

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

DIEGO DA SILVA LIMA

AVALIAÇÃO DA FUNÇÃO VISUAL INFANTIL A PARTIR DE SOLUÇÃO AUTOMATIZADA
DE RASTREAMENTO OCULAR BASEADA EM VÍDEO

São Paulo

2023

DIEGO DA SILVA LIMA

Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo

Versão original

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo, como requisito para obtenção do grau de Doutor em Ciências. Área de concentração: Neurociências e Comportamento
Orientadora: Dora Fix Ventura

São Paulo

2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE
TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO,
PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na publicação
Biblioteca Dante Moreira Leite
Instituto de Psicologia da Universidade de São Paulo
Dados fornecidos pelo(a) autor(a)

Lima, Diego da Silva

Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo / Diego da Silva Lima; orientador Dora Fix Ventura. -- São Paulo, 2023.

120 f.

Tese (Doutorado - Programa de Pós-Graduação em Neurociências e Comportamento) -- Instituto de Psicologia, Universidade de São Paulo, 2023.

1. Desenvolvimento perceptivo. 2. Psicofísica. 3. Visão. I. Fix Ventura, Dora , orient. II. Título.

Nome: Lima, Diego da Silva

Título: Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo, como requisito para obtenção do grau de Doutor em Ciências.

Aprovado em: _____

Banca examinadora

Prof. Dr. _____

Instituição _____

Julgamento _____

Prof. Dr. _____

Instituição _____

Julgamento _____

Prof. Dr. _____

Instituição _____

Julgamento _____

Prof. Dr. _____

Instituição _____

Julgamento _____

Prof. Dr. _____

Instituição _____

Julgamento _____

Dedicatória

Dedico este trabalho à minha esposa, Karollinny Morais Silveira, pelo suporte durante o desafiador processo do desenvolvimento do presente trabalho. Esse suporte se mostrou essencial dadas as difíceis circunstâncias da pandemia de Covid-19, que penso terem sido experienciadas por muitos pesquisadores. Espero poder, no futuro, retribuir a atenção, cuidado, paciência e compreensão que tão prontamente mostrou durante os últimos quatro anos.

Agradecimentos

Agradeço inicialmente à Prof^a. Dora Ventura, pela disposição e abertura para inclusão deste projeto no programa de pesquisa do Labvis-IPUSP e por todo o apoio durante sua execução. Agradeço ainda ao ao Prof. Dr. Saulo Duarte Passos, pela colaboração de muitos anos e por abrir a porta do Ambulatório de Pediatria do HU-Jundiaí para coleta de dados para o presente projeto. Estendo o agradecimento ao ao restante da equipe do projeto Zika-Jundiaí, principalmente à Adriana pela atenção e empenho ao ao encaminhar os pacientes. Agradeço ainda ao colega de laboratório Msc. Leonardo Aparecido Silva pela parceria durante a coleta de dados, e a todas as mães e crianças realizando acompanhamento no ambulatório que se dispuseram a participar da pesquisa.

Apoio financeiro

O presente trabalho foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) através dos programas Projeto Temático (números de processo 2014/26818-2 e e 2022/00191-0) e Bolsas de Doutorado de Fluxo Contínuo (número de processo 2019/12973-0).

Resumo

Lima, D. S. (2023). Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo (Tese de Doutorado). Instituto de Psicologia, Universidade de São Paulo, São Paulo.

A avaliação visual infantil é geralmente feita através da técnica de olhar preferencial (PL), na qual um observador treinado registra fixações do olhar da criança a estímulos variando em alguma dimensão de interesse como contraste ou frequência espacial. Apesar da popularidade da técnica, ela apresenta algumas limitações, como a impossibilidade de tornar o observador cego à condição testada, a variabilidade entre exames realizados por diferentes observadores, e a falta de controle preciso sobre parâmetros do estímulo. O objetivo principal do presente trabalho é desenvolver uma solução de software para avaliação da acuidade visual (AV) de grades utilizando vídeo-oculografia. Crianças entre 3-36 meses passaram por avaliação da AV pelo procedimento proposto e pelo procedimento clássico de olhar preferencial dos Cartões de Acuidade de Teller (TAC-II). O paciente, acomodado no colo da mãe ou responsável, observa os estímulos no monitor de vídeo enquanto uma câmera registra seus movimentos oculares em tempo real. A tarefa proposta consiste na apresentação de um padrão gradeado de frequência espacial variável que aparece contra um fundo cinza, do lado esquerdo ou direito da tela, aleatoriamente, durante um segundo. Um total de 38 pacientes foram examinados com sucesso pelos dois procedimentos. O resultado do procedimento proposto se mostrou positivamente relacionado com o resultado do procedimento TAC-II ($\beta = 0,25$; $p = 0,02$). Uma relação positiva entre a idade do participante e o valor de AV medido pelo teste de vídeo-oculografia também foi constatada ($\beta = 0,018$ decremento na medida em logMAR por mês de idade; $p < 0.001$). A diferença média do valor de AV medido pelos testes, entretanto, não foi nula, mas de 0,38 logMAR (IC = 0,27-0,48 logMAR; $t=7,11$; $p < 0.0001$). Diferenças metodológicas como tempo de apresentação do estímulo e engajamento da criança com a tarefa podem explicar as discrepâncias. Apesar destas diferenças, foram encontradas relações positivas entre a idade dos participantes e o valor de AV medido, bem como entre os valores medidos com as duas técnicas, o que sugere que o teste consegue capturar a dinâmica de desenvolvimento da AV.

