesmas dimensões do teclado utilizado
65 Definição do traçado efetuado na interface: forma, cor, espessura
66 Função escore_forma(pts_gesto_X, pts_gesto_Y, pts_template_X, pts_template_Y):
67   L = 1
68   media_gesto_X = ( $\sum$  pts_gesto_X) / length(pts_gesto_X)
69   media_gesto_Y = ( $\sum$  pts_gesto_Y) / length(pts_gesto_Y)
70   pts_gesto_X1 = (pts_gesto_X - media_gesto_X) * (L / max(width, height))
71   pts_gesto_Y1 = pts_gesto_Y - media_gesto_Y * (L / max(width, height))
72   para  $i \in$  pts_template_X faça
73     | media_template_X = ( $\sum$  pts_template_X) / length(pts_template_X)
74     | media_template_Y = ( $\sum$  pts_template_Y) / length(pts_template_Y)
75     | pts_template_X1 = (pts_template_X - media_template_X) * (L / max(width,
76     |   height))
76     | pts_template_Y1 = pts_template_Y - media_template * (L / max(width, height))
77   fim
78   para  $i \in$  pts_template_X faça
79     | dist_euclidiana = 0
80     | para  $i \in$  pts_template_X faça
81     |   | dist_euclidiana += distancia(pts_gesto_X1[ $j$ ], pts_gesto_Y1[ $j$ ],
81     |   |   pts_template_X1[ $j$ ], pts_template_Y1[ $j$ ])
82     |   | escore_forma  $\leftarrow$  dist_euclidiana / length(pts_template_X)
83     |   fim
84   fim
85   retorna escore_forma
86 fim

```

□ Cálculo do escore de localização

(KRISTENSSON; ZHAI, 2004) descrevem o cálculo do escore de localização, necessário para fazer a distinção entre gestos de formas similares mas deslocados no espaço. A equação 18 estabelece o método de cálculo do escore de localização x_L (KRISTENSSON; ZHAI, 2004, p. 47):

$$x_L = \sum_{i=1}^N \alpha(i)\delta(i) \quad (18)$$

onde $\alpha(i)$ e $\delta(i)$ são fatores de ponderação definidos pelo algoritmo e N é o número de pontos amostrais dos traçados - igual a 100 neste caso. O fator $\delta(i)$ e consequentemente o escore de localização x_L serão iguais a zero caso os contornos comparados estejam próximos até um limiar denotado raio r , definido de acordo com o tamanho das teclas do teclado virtual e aqui definido como $r = 15$.

O fator $\alpha(i), i \in (1, N)$ são pesos para cada comparação ponto-a-ponto entre os traçados de forma que $\sum_{i=1}^N \alpha(i) = 1$. Pode ser definido de diferentes formas mas foi definido uniformemente como $\alpha(i) = 0,01$, visto que foram adotados 100 pontos amostrais por traçado - $N = 100$.

□ Cálculo do escore final e retorno das palavras mais prováveis

Os escores de forma e localização exprimem a distância entre gestos de palavra desconhecidos do usuário e *templates* do dicionário. Os escores não estão na mesma escala, e para serem comparados diretamente são ponderados por dois coeficientes, de forma e de localização. Ambos os coeficientes foram definidos neste trabalho com o valor 0,5, e o escore final pode ser expresso como $escore.final = coef.forma \times escore.forma + coef.localização \times escore.localização$.

A lista final de palavras prováveis é obtida organizando-se as palavras em ordem decrescente de escore final. A lista de palavras sugeridas no teclado virtual possui tamanho arbitrário, definida na interface desenvolvida para apresentar as cinco palavras mais prováveis.

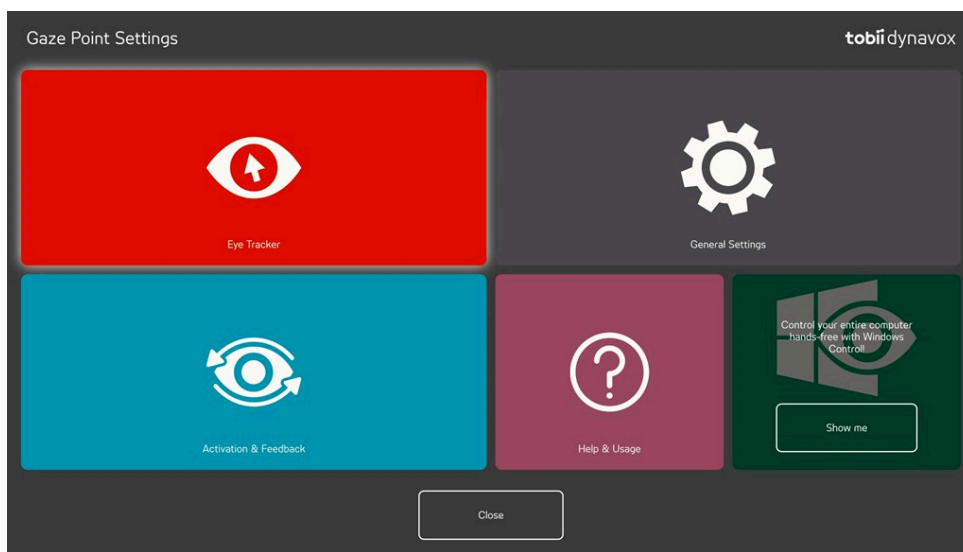
6.5 Integração com rastreador ocular

O algoritmo de decodificação de gestos de teclado adotado foi pensado para utilização em telas sensíveis ao toque. Foi adaptado neste trabalho para receber uma sequência de coordenadas de gesto em um teclado virtual. A sequência de coordenadas na presente situação é obtida e transmitida à rotina do decodificador desenvolvida em Python por uma interface Web com elementos Javascript. A interface recebe a entrada do contorno de palavra a partir do cursor do mouse. A integração do teclado virtual com o rastreador ocular se resume, portanto, ao controle da posição do cursor utilizando o olhar.

O rastreador ocular utilizado, Tobii 4c, funciona automaticamente ao se conectar ao computador. Requer somente a instalação do driver Tobii Core Software. Para controlar a posição do cursor utilizando o *stream* de coordenadas do olhar fornecido pelo oculômetro foi utilizado o software Gaze Point, desenvolvido pela mesma fabricante do rastreador Tobii Dynavox.

O software Gaze Point³ permite que o usuário controle o cursor, faça seleções e cliques utilizando o olhar. É um software gratuito e de fácil configuração. A Figura 49 exibe a interface de configuração do software.

Figura 49 – Interface de configuração do software Gaze Point de controle do cursor através do olhar.



Fonte – Tobii (2016, online)

Após instalação do rastreador ocular foram realizados testes de controle do cursor com os olhos e da interface assistiva desenvolvida na forma de página Web. Os resultados foram satisfatórios, e foi necessário efetuar a calibração do rastreador e ajuste de coeficientes e posições dos centroides de caracteres no teclado virtual.

6.6 Considerações Finais

Esta seção apresentou de forma progressiva e detalhada o desenvolvimento do teclado virtual assistivo, considerando-se separadamente cada módulo do sistema, nomeadamente a interface, o léxico adotado, o algoritmo de decodificação dos traçados do olhar em palavras, a codificação do algoritmo e a integração com o rastreador ocular Tobii 4c.

O próximo capítulo descreve os experimentos realizados para avaliação do desempenho do sistema de comunicação assistiva. É apresentado o planejamento das sessões de teste,

³ Informações e link para download do software Gaze Point estão disponíveis no endereço <<https://www.tobiidynavox.com/software/free-resources/gaze-point-1>>.

da amostragem e recrutamento de voluntários para os testes de usabilidade. São apresentados os resultados obtidos nos experimentos em condições de simulação e condições de utilização real por usuários sem novatos na utilização de rastreadores oculares.

Experimentos e Análise dos Resultados

7.1 Introdução

Neste capítulo são descritos experimentos realizados para avaliação objetiva e subjetiva do sistema de entrada de texto através de gestos do olhar desenvolvido. Os capítulos anteriores forneceram motivos os norteadores do desenvolvimento do sistema de digitação através de gestos oculares, assim como conceitos e fundamentos técnicos que possibilitam a digitação por gestos.

Apesar de todo o embasamento técnico e processo de desenvolvimento apresentado, a utilidade e eficácia do sistema de digitação desenvolvido neste trabalho só pode ser aferida empiricamente, através de testes de usabilidade que simulam condições reais.

A tarefa de digitação depende diretamente da habilidade do usuário, um fator complexo e ainda não compreendido profundamente. Capacidades de digitação de um indivíduo como velocidade e precisão estão associadas a suas habilidades motoras e cognitivas, sobretudo em condições de restrição de comunicação aqui abordadas.

Fatores como experiência prévia, idade e até mesmo língua nativa influenciam no desenvolvimento de habilidades motoras e escrita (KRISTENSSON, 2007). Na escrita através de formas alguns usuários possuem melhor técnica de escrita manual e desenho que outros.

A estimação da velocidade média de digitação é um processo complexo, devido à diferença de habilidade, estilo de digitação e neste caso até mesmo à acuidade visual dos usuários, que podem afetar o desempenho de digitação. A velocidade real de digitação é altamente individual e praticamente impossível de se aferir em um experimento controlado, como salienta Kristensson (2007).

Devido a esses fatores e ao escopo desta tese, além de restrições provocadas pela pandemia do novo coronavírus no período de desenvolvimento do trabalho, investigações e experimentos empíricos mais aprofundados acerca do desempenho de digitação ocular com usuários reais serão desenvolvidos futuramente. São descritos neste capítulo, entretanto, experimentos com escala reduzida mas que possibilitam a avaliação do sistema

desenvolvido em diferentes condições.

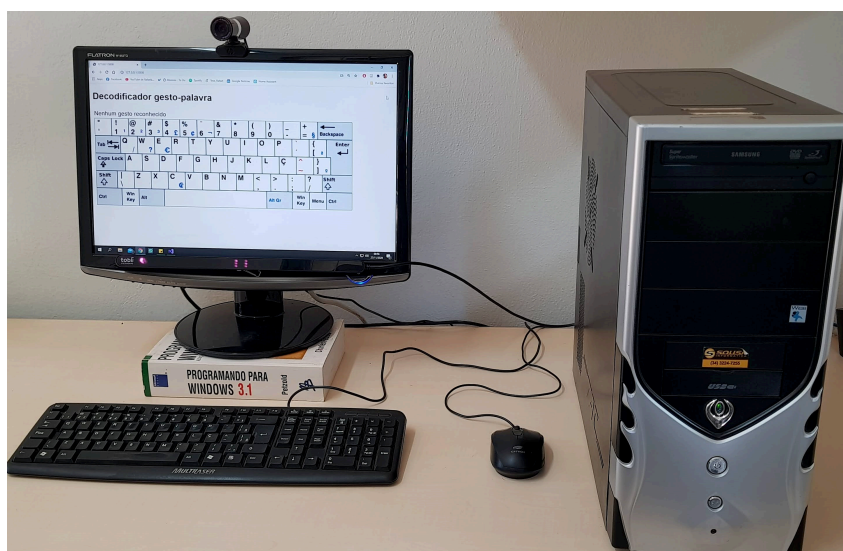
Este capítulo apresenta, portanto, os protocolos do experimento, incluindo condições dos testes, equipamentos utilizados, condições de calibração, indivíduos sujeitos aos testes e frases de teste utilizadas. São apresentados os resultados obtidos, de forma a averiguar a eficiência do método de digitação através do olhar proposto.

Os experimentos foram planejados com o intuito de avaliar dois aspectos principais do teclado virtual desenvolvido: (a) a precisão da decodificação dos gestos do usuário e (b) a velocidade efetiva de digitação obtida utilizando-se a interface assistiva.

7.1.1 Equipamentos

Os testes foram realizados em um computador de mesa com configuração que inclui processador Intel i5-2500k com 4 núcleos a 3,3GHz, 6 gigabytes de memória DDR3-RAM, placa de vídeo NVIDIA GeForce GT710, sistema operacional Windows 10 64-bit. O teclado virtual foi exibido em um monitor LCD de 19 polegadas com resolução de 1440x900 pixels e taxa de atualização de 60Hz. O rastreador ocular Tobii 4C foi utilizado nos testes de usabilidade, instalado na parte inferior do monitor como pode ser observado na Figura 50. A varredura da posição do olhar é efetuada à taxa de atualização de 90Hz, e todos os testes foram efetuados após a certificação de que os *drivers* correspondentes - *Tobii Eye Tracking Core* - estavam atualizados e o rastreador recalibrado para cada usuário. O software Gaze Point, desenvolvido pela *Tobii Dynavox*, é responsável pelo controle do cursor a partir das coordenadas do olhar rastreadas pelo dispositivo.

Figura 50 – Sistema utilizado nos testes de usabilidade, com rastreador ocular acoplado ao monitor.



Fonte – O autor.

A interface do teclado virtual foi desenvolvida como descrito na subseção 6.2, tendo sido projetada para ser executada em navegadores web básicos com compatibilidade *javas-*

cript. Os gestos de palavras são efetuados no teclado virtual e registrados pelo rastreador ocular conectado ao sistema ou através do próprio cursor do mouse.

Todos os experimentos foram realizados utilizando-se o léxico do português brasileiro contendo 3.000 palavras - seção 6.3 - com as quais foram formuladas frases de teste. Os experimentos foram realizados em três etapas, como descrito nas próximas subseções.

7.2 Experimento 1: Avaliação da taxa de reconhecimento de traçado de palavras

7.2.1 Método

O algoritmo de decodificação de gestos em palavras é o elemento central do sistema de comunicação assistiva proposto. Os primeiros testes realizados objetivaram aferir a capacidade de detecção dos gestos de palavras pelo algoritmo.

O desempenho do reconhecimento de gestos foi avaliado inicialmente utilizando o algoritmo alimentado com léxico de tamanho variável na disposição de teclas do tipo QWERTY, utilizando-se 5, 10, 50, 100 e 500 palavras classificadas em ordem decrescente de acordo com sua frequência de ocorrência. Os gestos das palavras dos léxicos utilizados foram efetuados utilizando-se o mouse do computador, de forma a simular os traçados efetuados pelo olhar e possibilitar o desenvolvimento e ajuste do algoritmo antes de avançar para os testes em condições reais com o rastreador ocular.

Em seguida, o mesmo teste de utilização do algoritmo com léxico de tamanho variável foi realizado com layouts de teclado alternativo, Fitaly e OPTI. Os testes sintéticos utilizando o cursor foram necessários para avaliação preliminar e ajuste fino de parâmetros do sistema, em diferentes disposições de teclas.

O sistema foi então avaliado de forma integrada, digitando-se através de gestos oculares as 100 primeiras palavras do léxico de treinamento contendo todas as 3.000 palavras do dicionário, utilizando o rastreador ocular Tobii 4c.

Os resultados obtidos nesta fase dos experimentos possibilitam avaliar se o sistema é utilizável em condições de simulação e em condições reais, e se a taxa de acertos é suficientemente aceitável.

7.2.2 Análise dos Resultados

Ao se utilizar o mouse para simular gestos de teclado obteve-se taxas de acerto acima de 90% demonstrando a efetividade da detecção de traçados. A tabela 8 apresenta a taxa de reconhecimento aferida digitando-se extensivamente todas as palavras de léxicos de diferentes tamanhos.

Tabela 8 – Taxa de reconhecimento de gestos de palavras feitos com o mouse para tamanho crescente do léxico utilizado.

Tamanho do léxico	Taxa de reconhecimento
5	100 %
10	100 %
50	98 %
100	96 %
500	93 %

Para léxicos pequenos é esperado que a taxa de reconhecimento de formas seja maior em relação a léxicos contendo muitas palavras, uma vez que uma quantidade menor de formas precisa ser comparada, diminuindo-se a probabilidade de ambiguidade. Como pode-se observar pela tabela obteve-se uma taxa de decodificação acima de 90% mesmo em um léxico contendo 500 palavras.

Os experimentos utilizando gestos sintéticos efetuados pelo mouse do computador foram necessários para o ajuste fino dos parâmetros de interface e de reconhecimento de forma. O mesmo teste de reconhecimento utilizando diferentes tamanhos do léxico foi efetuado com layouts alternativos de teclado Fitaly e OPTI, comprovadamente mais eficiente em diversas situações. A avaliação comparativa da taxa de acertos pode ser observada na tabela 9.

Tabela 9 – Taxa de reconhecimento de gestos de palavras feitos com o mouse para diferentes tamanhos de léxico e layouts de teclado.

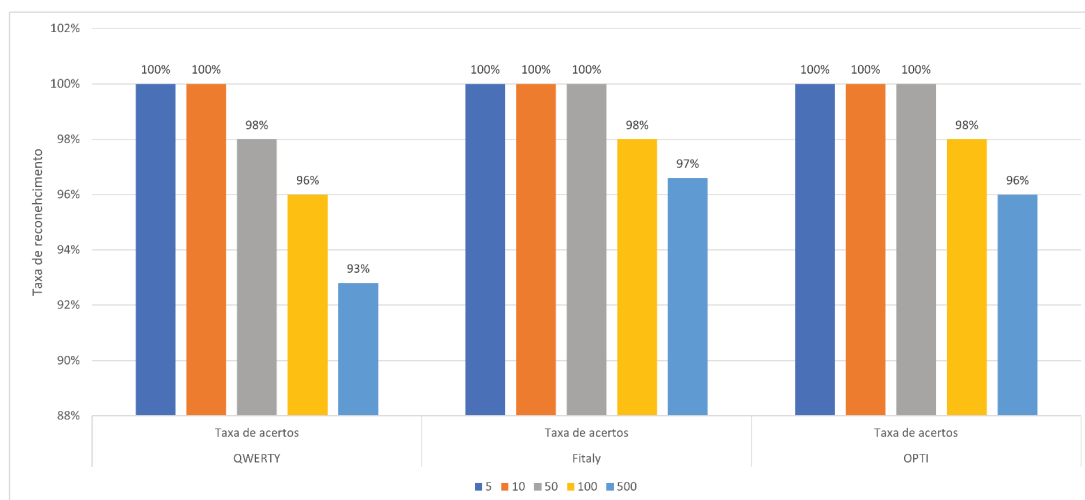
Tamanho do léxico	QWERTY	Fitaly	OPTI
5	100 %	100%	100%
10	100 %	100%	100%
50	98 %	100%	100%
100	96 %	98%	98%
500	93 %	97%	96%

O gráfico da Figura 51 representa a taxa de reconhecimento comparativa para cada layout de teclado.

Como esperado, o aumento na quantidade de *templates* em léxicos maiores resultam em desempenho ligeiramente inferior de comparação de gestos. Um fator que contribui para a maior taxa de acertos com os léxicos menores é o menor tamanho das primeiras palavras do léxico - como observado na Tabela 6 - que resultam em gestos menos complexos com maior capacidade de diferenciação.

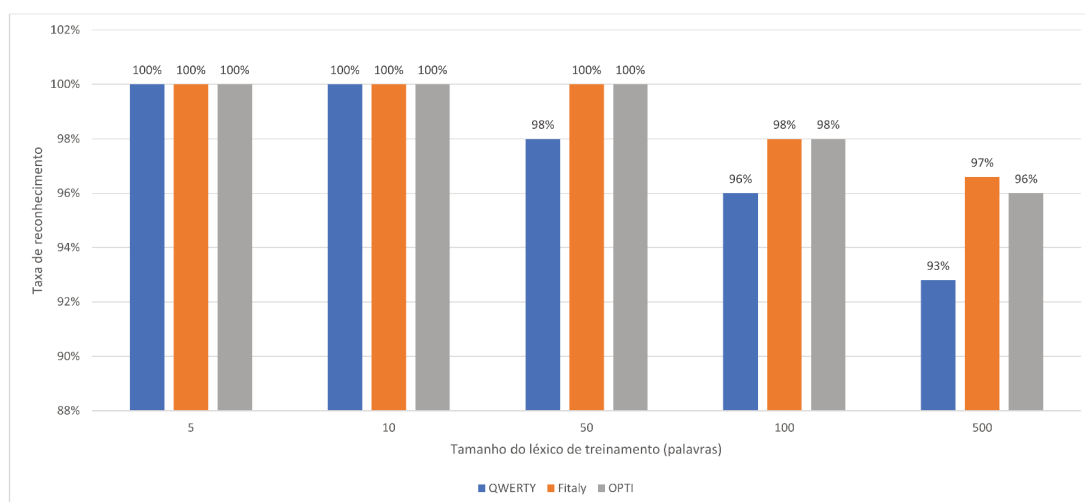
Pode-se agrupar os dados por em relação ao tamanho do léxico de treinamento como disposto na Figura 52, para que se possa observar melhor o decréscimo da taxa de reconhecimento bem sucedido com o aumento na quantidade de *templates* do léxico a serem comparados.

Figura 51 – Taxa de reconhecimento de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.



Fonte – O autor.

Figura 52 – Taxa de reconhecimento de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.



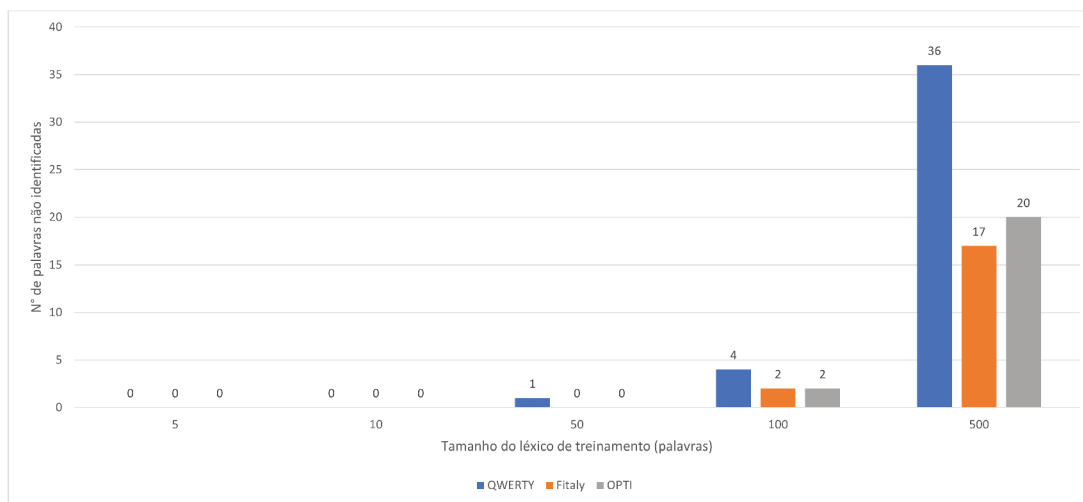
Fonte – O autor.

A quantidade de palavras não identificadas pelo algoritmo é inversamente proporcional à taxa de acertos, e está representada no gráfico da Figura 53.

A tabela 10 exemplifica os vetores de palavras sugeridas pelo algoritmo - contendo até cinco palavras classificadas em ordem decrescente de probabilidade - e o tempo de processamento, para cinco palavras do léxico selecionadas de forma aleatória.

Testes de digitação de palavras aleatórias indicam que não há uma relação direta entre tamanho de palavra e tempo de processamento, como pode-se pensar a princípio. A palavra 'inconstitucionalidade', por exemplo, apresentou um tempo de processamento médio de 286 milissegundos, menor que o tempo de processamento de palavras menores

Figura 53 – Quantidade de palavras não identificadas pelo algoritmo a partir de gestos feitos pelo mouse para diferentes tamanhos de léxico de treinamento no layout QWERTY, Fitaly e OPTI.



Fonte – O autor.

Tabela 10 – Exemplo de palavras sugeridas na decodificação e tempo de processamento para cinco gestos de palavras efetuados.