Abstract

Lima, D. S. (2023). Assessment of infant visual function with an automated video-based eye tracking solution (Doctoral Thesis). Instituto de Psicologia, Universidade de São Paulo, São Paulo.

Infant visual assessment is usually performed through the preferential looking technique (PL), in which a trained observer records children's fixations to stimuli varying in some relevant dimension such as contrast or spatial frequency. Despite the technique's popularity, it presents some limitations, such as not being possible to use blind observers, the variability between examinations from different observers, and the lack of precise control over parameters in stimulus presentation. The main objective of the present work is to develop a software solution to assess grating visual acuity (VA) using video-oculography. Children aging between 3-36 months-old underwent evaluation of their VA through the proposed procedure and the classical Teller Acuity Cards (TAC-II) preferential looking procedure. The proposed task consists of the presentation of a grating pattern of variable spatial frequency against a gray background at the left or right side of the screen, randomly, during one second. A total of 38 patients were evaluated successfully with both procedures. The result of the proposed procedure was positively correlated with the TAC-II procedure ($\beta = 0.25$; $p = 0.02$). A positive relationship between the patient's age and the measured acuity value with the eye tracking test was also observed ($\beta = 0.018$ decrease in logMAR by month; $p < 0.001$). The mean difference in acuity values measured by the procedures wasn't zero, but of 0.38 logMAR (CI = 0.27-0.48 logMAR; $t=7.11$; $p < 0.0001$). Methodological differences such as stimulus presentation time and child engagement with the task might explain the discrepancies. Despite those differences, positive relationships between patient age and measured VA value, as well as between the values measured by both techniques, suggest the test is able to capture VA development dynamics.

Lista de tabelas

1	Resumo dos estudos dedicados à aplicação da vídeo-oculografia para estudo das funções visuais a partir do olhar preferencial.	40
2	Valores de acuidade medidos pela técnica de vídeo oculografia (VOG) e Cartões de Acuidade de Teller	59

Lista de figuras

1	Técnicas de vídeo-oculografia baseadas em modelagem tridimensional	32
2	Esquema da aplicação dos cartões de Acuidade de Teller(TAC-II)	48
3	Esquema da situação de avaliação com a vídeo-oculografia	50
4	Montagem dos equipamentos para avaliação no ambulatório de pediatria do HU-Jundiaí	54
5	Comparação dos valores de acuidade visual obtidos com as duas técnicas . .	60
6	Comparação dos valores de acuidade visual obtidos com as duas técnicas . .	61
7	Mudanças na medida do valor de acuidade pela análise automatizada (APL) com o critério de duração da fixação no estímulo de teste.	62
8	Curva ROC resultante da variação do critério de detecção (duração da fixação na região de interesse, em milissegundos) por todo o intervalo de ≥ 0 ms a ≥ 1500 ms.	64
9	Comparação dos valores de acuidade medidos pelo critério de detecção automatizado (APL) versus a detecção por inspeção visual do registro de vídeo da sessão (VPL).	65
10	Uma arquitetura de aprendizado profundo para extração de pupilas	112
11	Processo de treino da rede	113
12	Diagrama de componentes do sistema implementado para aquisição de análise dos registros de rastreamento ocular	117
13	Máquina de estados representando a <i>pipeline</i> de aquisição	118
14	Exemplo de tela da aplicação "Attend" de rastreamento ocular	119

Lista de abreviaturas

AV	Acuidade visual
cd/m ²	Candelas por metro quadrado
LED	Diodo emissor de luz
PL	Olhar preferencial
APL	Olhar preferencial automatizado
VPL	Olhar preferencial por vídeo
PVCP	Potencial Visual Cortical Provocado
TAC-II	Cartões de Acuidade de Teller
VOG	Vídeo-oculografia
V1	Área visual primária
cpcm	Ciclos por Centímetro
cpg	Ciclos por Grau de Ângulo Visual
logMAR	Logaritmo do Ângulo Mínimo de Resolução
IC	Intervalo de Confiança
P-CR	Vetor pupila-reflexo corneal
GenICam	<i>Generic Interface for Cameras</i>
A'	Estatística de área sob a curva.
ROC	Característica de Operação do Receptor

Sumário

1	Introdução	12
1.1	Avaliação visual infantil	12
1.2	Desenvolvimento do sistema visual	18
1.3	Desenvolvimento do sistema oculomotor	20
1.3.1	Sacadas e fixações	21
1.3.2	Reflexo de acomodação e movimento de vergência	24
1.3.3	Perseguição lenta	25
1.3.4	Reflexo vestibulo-ocular e nistagmo optocinético	26
1.4	A técnica de vídeo-oculografia	27
1.5	Uso clínico da vídeo-oculografia	33
1.5.1	Aplicações clínicas da técnica de vídeo-oculografia em adultos	33
1.5.2	Utilização da vídeo-oculografia para tarefas de olhar preferencial	34
2	Justificativa	39
3	Objetivos	43
3.1	Objetivo Geral	43
3.2	Objetivos específicos	43
4	Metodologia	44
4.1	População	44
4.2	Materiais	45
4.3	Procedimento	46
4.3.1	Cartões de Acuidade de Teller (TAC-II)	46
4.3.2	Tarefa de olhar preferencial automatizado (APL)	47
4.4	Análise de dados	54
5	Resultados	57

6	Discussão	66
6.1	Aplicabilidade do programa para testagem da acuidade visual	66
6.2	Comparação entre os valores de acuidade medidos pelas duas técnicas	67
6.3	Testabilidade	69
6.4	Limitações	71
6.5	Relação dos resultados com o critério de detecção do estímulo	73
6.6	Variações procedimentais na técnica de rastreamento ocular	77
6.7	Engajamento do participante com a tarefa	83
6.8	Desafios para a aplicabilidade clínica	85
7	Conclusão	90
8	Considerações finais	92
	Referências	93
	ANEXO A - Termo de Consentimento Livre e Esclarecido	106
	APÊNDICE A - Técnica de segmentação de pupila baseada em aprendizagem profunda	110
	APÊNDICE B - Especificações do software de vídeo-oculografia	116

1 Introdução

1.1 Avaliação visual infantil

Perdas da função visual nos primeiros anos de vida são em grande parte passíveis de prevenção se descobertas precocemente, e quando não podem prevenidas ou tratadas, podem ter seus prejuízos minimizados por tratamentos ou soluções de reabilitação (Fulton & Mayer, 1988; Maurer, 2017; Sjostrand & Abrahamsson, 2022). As condições que podem levar a estas perdas incluem a retinopatia de prematuridade, o glaucoma infantil, as cataratas congênitas, as infecções pré ou perinatais, as malformações e distrofias retinianas de origem genética e o retinoblastoma (Gogate, Gilbert, & Zin, 2011). Esses quadros podem levar desde perdas sutis, detectáveis apenas em exames especializados e que possuem impacto limitado na vida do paciente, até perdas mais severas, que podem evoluir para condições de baixa visão ou cegueira, impactando severamente o desenvolvimento da criança. Além das condições médicas citadas, problemas refrativos ou de alinhamento e motilidade ocular sem causa médica definida podem impactar severamente o desenvolvimento do sistema visual (Holmes & Clarke, 2006).

Disfunção visual cortical é a denominação que se dá às perdas visuais causadas por algum insulto mecânico, químico ou biológico ao sistema nervoso central, com demonstrado prejuízo às regiões dedicadas ao processamento da informação visual, notadamente à porção occipital do córtex, onde estão situadas as regiões visuais primárias e secundárias (Chang & Borchert, 2020). Causas subjacentes da disfunção visual cortical incluem: encefalopatia hipóxico-isquêmica (geralmente relacionada à prematuridade), convulsões, hidrocefalia, trauma e infecções congênitas do grupo TORCH (toxoplasmose, rubéola, citomegalovírus, herpes) (Chang & Borchert, 2020), além do mais recentemente descoberto Zika Vírus (de Oliveira Dias et al., 2018; Zin et al., 2018; Portnoi Baran et al., 2019; D. d. S. Lima et al., 2020). O diagnóstico de disfunção visual cortical se baseia na baixa acuidade visual (AV) binocular ou em perdas de campo visual que não podem ser atribuídas à patologia ocular (Chang & Borchert, 2020).