Palavra digitada	Palavras sugeridas	Tempo
de	['de', 'desse', 'desde', 'deste', 'deve']	319 ms
que	['que', 'quase', 'quantidade']	67 ms
com	['com', 'comum']	201 ms
para	['para', 'passada', 'podera', 'praca', 'parada']	1173 ms
segundo	['segundo', 'sendo', 'sido', 'sandro', 'servico']	838 ms

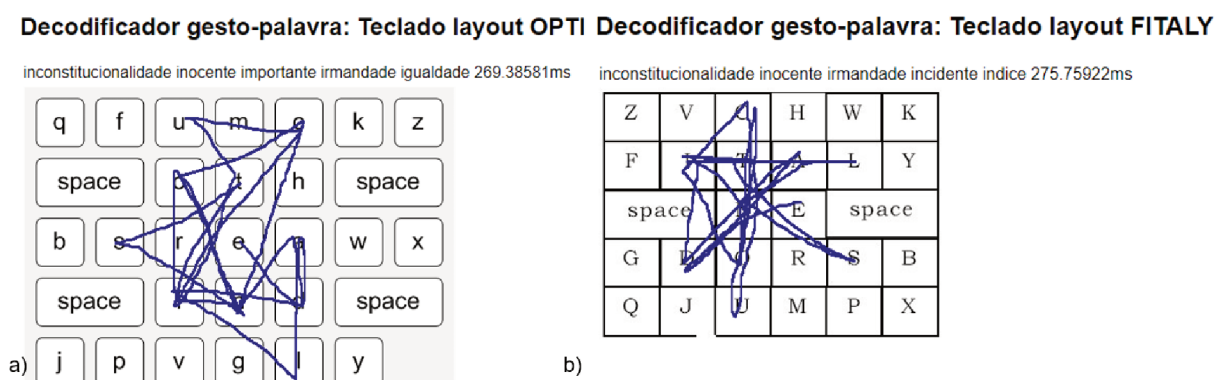
apresentadas na Tabela 10, por exemplo.

Deve-se salientar o bom desempenho do decodificador obtido digitando-se palavras extensas como por exemplo as entradas 'encaminhados', 'administrativamente', 'inconstitucionalidade'. Essas foram corretamente reconhecidas através de seus gestos de palavras, inclusive nos layouts de teclado alternativos OPTI e Fitaly. A Figura 54 representa a entrada de gestos para a palavra 'inconstitucionalidade' no teclado assistivo com layouts alternativos, onde observa-se a correta detecção da palavra desejada e o tempo de processamento no reconhecimento.

Em seguida foi efetuada uma análise comparativa da taxa de acertos do algoritmo utilizando-se todo o dicionário contendo 3.000 palavras, realizando-se os gestos de 100 palavra através do cursor do mouse em um layout do tipo QWERTY - amostra de controle - e da posição do olhar nos layouts QWERTY, Fitaly e OPTI. Cada gesto foi efetuado em no máximo três tentativas de acerto. A tabela 11 apresenta os resultados obtidos pelos dois métodos de entrada.

Pode-se observar pela tabela que a utilização do léxico completo contendo três mil

Figura 54 – Entrada de gestos para a palavra 'inconstitucionalidade' no teclado assistivo, no layout (a)OPTI e (b)Fitaly.



Fonte – O autor.

Tabela 11 – Acurácia na decodificação de gestos oculares para 100 palavras, de léxico contendo 3.000 palavras.

Método de entrada	Taxa de reconhecimento
Gestos do Mouse - layout QWERTY	99 %
Gestos do Olhar - layout QWERTY	89 %
Gestos do Olhar - layout Fitaly	96 %
Gestos do Olhar - layout OPTI	98 %

palavras não provocou queda significativa na taxa de reconhecimento. A simulação de gestos utilizando o mouse resultou em uma taxa de erros de somente 1%, enquanto o controle da interface com os olhos resultou em taxa de erros de reconhecimento de 11%, ambos no layout QWERTY. Os layouts alternativos obtiveram desempenho superior na precisão de reconhecimento, correspondendo a 96% e 98% para os layouts Fitaly e OPTI, respectivamente.

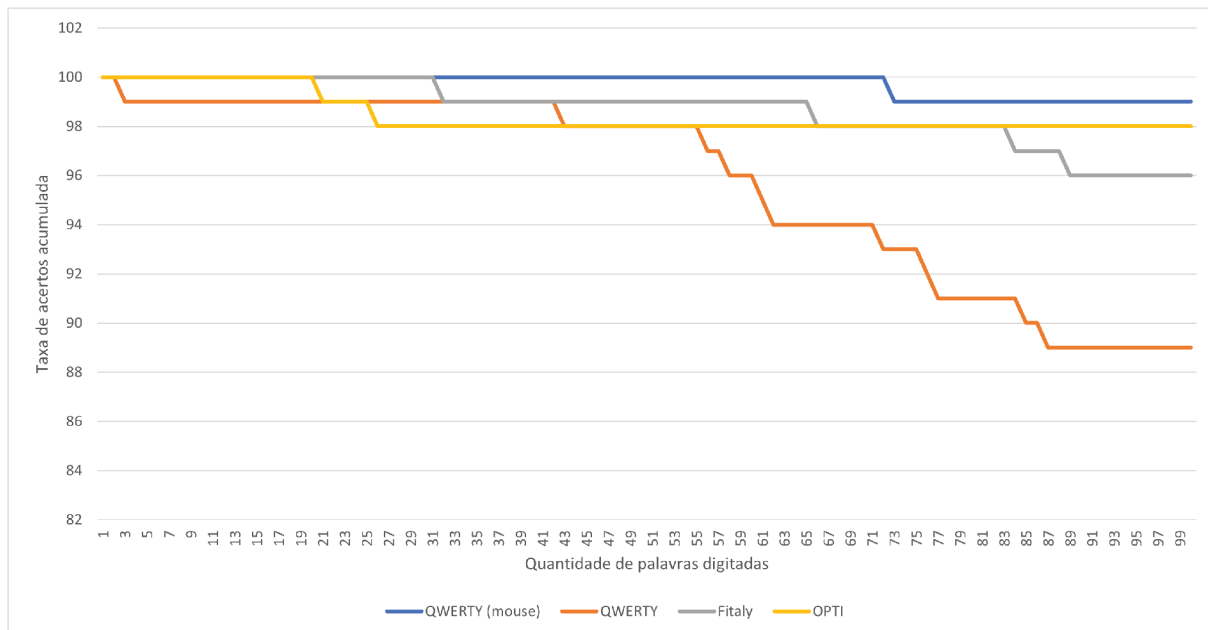
A Figura 55 representa a acurácia de forma cumulativa para as quatro condições de teste, ao se inserir sucessivamente as 100 palavras de teste. Pode-se observar que o layout QWERTY apresentou maior decréscimo na taxa e reconhecimento, em comparação com os layouts alternativos.

7.3 Experimento 2: Avaliação comparativa da digitação ocular com usuários sem experiência prévia

7.3.1 Método

Após comprovar a eficácia do sistema assistivo de digitação através dos gestos do olhar, o próximo passo naturalmente consiste na avaliação do desempenho de digitação

Figura 55 – Taxa de reconhecimento acumulada na entrada de 100 gestos de palavras, de léxico contendo 3.000 palavras.



Fonte – O autor.

de usuários reais que não possuem experiência prévia de utilização de rastreadores oculares ou interfaces assistivas. Simultaneamente foi realizada uma avaliação comparativa entre o método de digitação por gestos proposto e o método tradicional de digitação ocular utilizando fixações.

Os usuários participantes dos testes digitaram uma frase selecionada em quatro diferentes situações: utilizando digitação ocular através de gestos do olhar em três layouts de teclas diferentes e utilizando a digitação ocular letra-a-letra, de forma a avaliar a diferença na taxa de reconhecimento correto e velocidade de digitação entre diferentes *layouts* de teclado.

Os resultados obtidos envolvem métricas relevantes como precisão de digitação, quantidade de erros de palavra, quantidade de teclas por caractere (KSPC), velocidade de digitação, além de opiniões subjetivas dos usuários relacionadas ao (des)conforto de utilização, fadiga visual, preferência entre os métodos oferecidos.

7.3.2 Participantes

Foram recrutados 5 voluntários para os testes de usabilidade (4 homens e 1 mulher), adultos saudáveis com idade entre 25 e 33 anos - média de idade de 30,2 anos - que não apresentavam condições de restrição de comunicação. Dois dos indivíduos utilizaram óculos para correção de miopia e/ou astigmatismo durante a sessão do experimento. Devido à condição de distanciamento imposta pela pandemia do novo coronavírus durante o período de testes foi necessário restringir o recrutamento de voluntários externos à Uni-

versidade, assim como diminuir o tempo e a quantidade de sessões de teste, obedecendo-se as diretrizes sanitárias impostas pela instituição.

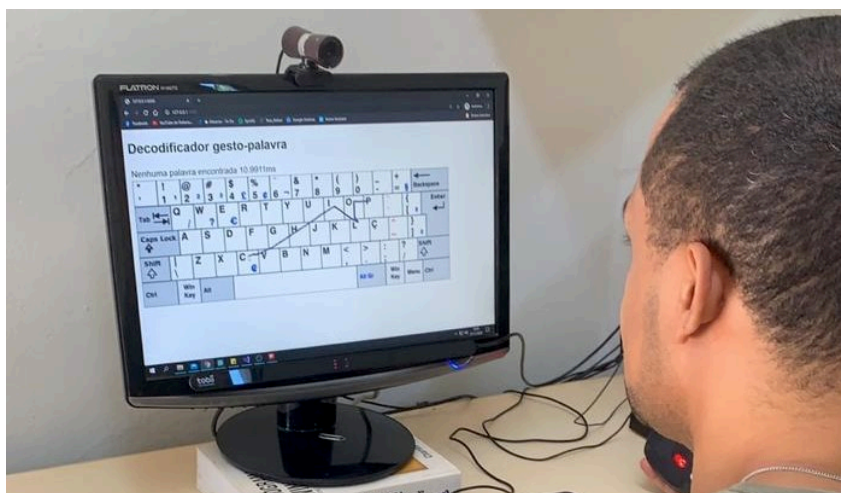
Todos os participantes do experimento possuem experiência na utilização de teclados QWERTY e possuem o português brasileiro como língua materna. Nenhum dos indivíduos recrutados possuíam experiência prévia de utilização de rastreadores oculares ou teclados assistivos, de forma a representar o primeiro contato com um teclado assistivo na primeira sessão de treinamento.

Os usuários foram convocados para uma sessão de testes envolvendo a digitação utilizando os olhos. As sessões de aproximadamente 30 minutos foram agendadas individualmente, não havendo contato direto ou comunicação entre eles a respeito do procedimento.

7.3.3 Procedimento

Ao início da sessão experimental foram apresentados a cada participante os fundamentos da pesquisa, os objetivos do experimento, o método de utilização da interface. Os participantes do experimento foram instruídos a se sentar à frente do monitor a uma distância de aproximadamente 60 centímetros, posicionados com o centro da tela ao nível dos olhos. O rastreador ocular estava instalado na borda inferior da tela, produzindo pulsos de varredura infravermelho invisíveis e inofensivos a seres humanos. Pode-se observar na Figura 56 a condição de teste de um usuário na bancada.

Figura 56 – Disposição de usuário na bancada de testes.



Fonte – O autor.