A ambliopia, por sua vez, é um conjunto de perdas visuais decorrentes de prejuízos no desenvolvimento das vias neurais de formação de imagem. Como esse desenvolvimento é dependente da estimulação adequada (Maurer, 2017), qualquer condição médica que

impeça a formação de imagem na retina, gere uma imagem borrada ou com características ópticas distintas em cada uma das retinas, têm potencial ambliogênico. A ambliopia é classificada, de acordo com Holmes & Clarke (2006), em: estrabísmica (devido ao desalinhamento dos eixos ópticos dos olhos, que não convergem para a formação adequada da imagem binocular, gerando alta disparidade entre as imagens formadas em cada uma das retinas); anisométrica (diferenças refrativas entre os olhos levam à condição em que uma das imagens formadas é relativamente mais borrada do que a outra), ou de privação, que ocorre quando há oclusão causada por condições médicas como catarata ou ptose (disfunção muscular que prejudica a sustentação da pálpebra), ou ainda quando há erro refrativo binocular não corrigido (Holmes & Clarke, 2006). Estimativas de prevalência de ambliopia na população geral giram em torno de 1-5% (Holmes & Clarke, 2006). Seu diagnóstico se dá pela detecção de AV baixa ou perdas de sensibilidade ao contraste, acrescido de diagnóstico de pelo menos um dos fatores ambliogênicos citados acima, sempre considerando a correção refrativa mais adequada para o paciente. Comumente, a ambliopia é diagnosticada por exclusão, quando algum grau de perda visual residual persiste mesmo na ausência de qualquer fator orgânico que a explique (LaRoche, 2011).

Tanto na disfunção visual cortical quanto na ambliopia, a avaliação visual psicofísica é componente crítico no diagnóstico e acompanhamento do quadro clínico. A acuidade visual (limiar de discriminação de detalhes) é a função mais comumente testada em contexto clínico. É a partir da caracterização quantitativa desta função que são delimitadas as condições de visão normal, baixa visão e cegueira legal (LaRoche, 2011), apesar da sensibilidade ao contraste (limiar de discriminação de variação de luminância) oferecer capacidade preditiva do desempenho do paciente em tarefas do dia a dia que é equiparável ou até melhor que a AV (Bennett, Bex, Bauer, & Merabet, 2019; Owsley & Sloane, 1987). As funções visuais básicas como AV e sensibilidade ao contraste, apesar de serem fundamentais para a caracterização das capacidades funcionais de um paciente por oferecerem uma medida quantitativa desta, não são a única fonte de informação sobre a condição visual de um paciente. Bennett et al. (2019) distinguem "função visual" (determinada por testagem psicofísica para detecção de limiares) de "visão funcional" (desempenho do paciente em desempenhar tarefas de seu dia a dia). A testagem psicofísica pode demonstradamente prever o desempenho em tarefas de visão funcional

(Owsley & Sloane, 1987); e o conhecimento desta, por sua vez, pode sugerir qual das dimensões da função visual e quais técnicas de testagem são adequadas para cada paciente. A importância de testar a visão funcional está no fato de as tarefas aplicadas avaliarem não somente as funções básicas (acuidade visual, sensibilidade ao contraste e discriminação de cor), como também as funções integrativas (movimento, profundidade, textura, hiperacuidade, agrupamento e forma), mostrando o quanto o desenvolvimento dessas dimensões mais complexas pode estar afetado em cada quadro clínico. A caracterização da natureza, grau e impacto das perdas visuais na infância, bem como de sua progressão ao longo do tempo, é requisito básico para a tomada de decisão sobre o início e a condução de quaisquer intervenções que se mostrem necessárias. Essa caracterização requer técnicas objetivas e rigorosas de avaliação, mas que não deixem de ser aplicáveis em contextos clínicos variados.