Após a apresentação do protocolo de teste foi efetuada a calibração do rastreador ocular para cada indivíduo e foram dados 3 minutos para que o usuário se familiarizasse com a interface assistiva do teclado. O processo de calibração do rastreador ocular foi feito individualmente para cada usuário, de forma a minimizar imprecisões e erros de escorregamento - deve ser criado e configurado um perfil de calibração personalizado para

cada indivíduo. O processo de calibração envolve a fixação do olhar pelo usuário em sete diferentes pontos em extremidades da tela. As coordenadas de posição do olhar são automaticamente mapeadas para o cursor do mouse, aplicando correções de posição e algoritmos de suavização embarcados no software do dispositivo. A sessão do experimento foi dividida em duas etapas:

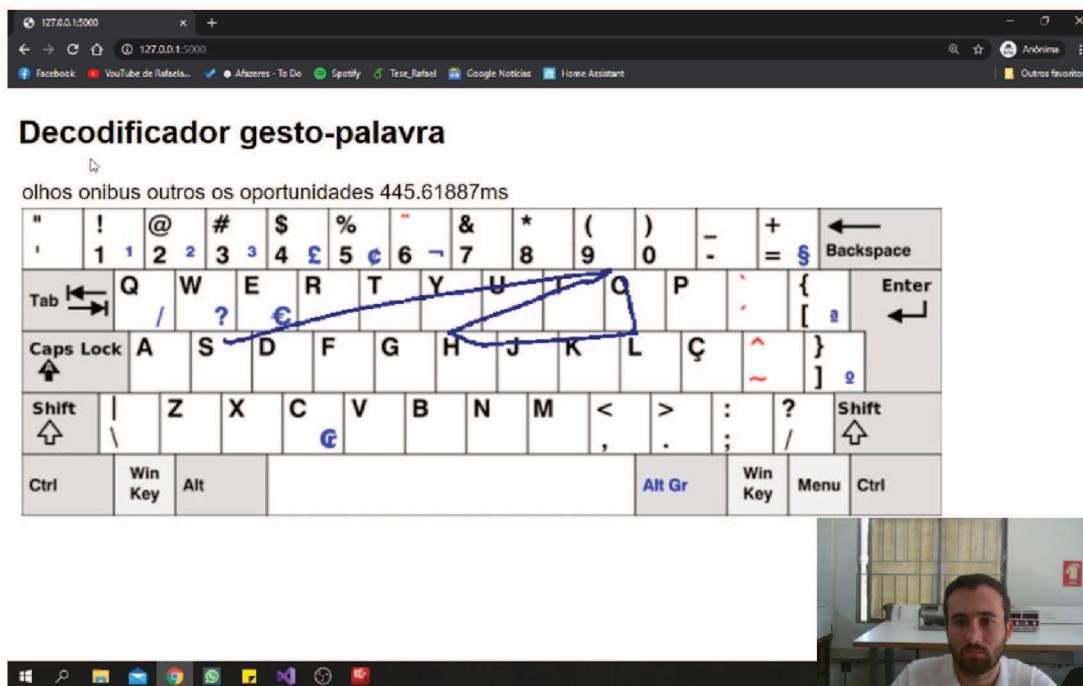
- Avaliação da taxa de acertos e da quantidade de erros de palavra:

Os usuários foram instruídos a digitar 20 palavras através de gestos do olhar na interface do teclado virtual apresentado. As palavras foram selecionadas de forma aleatória e independente, a partir do corpus adotado de 3000 palavras. O conjunto de palavras aleatórias era único para cada sujeito do experimento.

Cada palavra era apresentada ao indivíduo, que deveria lê-la em voz alta e então realizar a digitação utilizando o movimento dos olhos, em três tentativas para cada palavra. Após a realização de cada gesto do olhar, cinco prováveis palavras eram exibidas na parte superior da interface, e caso a primeira sugestão correspondesse à palavra desejada a próxima palavra era apresentada.

A Figura 57 reproduz uma situação de utilização da interface assistiva durante o experimento, com digitação ocular por gestos no teclado QWERTY. A sessão experimental foi registrada em vídeo, contendo a captura de tela com a interface e imagem do usuário durante a utilização.

Figura 57 – Usuário durante realização do experimento.



Fonte – O autor.

□ Avaliação da velocidade de digitação em diferentes layouts:

Para avaliação da velocidade de digitação os indivíduos foram instruídos a digitar a mesma frase contendo 45 caracteres: "tentativa de escrita com os olhos sem paradas". Os participantes digitaram a frase utilizando gestos do olhar em três diferentes configurações de teclas e utilizando a digitação ocular letra-a-letra na interface existente OptiKey com layout QWERTY.

Os participantes do experimento utilizaram as interfaces sem treinamento prévio. Foi dado 1 minuto para familiarização com cada layout antes de iniciar a digitação da frase. Foi registrado o tempo decorrido desde o início do primeiro gesto efetuado até o fim do gesto bem sucedido para a última palavra. Foi registrado também a quantidade de correções de palavra efetuadas.

7.3.4 Análise dos Resultados

Os quatro participantes do experimento digitaram 20 palavras aleatórias utilizando gestos do olhar em um teclado virtual com layout QWERTY. Os resultados obtidos na forma de quantidade de correções de palavra e sucesso no reconhecimento de gesto estão presentes na Tabela 12.

A taxa de reconhecimento - porcentagem de palavras apresentadas pela interface retornadas como primeira palavra provável na lista de resultados - obtida pelos participantes foi de 95%, 95%, 85% e 60% para os participantes do experimento P1, P2, P3 e P4 respectivamente. A taxa de reconhecimento média para o experimento com usuários sem treinamento corresponde, portanto, a 83,75%. A Figura 58 apresenta a taxa de reconhecimento para os quatro participantes do experimento.

Os usuários tiveram registrados a quantidade de erros de palavra durante a digitação por gestos, respectivamente 16, 17, 15 e 29 palavras corrigidas para os participantes P1, P2, P3 e P4, resultando em taxa média de erros de palavra de 19,25 palavras para uma sequência de 20 palavras, como representado no gráfico da Figura 59. Salienta-se que o resultado se mostra bastante promissor, considerando-se a inexperiência dos usuários com a utilização de rastreadores oculares e digitação por gestos.

Na segunda etapa do experimento, ao se digitar a frase estipulada contendo 45 caracteres utilizando quatro interfaces diferentes, os usuários obtiveram as métricas observadas na Tabela 13.

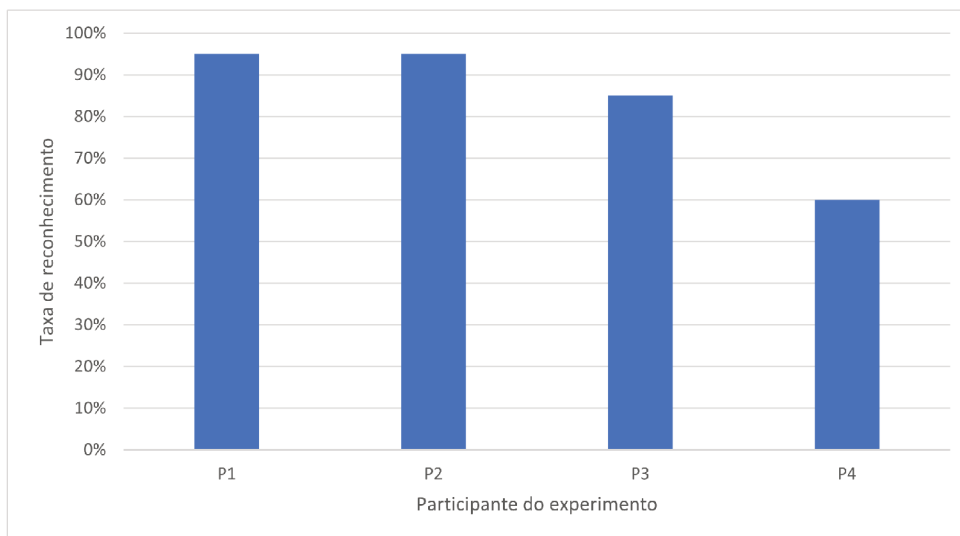
Pode-se identificar na tabela o tempo médio de digitação da frase em quatro diferentes situações: os participantes levaram em média 95,75 segundos na interface QWERTY, 151,5 segundos na interface Fitaly, 127,5 segundos na interface OPTI e 118,25 segundos na interface QWERTY utilizando fixações letra-a-letra.

Utilizando-se a equação 9 calcula-se a velocidade de digitação, em palavras por minuto (PPM). Os valores obtidos foram 7,41 PPM na interface QWERTY, 3,56 PPM na inter-

Tabela 12 – Teste de digitação por gestos do olhar de 20 palavras selecionadas aleatoriamente, por quatro indivíduos.

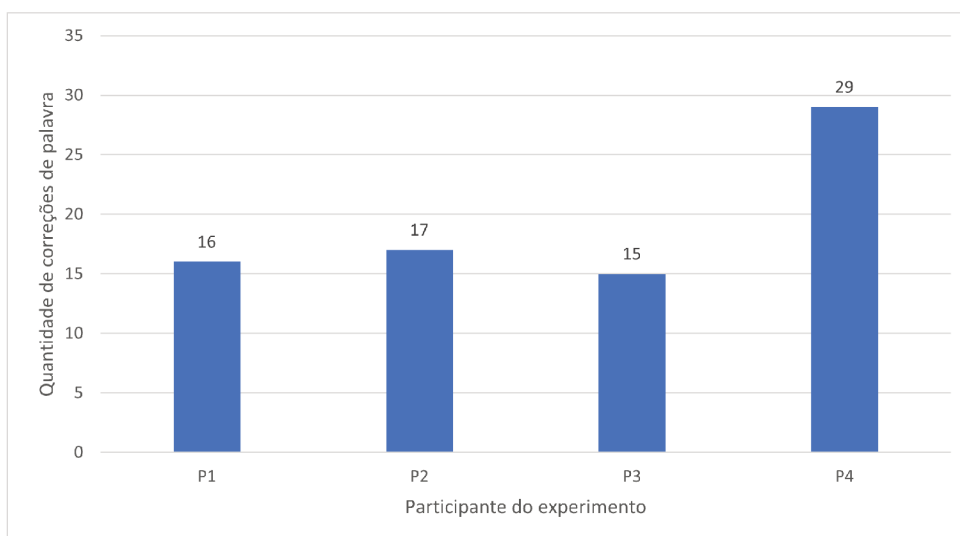
P1			P2			P3			P4		
Palavra	Erros	Recon.	Palavra	Erros	Recon.	Palavra	Erros	Recon.	Palavra	Erros	Recon.
caja	2	✓	amigos	0	✓	garante	0	✓	matadores	3	×
pescar	0	✓	salta	2	✓	nova	2	✓	pescar	1	✓
protestos	2	✓	sofreu	1	✓	divergencias	0	✓	acontecer	2	✓
descargas	1	✓	voltar	0	✓	detalhes	0	✓	sub	0	✓
pensar	0	✓	pelo	1	✓	embora	3	×	blitzes	3	×
coracao	1	✓	gasolina	1	✓	suspeitos	1	✓	proprio	3	×
fazer	0	✓	ressaltou	0	✓	revela	3	×	solidos	0	✓
nesses	0	✓	incendiado	0	✓	chao	0	✓	tomou	0	✓
descumpriu	1	✓	ouvido	2	✓	protetor	3	×	estabelecida	3	×
mamede	2	✓	palco	1	✓	museu	0	✓	cairo	0	✓
invadir	0	✓	paulista	2	✓	acionamento	0	✓	selecoes	0	✓
busca	0	✓	deslocamento	2	✓	considerou	0	✓	piedade	3	×
suposto	2	✓	judicial	0	✓	mortas	0	✓	atrasados	2	✓
percussao	0	✓	inconstitucionalidade	3	×	demonstra	0	✓	comprar	0	✓
cozinheira	0	✓	raciocinio	1	✓	autorizada	2	✓	rosangela	3	×
fornecimento	1	✓	humanos	0	✓	conduta	0	✓	jovem	0	✓
atraso	3	×	pedradas	0	✓	completou	0	✓	inconstitucional	0	✓
acrescentou	0	✓	londres	1	✓	lima	1	✓	zona	0	✓
conversado	1	✓	anteontem	0	✓	infancia	0	✓	china	3	×
aproximacao	0	✓	pregoes	0	✓	fantasia	0	✓	pediram	3	×
N. Correções	16			17			15			29	
Reconhecimento (%)		95%		0	95%		0	85%			60%

Figura 58 – Taxa de reconhecimento de gestos oculares de 20 palavras aleatórias no layout QWERTY.



Fonte – O autor.

Figura 59 – Número de correções de palavra na digitação por gestos oculares de 20 palavras aleatórias no layout QWERTY.



Fonte – O autor.

face Fitaly, 4,29 PPM na interface OPTI e 4,65 PPM na interface QWERTY utilizando fixações letra-a-letra, como disposto na Tabela 14.

A Figura 60 representa a medida comparativa da velocidade de digitação nas condições de teste agrupadas por layout de teclado.

Nota-se que os layouts alternativos Fitaly e OPTI resultaram em velocidades de digitação comparáveis ao sistema amplamente utilizado atualmente baseado em fixações, mesmo sendo configurações de teclas vistos pela primeira vez pelos participantes. O teclado OPTI até mesmo superou o desempenho em palavras por minuto do teclado OptiKey. A Figura

Tabela 13 – Tempo de digitação (t) e quantidade de erros de palavra na digitação uma frase de 45 caracteres em diferentes teclados oculares.

	QWERTY		Fitaly		OPTI		OptiKey	
	t (s)	N. Erros	t (s)	N. Erros	t (s)	N. Erros	t (s)	N. Erros
P1	58	7	129	8	98	2	99	0
P2	53	3	187	18	163	10	117	0
P3	206	10	154	14	118	5	160	3
P4	66	4	136	8	130	6	97	2
Média	95,75	6	151,5	12	127,5	5,75	118,25	1,25

Tabela 14 – Tempo médio de digitação de caractere e velocidade de digitação para uma frase, em diferentes layouts de teclado.

Layout de teclado	Tempo médio por caractere (s)	Velocidade (PPM)
Layout QWERTY	2,18	7,41
Layout Fitaly	3,44	3,56
Layout OPTI	2,89	4,29
Layout QWERTY - Fixação	2,69	4,65

61 apresenta os dados agrupados por participante do experimento.

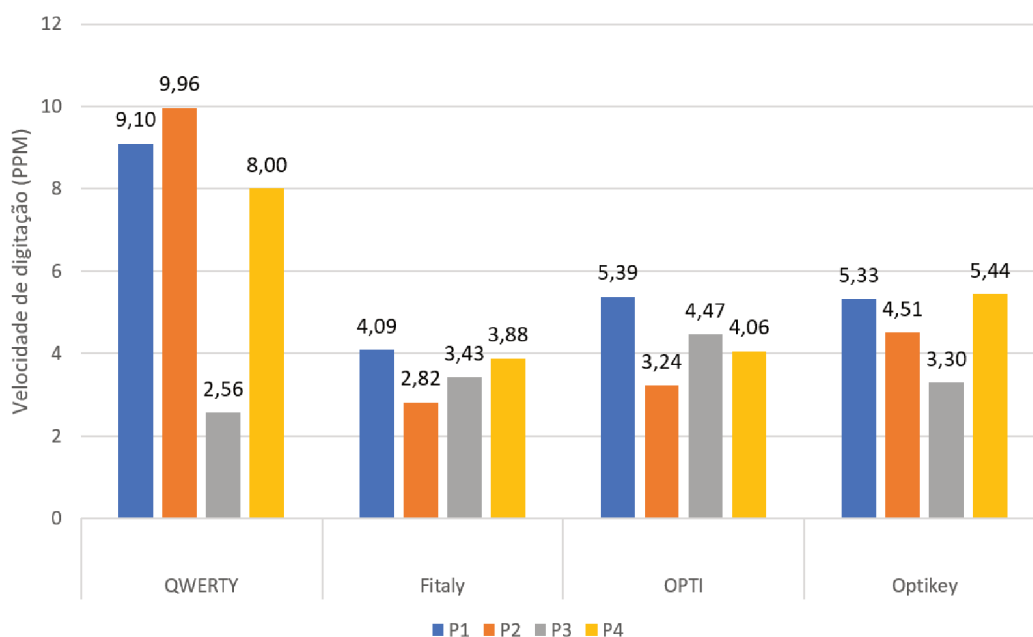
O melhor desempenho foi obtido digitando-se através de gestos do olhar na interface QWERTY. Os resultados ligeiramente superiores podem ser atribuídos à familiaridade dos participantes com o layout, e talvez à prática de digitação na primeira parte do experimento. O boxplot da Figura 62 representa estatísticas de desempenho de forma comparativa entre os diferentes layouts de teclado.

O teclado ocular convencional OptiKey foi responsável pela menor taxa de erros de palavra do teste, ainda que não tenha sido o mais rápido.

Deve-se notar que a velocidade média de digitação pode parecer inferior à encontrada na literatura relacionada (KRISTENSSON; VERTANEN, 2012; LIU et al., 2016; PEDROSA, 2015) mas se refere ao resultado de uma sessão inicial de digitação, uma vez que os indivíduos participantes do experimento nunca haviam utilizado um rastreador ocular para qualquer fim ou utilizado teclados com layout alternativo. Sessões posteriores de treinamento certamente conduziram a resultados mais satisfatórios e próximos de condições de utilização real.

O desempenho do rastreamento do olhar não é prejudicado pela utilização de lentes de correção - óculos ou lentes de contato - como informado pelo fabricante do rastreador ocular e reportado pelos participantes do experimento que utilizaram óculos. Não foram identificadas, portanto, alterações perceptíveis de usabilidade da interface assistiva ao se utilizar óculos. A Figura 63 representa a execução do experimento por participante que utiliza óculos ocasionalmente e reportou não haver diferença perceptível na utilização do teclado virtual com ou sem óculos.

Figura 60 – Análise comparativa da velocidade de digitação em diferentes configurações de tecla.



Fonte – O autor.

7.4 Considerações Finais

Este capítulo apresentou o planejamento e execução dos experimentos de avaliação do sistema de digitação ocular por gestos desenvolvido. inicialmente foram realizados testes com o algoritmo, antes de passar à etapa de testes com indivíduos reais sem experiência com digitação ocular.

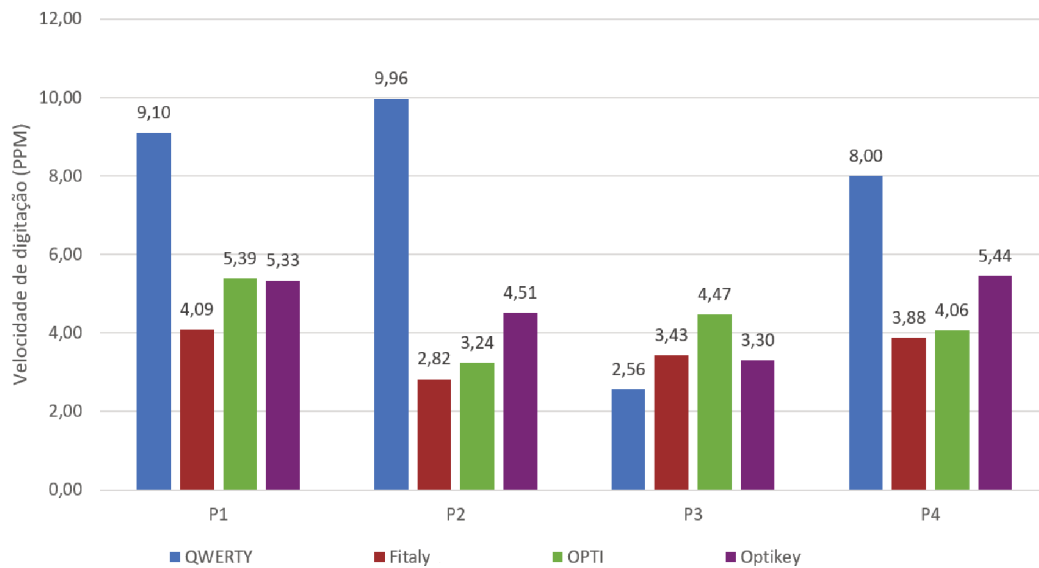
Foram apresentados indicadores de eficiência do algoritmo e da integração entre software e rastreador ocular, incluindo a taxa de reconhecimento preciso de gestos do olhar em palavras, a velocidade de digitação e quantidade de erros de digitação de indivíduos inexperientes com a interface.

Pode-se concluir que o método e algoritmo propostos nesta tese apresentam resultados comparáveis a soluções de referência na digitação ocular, mesmo em uma sessão inicial de treinamento, apresentando um potencial promissor de aplicação em sistemas de comunicação assistiva.

Após avaliação própria e considerações dos participantes dos experimentos, aponta-se o teclado com *layout* OPTI como o mais promissor para utilização utilizando-se gestos do olhar, após sessões de treinamento utilizando o *layout* pouco conhecido. O *layout* OPTI foi apontado como sendo mais confortável para digitação por gestos devido a seu formato mais próximo de um quadrado, teclas maiores e caracteres mais utilizados ao centro da interface.

Entretanto os usuários salientaram a dificuldade de habituação inicial com a utilização

Figura 61 – Análise comparativa da velocidade de digitação em diferentes configurações de tecla.

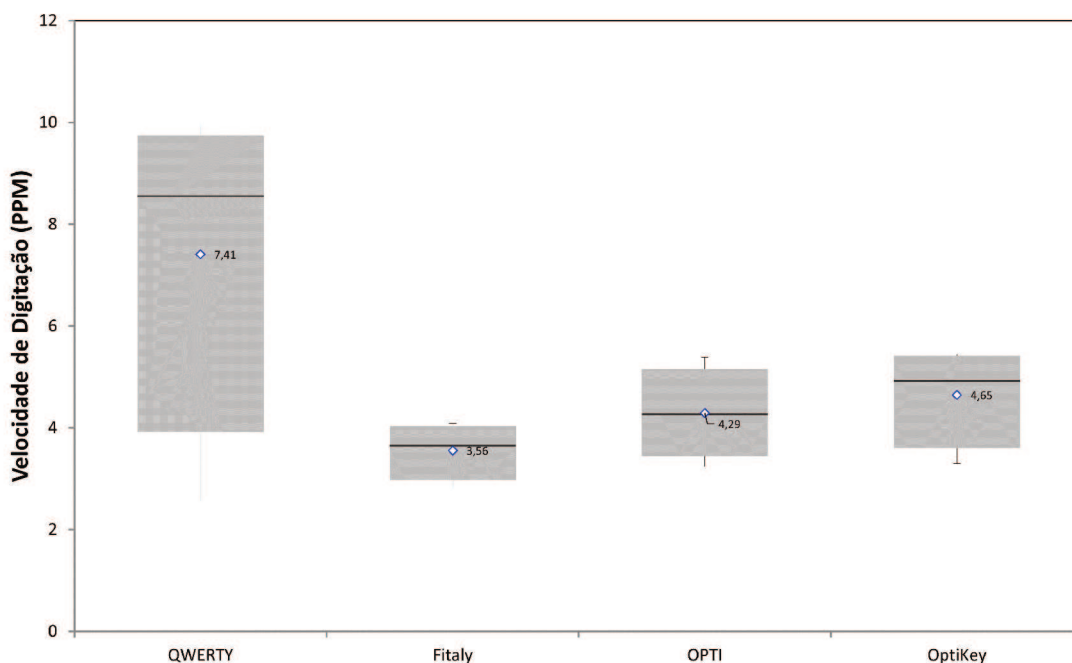


Fonte – O autor.

do olhar como controle, necessidade de atenção e foco constante, que tornam a utilização da interface exaustiva. Apesar de apontarem o potencial do *layout* OPTI, os participantes preferiram o *layout* QWERTY devido à familiaridade com o *layout* de teclas.

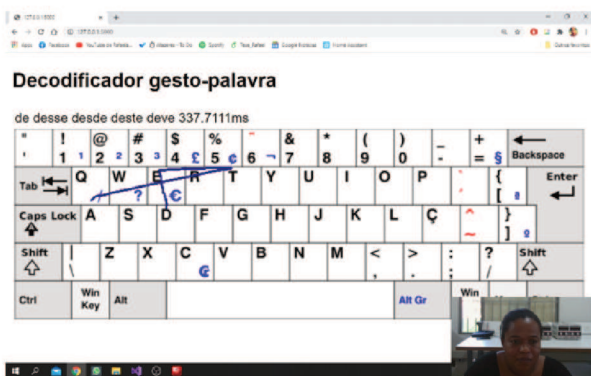
O próximo capítulo apresenta as considerações finais, contribuições desta tese e sugestões de trabalhos futuros.