Quando o paciente já atingiu idade em que pode colaborar comunicando-se com o avaliador, mas ainda não foi alfabetizado (na faixa etária dos 2-5 anos), testes de AV baseados no emparelhamento de símbolos podem ser empregados. Estes testes não exigem nomeação dos símbolos, mas exigem apenas que a criança indique um dos símbolos de um conjunto apresentado que seja igual a um modelo. O exame de emparelhamento com os optótipos HOTV ou Símbolos de LEA (The Vision in Preschoolers Study Group, 2004; Holmes & Clarke, 2006) são comumente utilizados para avaliação da AV infantil. Acima dos 5 anos, para pacientes que já passaram pela fase inicial do processo de alfabetização e já são capazes de nomear letras, testes de optótipos que usam essa habilidade (como os que empregam as tabelas de optótipos de Snellen ou logMAR) já podem ser utilizados (Holmes & Clarke, 2006).

Já para a testagem de crianças pré-verbais, duas técnicas de avaliação são comumente utilizadas: a técnica eletrofisiológica de Potencial Visual Cortical Provocado (PVCP) e a técnica psicofísica do olhar preferencial (PL) (Dobson & Teller, 1978). Ambas as técnicas foram desenvolvidas em paralelo durante as décadas de 1960 e 1980, e exploram aspectos da função visual ligeiramente diferentes. A técnica de PVCP envolve a medida de potenciais elétricos detectáveis de forma não-invasiva quando o paciente observa um padrão oscilando em um monitor. A frequência temporal geralmente usada é em torno de 6 Hz. Ao variar alguma propriedade do padrão, é possível estimar um limiar de detecção (Norcia &

Tyler, 1985; Norcia, Tyler, Hamer, & Wesemann, 1989).

A técnica de potencial visual provocado foi rapidamente adotada para avaliação clínica, e demonstrou-se que sua testabilidade (proporção de pacientes com os quais a técnica é bem-sucedida) com população pré-verbal ou com dificuldades cognitivas e comunicativas resultantes de condições neurológicas é equiparável a outras técnicas de avaliação da AV (Odom & Green, 1984). Tanto a técnica de potencial visual provocado quanto a técnica de olhar preferencial permitem a identificação de tendências de melhora significativas nos valores de AV relacionados ao aumento de idade em populações não-clínicas e permitem a inferência estatística de efeitos sobre a AV em desenhos experimentais transversais, mas é conhecido que as técnicas reportam diferenças significativas de AV de um mesmo paciente quando comparadas entre si (Prager et al., 1999). Isto se deve ao fato de que elas avaliam aspectos ligeiramente distintos da função visual. Dentre as vantagens da técnica eletrofisiológica, estão a objetividade das medidas (por não dependerem de um observador e nem do desempenho comunicativo da criança) e a garantia de que quaisquer alterações identificadas podem ser atribuídas a problemas na via retino-estriada, ou seja, até o nível de processamento do córtex visual primário (V1). Apesar dessas vantagens, tal técnica exige um grau elevado de investimento em instalação e manutenção de equipamento, a ponto de não ser prática em todos os contextos clínicos de avaliação.

A técnica de olhar preferencial, por outro lado, atingiu ampla disseminação para pesquisa e avaliação clínica, evidenciada pela multiplicidade de testes comercialmente disponíveis que se baseiam nela, como os Cartões de Acuidade de Teller (TAC) (Teller, McDonald, Preston, Sebris, & Dobson, 1986), os Cartões de Acuidade de Cardiff (Adoh, Woodhouse, & Oduwaiye, 1992) e o teste de sensibilidade ao contraste Hiding Heidi (Chen & Mohamed, 2003). A origem desta técnica está nos experimentos psicofísicos iniciados por Fantz (1958), que exploram a tendência espontânea de observadores – incluindo primatas não-humanos e crianças pré-verbais – de fixar com alta probabilidade em regiões mais complexas de um estímulo, como um padrão gradeado sobreposto a um fundo homogêneo. Ao ajustar alguma propriedade de interesse do padrão (como o tamanho das grades ou o seu contraste) até um valor que esteja acima do limiar do observador, nota-se a reduzida frequência da resposta de orientação do olhar para o estímulo.