Figura 62 – Estatística comparativa da velocidade de digitação para diferentes layouts de teclado.

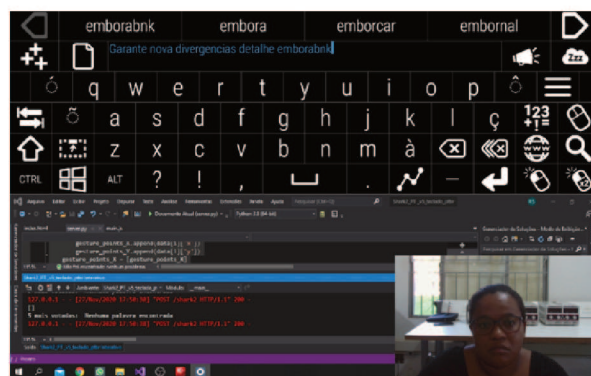


Fonte – O autor.

Figura 63 – Avaliação de utilização da interface por participante do experimento (a) sem óculos e (b) utilizando óculos.



(a) Sem óculos



(b) Com óculos

Fonte – O autor.

Considerações Finais

8.1 Introdução

Este capítulo apresenta uma visão geral sobre o que foi desenvolvido nesta tese, considerações sobre o cumprimento dos objetivos e possibilidades de trabalhos futuros que poderão ser desenvolvidos a partir desta tese.

8.2 Contribuições

Este trabalho propôs o desenvolvimento de um teclado virtual assistivo controlado por gestos do olhar para entrada de texto por pessoas com restrições severas de comunicação. A interface de comunicação assistiva foi desenvolvida a partir de módulos interconectados que consistem na interface de usuário contendo teclado ocular, algoritmo de decodificação de traçados do olhar em palavras, modelo linguístico do português brasileiro e captura de coordenadas da posição do olhar a partir de um rastreador ocular.

O capítulo introdutório desta tese apresenta o objetivo e hipótese central do trabalho: a possibilidade de melhoria na ergonomia e velocidade de digitação utilizando os olhos e a diminuição da carga cognitiva ao se inserir palavras utilizando gestos do olhar. A possibilidade foi vislumbrada como uma integração entre teclados oculares convencionais atualmente utilizados e teclados por gestos como os presentes em *smartphones*.

Foram apresentadas revisões de literatura acerca da Tecnologia Assistiva e Comunicação Aumentativa e Alternativa, do desenvolvimento de teclados virtuais, assim como da utilização da direção do olhar como fonte de informação. A incursão em áreas de estudo distintas porém correlatas foi necessário para o desenvolvimento de diferentes componentes da interface assistiva, consistindo em um desafio multicêntrico e complexo.

O desenvolvimento do teclado virtual controlado pelo olhar foi então descrito de forma modularizada, incluindo a interface embutida em uma página web leve, responsiva e multiplataforma, o dicionário inicial derivado de um corpus do português, o algoritmo de decodificação de traçados realizados pelo olhar sobre o teclado virtual, a integração

e configuração de um rastreador ocular genérico acessível no sistema de comunicação assistiva.

Foram realizados experimentos com o objetivo de aferir a efetividade do sistema desenvolvido e condições de usabilidade real da solução pretendida. O primeiro experimento demonstrou através de entradas de gesto sintéticas e também gestos reais produzidos pelo rastreador ocular que o algoritmo possui capacidade real de decodificação de gestos do teclado em palavras do léxico, obtendo taxas de reconhecimento acima de 90% para três diferentes layouts de teclado experimentados.

O segundo experimento apresentado verificou condições de usabilidade real do teclado ocular assistivo com usuários sem experiência com digitação ocular. Os resultados obtidos apontam velocidade de digitação variando de 3,56 a 7,41 palavras por minuto em condições de utilização sem treinamento e taxa média de reconhecimento correto de palavras de 83,75%.

Os resultados apresentados são promissores, sobretudo ao considerar que os usuários realizaram os testes em uma única sessão de utilização, sem treinamento prévio e familiaridade com configurações de teclas alternativas utilizadas. O teclado por gestos oculares está em sua primeira versão, podendo ser otimizado e adaptado a partir dos resultados quantitativos e impressões subjetivas concedidas pelos participantes dos experimentos.

8.3 Trabalhos Futuros

Os testes de usabilidade com usuários reais foram efetuados com indivíduos sem restrições de comunicação ou condições de mobilidade reduzida. É importante que os experimentos sejam realizados com indivíduos que apresentam restrições severas de comunicação como portadores da esclerose lateral amiotrófica.

A primeira versão do teclado assistivo foi testada com dicionários de tamanho máximo de 3.000 palavras. Trabalhos futuros podem explorar o acréscimo na quantidade de palavras do léxico e seu impacto na efetividade do algoritmo de decodificação de forma.

Pode-se explorar a ambiguidade de gestual em diferentes layouts de teclado para o português brasileiro, de forma a investigar a configuração de teclas mais eficiente e possivelmente desenvolver um layout otimizado para digitação gestual em português brasileiro.

A decodificação de gestos do olhar pode ser explorada a partir de outras técnicas, sobretudo relacionadas ao aprendizado de máquina. Métodos utilizando redes neurais recorrentes como a *Long short-term memory* (LSTM) ou empregando máquinas de vetores de suporte (SVM) já foram relatadas na literatura com resultados promissores.

O teclado com disposição de teclas matricial - em linhas e colunas - pode não ser o mais adequado para digitação por gestos. É sugerido o desenvolvimento e avaliação de um teclado em formato radial - com fileiras de teclas concêntricas - para examinar compa-

rativamente o desempenho do algoritmo de reconhecimento de gestos e as características de usabilidade.

Pode-se apontar ainda a possibilidade de desenvolvimento de outros elementos de navegação assistida além do teclado virtual, que de forma integrada a este possam propiciar conforto, acessibilidade e conexão com o mundo exterior a indivíduos com limitações severas de comunicação.

8.4 Contribuições em Produção Bibliográfica

Esta seção apresenta as publicações originadas do desenvolvimento desta pesquisa. O desenvolvimento do algoritmo de decodificação de gestos do olhar em um teclado virtual em palavras e sua integração em um sistema completo de digitação ocular foi descrito em um artigo para o periódico *IEEE Latin America Transactions*, ISSN 1548-0992. Os testes preliminares e análise de resultados foram apresentados em estudo publicado no ano de 2020.

- SILVA, Rafael Augusto; VEIGA, Antônio Cláudio Paschoarelli. Algorithm for decoding visual gestures for an assistive virtual keyboard. **IEEE Latin America Transactions**, Vol. 100 No. 1e (2020): Early Access, 2020. ISSN 1548-0992. Disponível em: <<https://latamt.ieeer9.org/index.php/transactions/article/view/3858>>.

Referências

ABRAMS, R. A.; MEYER, D. E.; KORNBLUM, S. Speed and Accuracy of Saccadic Eye Movements: Characteristics of Impulse Variability in the Oculomotor System. **Journal of Experimental Psychology: Human Perception and Performance**, 1989. ISSN 00961523.

AHSAN, H. et al. Vision Based Laser Controlled Keyboard System for the Disabled. **Proceedings of the 7th International Symposium on Visual Information Communication and Interaction**, p. 200:200–200:203, 2014. Disponível em: <http://doi.acm.org/10.1145/2636240.2636863>.

ALSHARIF, O. et al. Long short term memory neural network for keyboard gesture decoding. In: **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**. IEEE, 2015. v. 2015-Augus, p. 2076–2080. ISBN 978-1-4673-6997-8. ISSN 15206149. Disponível em: <http://ieeexplore.ieee.org/document/7178336>.

ANDERSON, D.; BAILEY, C.; SKUBIC, M. Hidden Markov Model Symbol Recognition for Sketch-Based Interfaces. **AAAI Fall Symposium**, p. 15–21, 2004. Disponível em: <https://www.aaai.org/Papers/Symposia/Fall/2004/FS-04-06/FS04-06-003.pdf>.

ANTONIO, N. S. et al. Design of a Brazilian Portuguese Virtual Keyboard for People with Severe Motor Disability. **Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems**, v. 1, p. 14–20, 2014. Disponível em: <http://dl.acm.org/citation.cfm?id=2738055.2738061>.

ASHA. **American Speech-Language-Hearing Association**. 2019. Disponível em: <https://www.asha.org>.

BAKER, B. Using images to generate speech. **Byte**, v. 11, n. 3, p. 160–168, 1986.

BALL, L. J.; FAGER, S.; FRIED-OKEN, M. Augmentative and alternative communication for people with progressive neuromuscular disease. **Physical Medicine and Rehabilitation Clinics of North America**, Elsevier Inc, v. 23, n. 3, p. 689–699, 2012. ISSN 10479651. Disponível em: <http://dx.doi.org/10.1016/j.pmr.2012.06.003>.

BEDELL, H. E.; STEVENSON, S. B. Eye movement testing in clinical examination. **Vision Research**, 2013. ISSN 00426989.

- BEE, N.; ANDRÉ, E. Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2008. ISBN 3540693688. ISSN 03029743.
- BELLMAN, R.; KALABA, R. On adaptive control processes. **IRE Transactions on Automatic Control**, 1959. ISSN 0096-199X.
- BERSCH, R. Introdução à Tecnologia Assistiva. **Porto Alegre: CEDI**, p. 21, 2008.
- BERSCH, R.; SARTORETTO, M. L. **O que é Tecnologia Assistiva?** 2019. Disponível em: <<http://www.assistiva.com.br/tassistiva.html>>.
- BERTAZZI, R. N. et al. Esclerose Lateral Amiotrófica. **Revista de Patologia do Tocantins**, v. 4, n. 3, p. 54–65, 2017. ISSN 2318-8960.
- BEUKELMAN, D.; MIRENDA, P. Augmentative and Alternative Communication. **Handbook of Research on Developments in E-Health and Telemedicine**, Brookes, n. January 2012, p. 1070–1087, 2012. Disponível em: <<http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-61520-670-4.ch051>>.
- BEUKELMAN, D. R. et al. AAC for adults with acquired neurological conditions: A review. **AAC: Augmentative and Alternative Communication**, v. 23, n. 3, p. 230–242, 2007. ISSN 07434618.
- BHATTACHARYA, S.; LAHA, S. Bengali text input interface design for mobile devices. **Universal Access in the Information Society**, 2013. ISSN 16155289.
- BI, X. et al. Octopus. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2013. p. 543–552. ISBN 9781450318990. Disponível em: <<https://dl.acm.org/doi/10.1145/2470654.2470732>>.
- BI, X.; ZHAI, S. IJQwerty. In: **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2016. p. 49–58. ISBN 9781450333627. Disponível em: <<https://dl.acm.org/doi/10.1145/2858036.2858421>>.
- BRASIL. Lei nº 13.005, de 25 de junho de 2014. Aprova o Plano Nacional de Educação-PNE e dá outras providências. **Diário Oficial da União**, v. 26, 2014. Disponível em: <<http://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>>.
- _____. LEI Nº 13.146, DE 6 DE JULHO DE 2015. **Diário da República**, 1ª série - nº 116, 2015.
- BULLING, A.; ROGGEN, D.; TRÖSTER, G. It's in your eyes: towards context-awareness and mobile HCI using wearable EOG goggles. . . . **of the 10th international conference on . . .**, 2008.
- CARPENTER, R. H. S. **Movements of the eyes**, 2nd rev. London, England: Pion Limited, 1988. 593 p. ISBN 0-85086-109-8 (Hardcover).