A técnica desenvolvida por Fantz permitiu um significativo avanço no estudo da

função visual infantil nas décadas de 1970 e 1980, quando foram determinados os perfis quantitativos do desenvolvimento da AV (Teller, 1979), da sensibilidade ao contraste (Banks & Salapatek, 1981, 1978), da discriminação de cor (Teller, 1998) e da estereopsia (Birch, Shimojo, & Held, 1985), dentre outras funções. Diferentemente da técnica de PVCP, a técnica de PL exige da criança as respostas de orientação para o estímulo e manutenção da fixação, o que pode ser um indicador mais completo da função visual que a técnica de PVCP, que não garante por si só que a informação visual recebida no córtex visual primário seja utilizada por regiões neurais relacionadas à coordenação do comportamento (Dobson & Teller, 1978). O relativo baixo custo da técnica, sua praticidade e a existência de extensos estudos normativos que a embasam (D. L. Mayer & Dobson, 1982; Salomão & Ventura, 1995) explicam a popularidade desta técnica em contexto clínico. O exame dos Cartões de Acuidade de Teller (TAC), em particular, não exige extensas sessões de avaliação típicas dos procedimentos psicofísicos clássicos, e aproveita melhor a informação fornecida pela criança durante o momento de avaliação, permitindo ao avaliador realizar julgamentos subjetivos sobre o comportamento global da criança após apresentação do estímulo (Teller et al., 1986). A técnica de olhar preferencial não se restringe à avaliação visual, sendo empregada também em tarefas experimentais para o estudo de um amplo espectro de funções cognitivas na população pré-verbal (Tafreshi, Thompson, & Racine, 2014).

Apesar de sua fundamental importância clínica, a técnica de PL possui algumas limitações. A primeira delas é sua dependência exclusiva nos julgamentos subjetivos de um observador treinado. Alguns observadores podem ser mais conservadores ou mais liberais ao julgar o significado de um movimento ocular, resultado em concordância entre observadores máxima que varia de meia a uma oitava no caso da AV (Harvey, Dobson, Tung, Quinn, & Hardy, 1999; Mash & Dobson, 2005). Diferenças no processo de treinamento dos observadores, ou nos critérios adotados por diferentes avaliadores podem diminuir a reprodutibilidade da técnica quando o objetivo é realizar pesquisas clínicas e comparar os resultados de diferentes laboratórios. Além disso, em contextos de pesquisa, pode ser difícil tornar o testador cego à condição clínica da criança, seja por indisponibilidade de uma equipe extensa de testagem, seja pela condição da criança ser imediatamente aparente (por exemplo, em casos de estrabismo ou malformações oculares). O fato de a técnica depender tradicionalmente de cartões que precisam ser manipulados pelo observador é outra limitação,

pois o avaliador não possui controle exato sobre a posição do estímulo no campo visual da criança, da distância da criança ao estímulo, ou a características como a luminância refletida pelo cartão (Jones, Kalwarowsky, Atkinson, Braddick, & Nardini, 2014). Além disso, o custo de teste, validação e distribuição de novos instrumentos de avaliação é alto, pois depende de um cuidadoso processo de manufatura dos cartões.

O teste de pacientes pré-verbais e não-verbais (por exemplo, com condições neurológicas que impeçam a comunicação verbal com o avaliador) apresenta um desafio significativo, pois exige a quantificação de comportamentos não-verbais como direção da cabeça e/ou olhar, ações cuja relação com a detecção do estímulo podem não ser facilmente descritas de forma objetiva. Apesar deste desafio, perdas visuais nestas populações precisam ser diagnosticadas o quanto antes para melhores resultados de intervenções clínicas, e são necessárias mesmo em idades em que os primeiros testes exigindo instruções verbais ainda não podem ser aplicados. A testagem eletrofisiológica nestes casos sempre é uma opção que pode ser aplicada (e oferece boa capacidade preditiva). A detecção de uma resposta de potencial cortical provocado a um estímulo com determinadas características psicofísicas de interesse significa que este estímulo chegou a ser processado pelo córtex visual primário (Dobson & Teller, 1978), e serve como uma estimativa útil de limite superior para esta capacidade. Mas este limite não necessariamente garante que este processamento signifique que esta informação venha a ser utilizada para guiar o comportamento do paciente. O teste por olhar preferencial (que fornece um indicador do comportamento dirigido pelo estímulo) é uma garantia de que este é o caso, potencialmente oferecendo melhor capacidade preditiva do desempenho do paciente em sua vida diária. Por este motivo, a técnica eletrofisiológica não pode ser considerada como substituta da técnica de olhar preferencial, mas complementar a esta. Apesar da multiplicidade de testes desenvolvidos com base na técnica de olhar preferencial, a técnica ainda pode ser desenvolvida, seja em seus aspectos técnicos (técnica utilizada para detecção e medida do comportamento de detecção do estímulo) ou de procedimento (qual esquema de apresentação utilizar).