- CAT, C. d. A. T. Ata da Reunião VII, de dezembro de 2007 do Comitê de Ajudas Técnicas. **Secretaria Especial dos Direitos Humanos da Presidência da República (CORDE/SEDH/PR)**, 2007. Disponível em: <https://www.assistiva.com.br/Ata_VII_Reuni~ao_do_Comite_de_Ajudas_Técnicas.pdf>.

- CECOTTI, H. A Multimodal Gaze-Controlled Virtual Keyboard. **IEEE Transactions on Human-Machine Systems**, IEEE, v. 46, n. 4, p. 601–606, 2016. ISSN 21682291.
- CHALUPA, L. M.; WERNER, J. S. (Ed.). **The visual neurosciences, Vols. 1 & 2**. Cambridge, MA, US: MIT Press, 2004. xiv, 1694-xiv, 1694 p. ISBN 0-262-03308-9 (Hardcover).
- Chronos-Vision. **Scleral Search Coils 2D/3D**. 2019. Disponível em: <http://www.chronos-vision.de/downloads/CV_Product_SSC.pdf>.
- CLIFTON, C.; RAYNER, K. Eye Movements in Reading Words and Sentences Charles. v. 1, p. 1–61, 2015. ISSN 1098-6596.
- COGAIN. **Communication by Gaze Interaction Wiki**. 2019. Disponível em: <wiki.cogain.org>.
- COSTA, M. F. Movimentos oculares no bebê: o que eles nos indicam sobre o status oftalmológico e neurológico. **Psicologia USP**, 2017.
- COVRE, P. et al. Movimentos oculares e padrões de busca visual em tarefas de rotação mental. **Psico-USF**, v. 10, n. 1, p. 41–49, 2005. ISSN 1413-8271.
- DONAGHY, C. et al. **Eye movements in amyotrophic lateral sclerosis and its mimics: A review with illustrative cases**. 2011.
- DREWES, H. **Eye Gaze Tracking for Human Computer Interaction**. Tese (Doutorado) — LMU München, 2010. Disponível em: <<https://edoc.ub.uni-muenchen.de/11591/>>.
- DUCHOWSKI, A. T. **Eye Tracking Methodology**. Cham: Springer International Publishing, 2017. ISSN 1098-6596. ISBN 978-3-319-57881-1. Disponível em: <<http://link.springer.com/10.1007/978-3-319-57883-5>>.
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. **New York: John Wiley, Section**, 2001.
- ELMEZAIN, M.; AL-HAMADI, A.; MICHAELIS, B. Hand trajectory-based gesture spotting and recognition using HMM. In: **2009 16th IEEE International Conference on Image Processing (ICIP)**. IEEE, 2009. p. 3577–3580. ISBN 978-1-4244-5653-6. ISSN 15224880. Disponível em: <<http://ieeexplore.ieee.org/document/5414322/>>.
- EUSTAT, b. Educação em tecnologias de apoio para utilizadores finais: linhas de orientação para formadores. 1999. Disponível em: <<http://www.siva.it/research/eustat/eustgupt.html>>.
- FAN, X. et al. Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. **Journal of Autism and Developmental Disorders**, 2009. ISSN 01623257.
- FILHO, T. G. a Tecnologia Assistiva : De Que Se Trata ? **Porto Alegre: Redes Editora**, p. 207–235, 2009.
- FINATTO, M. J. B. Projeto PorPopular, frequência de verbos em português e no jornal popular brasileiro. **As Ciências do Léxico: lexicologia, lexicografia, terminologia**. Campo Grande: Ed. da UFMS, v. 6, p. 227–244, 2012.

- FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. **Journal of Experimental Psychology**, 1954. ISSN 00221015.
- GEGENFURTNER, A.; LEHTINEN, E.; SÄLJÖ, R. **Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains**. 2011.
- GIANNOTTO, E. C. Uso de rastreamento do olhar na avaliação da experiência do tele-usuário de aplicações de TV interativa. 2009. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3141/tde-15042009-151212/>>.
- GIBBONS, C.; BENETEAU, E. Functional Performance Using Eye Control and Single Switch Scanning by People With ALS. **Perspectives on Augmentative and Alternative Communication**, 2010. ISSN 1940-7475.
- GOLDBERG, D.; RICHARDSON, C. Touch-typing with a stylus. **Proceedings of the INTERACT'93 and CHI' . . .**, 1993.
- HAIN, T. C. **Eye Movement recording Devices**. 2012. Disponível em: <<https://www.dizziness-and-balance.com/practice/eyemove.html>>.
- HILL, K. Advances in Augmentative and Alternative Communication as Quality-of-Life Technology. **Physical Medicine and Rehabilitation Clinics of North America**, v. 21, n. 1, p. 43–58, 2010. ISSN 10479651.
- HOPPE, S.; LÖCHTEFELD, M.; DAIBER, F. Eype - Using eye-traces for eye-typing. In: **Workshop on grand challenges in text entry (chi 2013)**. [S.l.: s.n.], 2013.
- HORSTMANN, N.; AHLGRIMM, A.; GLÖCKNER, A. How Distinct are Intuition and Deliberation? An Eye-Tracking Analysis of Instruction-Induced Decision Modes. 2009.
- ISAAC. **International Society for Augmentative and Alternative Communication**. 2011. Disponível em: <<https://www.isaac-online.org/english/home/>>.
- ISO 9999:2016. **Assistive products for persons with disability – Classification and terminology**. 2016. 194 p. Disponível em: <<https://www.iso.org/standard/60547.html>>.
- ISOKOSKI, P. Text input methods for eye trackers using off-screen targets. In: **Proceedings of the symposium on Eye tracking research & applications - ETRA '00**. New York, New York, USA: ACM Press, 2000. p. 15–21. ISBN 1581132808. Disponível em: <<http://portal.acm.org/citation.cfm?doid=355017.355020>>.
- JACOB, R. J.; KARN, K. S. Eye Tracking in Human-Computer Interaction and Usability Research. Ready to Deliver the Promises. **The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research**, p. 531–553, 2003. ISSN 09248455.
- JEGERSKI, J.; VANPATTEN, B. **Research Methods in Second Language Psycholinguistics**. Taylor & Francis, 2013. (Second Language Acquisition Research Series). ISBN 9781136339141. Disponível em: <<https://books.google.com.br/books?id=jMdiAgAAQBAJ>>.

JOHNSON, R.; ZHANG, T. Supervised and Semi-Supervised Text Categorization using LSTM for Region Embeddings. v. 48, 2016. Disponível em: <<http://arxiv.org/abs/1602.02373>>.

JULIATO, M. et al. TFlex : Um Simulador de Teclado com Múltiplos Modos de Varredura To cite this version : HAL Id : hal-01093638 TFlex : Um Simulador de Teclado com Múltiplos Modos de Varredura. 2014.

Kang Wang; JI, Q. Real time eye gaze tracking with Kinect. In: **2016 23rd International Conference on Pattern Recognition (ICPR)**. IEEE, 2016. p. 2752–2757. ISBN 978-1-5090-4847-2. Disponível em: <<http://ieeexplore.ieee.org/document/7900052/>>.

KOWLER, E. Eye movements: The past 25years. **Vision Research**, Elsevier Ltd, v. 51, n. 13, p. 1457–1483, 7 2011. ISSN 00426989. Disponível em: <<http://dx.doi.org/10.1016/j.visres.2010.12.014>>.

KRISTENSSON, P.; VERTANEN, K. The potential of dwell-free eye-typing for fast assistive gaze communication. **Proceedings of the Symposium on Eye ...**, v. 1, n. 212, p. 241–244, 2012. Disponível em: <<http://dl.acm.org/citation.cfm?id=2168605>>.

KRISTENSSON, P. O. **Discrete and Continuous Shape Writing for Text Entry and Control**. 215 p. Tese (Doutorado) — Linköping University, 2007.

KRISTENSSON, P.-O.; WANG, J.; ZHAI, S. **System and Method for Recognizing Word Patterns in a Very Large Vocabulary Based on a Virtual Keyboard Layout**. 2010.

KRISTENSSON, P.-O.; ZHAI, S. SHARK: A Large Vocabulary Shorthand Writing System for Pen-Based Computers. **Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04**, v. 6, n. 2, p. 43–52, 2004. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1029632.1029640>>.

KUMAR, M.; PAEPCKE, A.; WINOGRAD, T. EyePoint: practical pointing and selection using gaze and keyboard. **Proceedings of the SIGCHI conference on Human factors in computing systems**, 2007.

KURAUCHI, A. et al. EyeSwipe: Dwell-free Text Entry Using Gaze Paths. **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**, p. 1952–1956, 2016. Disponível em: <<http://doi.acm.org/10.1145/2858036.2858335>>.

KURAUCHI, A. T. N. **EyeSwipe: text entry using gaze paths**. Tese (Doutorado) — Universidade de São Paulo, São Paulo, 7 2018. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45134/tde-03072018-151733/>>.

LANCIONI, G. E. et al. Three persons with multiple disabilities accessing environmental stimuli and asking for social contact through microswitch and VOCA technology. **Journal of Intellectual Disability Research**, 2008. ISSN 09642633.

LAND, M. F. Eye movements and the control of actions in everyday life. v. 25, p. 296–324, 2006.

- LEIGH, R. J.; KHANNA, S. Neuroscience of Eye Movements. **Advances in Clinical Neuroscience and Rehabilitation**, v. 5, n. 6, p. 4–6, 2006.
- LEIGH, R. J.; ZEE, D. S. **The Neurology of Eye Movements**. Oxford University Press, 2015. (Contemporary neurology series). ISBN 9780199969289. Disponível em: <<https://books.google.com.br/books?id=v2s0BwAAQBAJ>>.
- LEWIS, J. R.; KENNEDY, P. J.; LALOMIA, M. J. Development of a Digram-Based Typing Key Layout for Single-Finger/Stylus Input. **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, 1999. ISSN 1541-9312.
- LI, H. On-line and dynamic time warping for time series data mining. **International Journal of Machine Learning and Cybernetics**, v. 6, n. 1, p. 145–153, 2014. ISSN 1868808X.
- LIEGEL, L. A.; GOGOLA, M. M. R.; NOHAMA, P. Layout de teclado para uma prancha de comunicação alternativa e ampliada. **Revista Brasileira de Educação Especial**, v. 14, n. 3, p. 479–496, 2008. ISSN 1413-6538. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-65382008000300010&lng=pt&tlng=pt>.
- LIU, Y. et al. A robust recognition approach in eye-based dwell-free typing. **Proceedings of 2015 IEEE International Conference on Progress in Informatics and Computing, PIC 2015**, IEEE, p. 5–9, 2016.
- _____. GazeTry. In: **Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction on - OzCHI '15**. New York, New York, USA: ACM Press, 2015. p. 192–196. ISBN 9781450336734. Disponível em: <<http://doi.acm.org/10.1145/2838739.2838804http://dl.acm.org/citation.cfm?doid=2838739.2838804>>.
- LOJA, L. f. B. Tecnologia Assistiva: Um teclado virtual evolutivo para aplicação em sistemas de comunicação alternativa e aumentativa. 2015.
- MACKENZIE, I. S.; SOUKOREFF, R. W. Text Entry for Mobile Computing: Models and Methods, Theory and Practice. **Human-Computer Interaction**, Taylor & Francis, v. 17, n. 2-3, p. 147–198, 2002. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/07370024.2002.9667313>>.
- MACKENZIE, I. S.; ZHANG, S. X. The design and evaluation of a high-performance soft keyboard. **Conference on Human Factors in Computing Systems - Proceedings**, n. May, p. 25–31, 1999.
- MACKENZIE, I. S.; ZHANG, S. X.; SOUKOREFF, R. W. Text entry using soft keyboards. **Behaviour and Information Technology**, v. 18, n. 4, p. 235–244, 1999. ISSN 0144929X.
- MACKENZIE, S.; READ, J. C. Using paper mockups for evaluating soft keyboard layouts. **Proceedings of the 2007 Conference of the Center for Advanced Studies on Collaborative Research, CASCON '07**, n. January 2007, p. 98–108, 2007.