Aspectos do desenvolvimento visual e oculomotor precisam ser considerados para o desenvolvimento de técnicas de pesquisa da função visual infantil ou desenvolvimento de ferramentas diagnósticas de perdas visuais. A literatura que será revista nas próximas

duas sessões - que trata do desenvolvimento dos sistemas visual e oculomotor - sugere dois pontos importantes. O primeiro, é o de que as principais funções psicofísicas de interesse clínico (acuidade visual e sensibilidade ao contraste), variam amplamente em um curto intervalo de tempo durante seu curso do desenvolvimento (às vezes aumentando em mais de uma ordem de magnitude dentro de um intervalo de poucos meses). Ferramentas que se proponham a oferecer validade diagnóstica precisam, primeiramente, capturar o curso de desenvolvimento normal dessas funções, oferecendo uma referência normativa a partir da qual desempenhos na tarefa podem ser caracterizados como tendo significância clínica. O segundo ponto é o de que as técnicas que se baseiam no comportamento voluntário e espontâneo do paciente precisam considerar que mesmo o desempenho de linha de base (isto é, comportamentos apresentados pela criança em tentativas nas quais o estímulo não tenha sido detectado) pode variar de paciente para paciente, e também variar sistematicamente de acordo com a idade, a depender das capacidades funcionais do sistema motor e dos comportamentos que este já é capaz de sustentar. Para permitir que variáveis psicofísicas sejam detectadas, a tarefa precisa ser desenhada de tal forma que considere que a criança já esteja plenamente capaz de realizar os movimentos de orientação e/ou seguimento exigidos.

1.2 Desenvolvimento do sistema visual

Apesar de que tanto os fatores do controle do ajuste óptico da formação de imagem (capacidade de acomodação e de vergência) quanto os fatores da morfologia e organização celular retiniana e cortical contribuam para a reduzida capacidade funcional do sistema visual no início da vida, o segundo conjunto de fatores é dominante (Braddick & Atkinson, 2011). Mudanças morfológicas e de organização celular dos fotorreceptores são críticas para o aumento da resolução da imagem formada na retina no início da vida (Yuodelis & Hendrickson, 1986; Abramov et al., 1982): O alongamento dos segmentos externos dos cones e migração celular para a região central da retina. A densidade de cones na foveola passa de 18.921 células/mm² no nascimento para 40.794 células/mm² aos quinze meses de idade. Aos 45 meses de idade, esse valor aumenta para 112.373 células/mm². No mesmo período, os segmentos externos dos cones, regiões onde acontece a absorção de luz, aumentam de 3,1 µm de comprimento ao nascimento para 21-24 µm aos 15 meses e

depois para 37-32 μm aos 45 meses. O diâmetro dos cones diminui de 7.5 μm no início da vida para 2 μm aos 45 meses de idade, em função de seu agrupamento mais compacto na foveola.

O processo de concentração de cones na retina durante os primeiros anos de vida aumenta a capacidade de amostragem da imagem projetada, apesar de que modelos de correlação anátomo-funcionais apontarem para uma contribuição relativa desse fator de apenas 25% relativa à toda a melhora na medida de AV (Banks & Bennett, 1988; Braddick & Atkinson, 2011). Tais modelos sugerem que a melhora funcional restante deve ser atribuída a aspectos da organização celular do restante do sistema visual. O núcleo geniculado lateral (LGN) é uma estrutura da região do tálamo que desempenha função de relé sensorial, repassando os sinais neurais recebidos da retina para o córtex visual. Esta região que recebe aferências a partir das células ganglionares da retina, e seus neurônios fazem sinapse na área visual primária, completando as duas principais vias de formação de imagem (uma via com origem nos bastonetes, processando estímulos periféricos em alta resolução temporal, fazendo sinapse nas células magnocelulares; e outra via com origem nos cones, processando estímulos centrais no campo visual com oponência de cor). Esta região dobra em tamanho nos primeiros 6 meses de vida, atingindo o volume adulto ao sexto mês (Garey & De Courten, 1983). O córtex visual chega a quadruplicar de volume nos primeiros quatro meses de vida, quando também atinge o volume adulto (Garey & De Courten, 1983; De Courten & L. J., 1983; Huttenlocher, de Courten, Garey, & Van der Loos, 1982). Essas alterações anatômicas coincidem em seu curso temporal com melhoras em funções visuais mais complexas que dependem de processamento no sistema nervoso central e integração de informação de diferentes vias, como a estereopsia, e a percepção de forma. O número de sinapses do córtex visual cresce rapidamente nos primeiros meses de vida, atingindo seu valor máximo aos 8 meses idade a partir da qual começa a decrescer dirigido pelo processo de poda sináptica (que continua até aproximadamente os onze anos de idade) (Huttenlocher et al., 1982). O processo de poda sináptica está relacionado à formação de colunas de dominância ocular no córtex visual e ao conseqüente estabelecimento de circuitos visuais funcionais (Huttenlocher et al., 1982).