- MAJARANTA, P.; AHOLA, U.-K.; ŠPAKOV, O. Fast gaze typing with an adjustable dwell time. **Proceedings of the 27th international conference on Human factors in computing systems - CHI 09**, p. 357, 2009. Disponível em: <<http://dl.acm.org/citation.cfm?id=1518701.1518758>>.
- MAJARANTA, P. et al. Effects of feedback and dwell time on eye typing speed and accuracy. **Universal Access in the Information Society**, v. 5, n. 2, p. 199–208, 2006. ISSN 1615-5289. Disponível em: <<http://link.springer.com/10.1007/s10209-006-0034-z>>.
- MAJARANTA, P.; RÄIHÄ, K.-J. Twenty Years of Eye Typing: Systems and Design Issues. **Eye Tracking Research & Applications (ETRA) Symposium**, p. 15–22, 2002.
- MALMIVUO, J.; PLONSEY, R. **Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields**. [S.l.: s.n.], 2012. ISBN 9780199847839.
- MANKOFF, J.; ABOWD, G. G. D. Cirrin: A word-level unistroke keyboard for pen input. In: **Proceedings of the 11th annual ACM symposium on User interface software and technology - UIST '98**. [S.l.: s.n.], 1998. ISBN 1581130341.
- MANZINI, E. J. Tecnologia assistiva para educação: recursos pedagógicos adaptados. **Ensaio pedagógico: construindo escolas inclusivas**. Brasília: SEESP/MEC, p. 82–86, 2005.
- MARKUSSEN, A. Vulture : A Mid-Air Word-Gesture Keyboard. p. 1073–1082, 2014.
- MELCHIADES, A. M. **PARÂMETROS OCULARES NO RASTREAMENTO VISUAL DE CENAS COM CONTEÚDO EMOCIONAL**. Tese (Doutorado) — Universidade de Brasília, 2014.
- Michaelis. **Moderno dicionário da língua portuguesa**. Melhoramentos, 1999. ISBN 978-8506027592. Disponível em: <michaelis.uol.com.br/>.
- MILEKIC, S. The More You Look the More You Get: Intention-based Interface using Gaze-tracking. In: **Proceedings of Museums and the Web**. [S.l.: s.n.], 2003.
- MINIOTAS, D.; SPAKOV, O.; EVREINOV, G. Symbol creator: An alternative eye-based text entry technique with low demand for screen space. **Proceedings of INTERACT**, 2003.
- MINIOTAS, D.; ŠPAKOV, O.; MACKENZIE, I. S. Eye gaze interaction with expanding targets. In: . [S.l.: s.n.], 2004. ISBN 1581137036.
- MIRENDA, P. Autism, Augmentative Communication, and Assistive Technology: What Do We Really Know? **Focus on Autism and Other Developmental Disabilities**, 2001. ISSN 10883576.
- OUYANG, T. et al. Mobile Keyboard Input Decoding with Finite-State Transducers. 2017. Disponível em: <<http://arxiv.org/abs/1704.03987>>.

- PEDROSA, D.; PIMENTEL, M. d. G.; TRUONG, K. N. Filteryedping: A Dwell-Free Eye Typing Technique. **Extended Abstracts of the ACM CHI'15 Conference on Human Factors in Computing Systems**, v. 2, p. 303–306, 2015. Disponível em: <http://dx.doi.org/10.1145/2702613.2725458>.
- PEDROSA, D. D. C. **Data input and content exploration in scenarios with restrictions**. Tese (Doutorado) — Universidade de São Paulo, São Carlos, 12 2015. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-13042015-144651/>.
- PI, J.; SHI, B. E. Probabilistic adjustment of dwell time for eye typing. **Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017**, p. 251–257, 2017. ISSN 2301380X.
- PILZER, J.; LIU, S. Detecting Self-Interruptions during Reading. 2017.
- PLATZ, M. S.; ONG, J. K. Y.; HASLWANTER, T. Optimizing Video-oculography Systems by Simulating The Effect of Slippage Artifacts. In: . [S.l.: s.n.], 2009.
- PORTA, M.; TURINA, M. Eye-S: a full-screen input modality for pure eye-based communication. **Proceedings of the 2008 Symposium on Eye Tracking Research & Applications**, 2008.
- QUAID, P.; SIMPSON, T. Association between reading speed, cycloplegic refractive error, and oculomotor function in reading disabled children versus controls. **Graefe's Archive for Clinical and Experimental Ophthalmology**, 2013. ISSN 0721832X.
- QVARFORDT, P.; ZHAI, S. Conversing with the user based on eye-gaze patterns. In: . [S.l.: s.n.], 2005.
- RICK, J. Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop. **UIST 2010 - 23rd ACM Symposium on User Interface Software and Technology**, p. 77–86, 2010.
- RÖSLER, A. Using the Tobii Mobile Device - Stand in Usability Testing on Mobile Devices. 2012.
- SALVADOR, S.; CHAN, P. Toward accurate dynamic time warping in linear time and space. **Intelligent Data Analysis**, v. 11, n. 5, p. 561–580, 2007. ISSN 1088467X.
- SALVUCCI, D. D.; GOLDBERG, J. H. Identifying fixations and saccades in eye-tracking protocols. **Proceedings of the symposium on Eye tracking research & applications - ETRA '00**, p. 71–78, 2000. ISSN 10960384. Disponível em: <http://portal.acm.org/citation.cfm?doid=355017.355028>.
- SARCAR, S. et al. Virtual keyboard design: State of the arts and research issues. **TechSym 2010 - Proceedings of the 2010 IEEE Students' Technology Symposium**, IEEE, n. April, p. 289–299, 2010.
- SARCAR, S.; PANWAR, P.; CHAKRABORTY, T. EyeK. In: **Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction - APCHI '13**. New York, New York, USA: ACM Press, 2013. p. 215–220. ISBN 9781450322539. Disponível em: <http://dl.acm.org/citation.cfm?doid=2525194.2525288>.

- SENIN, P. Dynamic Time Warping Algorithm Review. **Science**, v. 2007, n. December, p. 1–23, 2008. ISSN 1557170X. Disponível em: <https://www.researchgate.net/publication/228785661_Dynamic_Time_Warping_Algorithm_Review>.
- SILVA, M. A. d. **Info Escola - Globo Ocular**. 2012. Disponível em: <<https://www.infoescola.com/visao/globo-ocular/>>.
- SILVA, R. **Desenvolvimento de um sistema embarcado para pupilometria**. 2016. 113 p. Disponível em: <<https://repositorio.ufu.br/handle/123456789/14621>>.
- SMITH, B. A.; BI, X.; ZHAI, S. Optimizing Touchscreen Keyboards for Gesture Typing. In: **Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15**. New York, New York, USA: ACM Press, 2015. p. 3365–3374. ISBN 9781450331456. ISSN 0192-513X. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2702123.2702357>>.
- SMITH, E.; DELARGY, M. Clinical review Locked-in syndrome. v. 330, n. February, p. 3–6, 2005.
- SOTO GLORIA; ZANGARI, C. **Practically speaking: Language, literacy, and academic development for students with AAC needs**. [S.l.]: Paul H. Brookes, 2009.
- STENDAHL, J.; ARNÖR, J. Gesture Keyboard Using Machine Learning. 2014.
- SWEETLAND, J. **OptiKey Wiki**. 2015. Disponível em: <<https://github.com/OptiKey/OptiKey/wiki>>.
- TANG, L. **Subtleties of Eyetracking Heat Maps and Gaze Plots**. 2016. Disponível em: <<https://medium.com/@TheRealTang/subtleties-of-eyetracking-heat-maps-and-gaze-plots-a7ba4207f20f>>.
- TAPPERT, C. C. Cursive Script Recognition by Elastic iy/latcling. v. 26, n. 6, p. 765–771, 1982.
- Tobii. **Tobii Eye Tracker 4C**. 2016. Disponível em: <<https://gaming.tobii.com/product/tobii-eye-tracker-4c>>.
- TORTORA, G. J. **Corpo Humano: Fundamentos de Anatomia e Fisiologia**. [S.l.: s.n.], 2003. ISSN 1748-7161. ISBN 9780870707339.
- TSAI, M. J. et al. Visual attention for solving multiple-choice science problem: An eye-tracking analysis. **Computers and Education**, 2012. ISSN 03601315.
- WADDINGTON, C. T. et al. Comparing a scanning ambiguous keyboard to the on-screen QWERTY keyboard. **HCI 2017: Digital Make Believe - Proceedings of the 31st International BCS Human Computer Interaction Conference, HCI 2017**, v. 2017-July, n. 1985, p. 1–6, 2017.
- WANG, Q. et al. An eye-tracking study of website complexity from cognitive load perspective. **Decision Support Systems**, 2014. ISSN 01679236.
- WARD, D. J.; MACKAY, D. J. C. Fast Hands-free Writing by Gaze Direction. **Nature**, v. 418, n. August, 8 2002. ISSN 0254-1106. Disponível em: <<http://arxiv.org/abs/cs/0204030>>.

- WATANABE, M. V. H. Eye Tracking E Suas Aplicações. p. 57, 2013.
- Wikipédia. **Esforço cognitivo** — Wikipédia, a enciclopédia livre. 2016. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Esforço_cognitivo&oldid=46264167>.
- WOBBROCK, J. O.; MYERS, B. A.; KEMBEL, J. A. EdgeWrite. In: **Proceedings of the 16th annual ACM symposium on User interface software and technology - UIST '03**. New York, New York, USA: ACM Press, 2003. v. 5, n. 2, p. 61–70. ISBN 1581136366. Disponível em: <<http://portal.acm.org/citation.cfm?doid=964696.964703>>.
- WOBBROCK, J. O.; RUBINSTEIN, J. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In: **Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08**. New York, New York, USA: ACM Press, 2008. p. 11. ISBN 9781595939821. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1344471.1344475>>.
- WOBBROCK, J. O. et al. Not Typing but Writing: Eye-based Text Entry Using Letter-like Gestures. **The 3rd Conference on Communication by Gaze Interaction**, n. Figure 1, p. 4–6, 2007.
- YANG, N.; MALI, A. D. Modifying Keyboard Layout to Reduce Finger-Travel Distance. In: **2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)**. IEEE, 2016. p. 165–168. ISBN 978-1-5090-4459-7. Disponível em: <<http://ieeexplore.ieee.org/document/7814594/>>.
- YARBUS, A. L. **Eye Movements and Vision**. Boston, MA: Springer US, 1967. v. 6. 389–390 p. ISSN 00283932. ISBN 978-1-4899-5381-0. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/0028393268900122>>.
- ZANGARI, C.; KANGAS, K.; LLOYD, L. Augmentative and alternative communication: A field in transition. **Augmentative and Alternative Communication**, Taylor & Francis, v. 4, n. 2, p. 60–65, 1 1988. ISSN 0743-4618. Disponível em: <<https://doi.org/10.1080/07434618812331274647>>.
- ZHAI, S.; HUNTER, M. Performance optimization of virtual keyboards. **Human-Computer Interaction**, v. 17, n. 2-3, p. 229–269, 2002. ISSN 07370024.
- ZHAI, S.; HUNTER, M.; SMITH, B. A. The metropolis keyboard - An exploration of quantitative techniques for virtual keyboard design. In: **UIST (User Interface Software and Technology): Proceedings of the ACM Symposium**. [S.l.: s.n.], 2000.
- ZHAI, S.; KRISTENSSON, P.-O. Shorthand writing on stylus keyboard. **Proceedings of the conference on Human factors in computing systems - CHI '03**, n. 5, p. 97, 2003. Disponível em: <<http://portal.acm.org/citation.cfm?doid=642611.642630>>.
- ZHAI, S.; KRISTENSSON, P. O. Introduction to Shape Writing. **Text Entry Systems**, n. September, p. 139–158, 2007.
- _____. Introduction to Shape Writing. **Text Entry Systems**, n. September, p. 139–158, 2007.

_____. The word-gesture keyboard. **Communications of the ACM**, v. 55, n. 9, p. 91–101, 9 2012. ISSN 0001-0782. Disponível em: <<https://dl.acm.org/doi/10.1145/2330667.2330689>>.

ZOETEN, R. d. et al. Recognizing input for swipe based keyboards. 2013.