A AV aumenta de forma rápida nos seis primeiros meses de vida, e após essa idade aumenta de forma mais branda, até atingir níveis próximos da acuidade adulta por volta de

quatro a seis anos de idade (Lewis & Maurer, 2005; D. L. Mayer & Dobson, 1982; Teller et al., 1986; Salomão & Ventura, 1995). A AV média de um recém-nascido é de 0,66 ciclos por grau (cpg,) ao passo em que a de uma criança de 6 meses é de 7,44 cpg, e de uma criança de 36 meses é de 17,82 cpg – um aumento relativo de 27 vezes nesse intervalo de tempo (Salomão & Ventura, 1995). Não existem idades definidas para quando funções visuais como a AV e a sensibilidade ao contraste atingem exatamente os níveis adultos, pois como as curvas de desenvolvimento atingem um platô a partir da segunda infância, variações inter-individuais, acrescidas da variabilidade inerente às técnicas de testagem, levam a grandes variações na idade em que o nível de função adulto é atingido. Algumas estimativas são a de que a AV atinge nível adulto entre os 5-15 anos de idade, e a sensibilidade ao contraste atinge o nível adulto entre os 9-19 anos de idade (Leat, Yadav, & Irving, 2009).

1.3 Desenvolvimento do sistema oculomotor

A função do sistema oculomotor está intrinsecamente ligada ao sistema visual (Wurtz, 2015), apesar de ser anatomicamente distinto deste. Este sistema desempenha uma série de funções que garantem a nitidez, estabilidade e qualidade da imagem formada. O sistema controla a contínua re-orientação conjugada dos olhos (reflexa ou voluntária), que garante a manutenção de diferentes regiões da cena visual na fóvea, região de maior AV, no processo de foveação dirigida por sacadas. Outra função do sistema é o ajuste do diâmetro pupilar para controle da adaptação ao nível de luz ambiente, mantendo o nível de incidência luminosa sobre a retina de modo a não saturar os fotorreceptores (reflexo pupilar), e o ajuste da capacidade refrativa do cristalino (resposta de acomodação), a fim de manter um nível de comprimento focal ideal para manter a nitidez da imagem de objetos localizados a diferentes distâncias. O sistema é importante na visão binocular, sendo responsável pelo ajuste do grau de alinhamento de ambos os olhos ao objeto de interesse no eixo nasal-temporal, que precisa ser distinto para objetos localizados a diferentes níveis de profundidade (movimento de vergência). O sistema também é responsável por garantir a estabilidade das fixações em um objeto de interesse, compensando pelos movimentos de cabeça e mantendo a imagem formada na mesma posição da retina a partir de aferências do sistema vestibular (reflexo vestibulo-ocular), e por fim realiza também a compensação de deslocamentos contínuos do objeto de interesse no campo visual (movimentos de perseguição lenta).

Os movimentos oculares são classificados pela sua dinâmica de aceleração e parada (Lappi, 2016). Do ponto de vista morfológico, a inervação dos músculos retos e oblíquos controlando a posição do olho na órbita parte dos mesmos núcleos de células no tronco encefálico (localizados na formação reticular pontina e mesencefálica), mas estes são por sua vez controlados por múltiplos sistemas subcorticais e corticais (Wurtz, 2015). Alguns movimentos estão sob controle voluntário, outros sob controle reflexo a partir da estimulação do sistema visual, e ainda outros sob controle reflexo do sistema vestibular. Os núcleos do tronco encefálico responsáveis pelo movimento ocular recebem aferências do colículo superior, estrutura que possui papel central no controle dos movimentos oculares. Este, por sua vez, recebe sinapses excitatórias e inibitórias dos núcleos da base e do cerebelo (que possuem função moduladora) e de uma série de regiões corticais, sendo as regiões corticais do campo visual frontal, campo visual complementar e área intraparietal lateral as regiões que têm recebido maior destaque na literatura. Estas regiões são responsáveis pela coordenação dos movimentos a partir de aferências dos sistemas sensoriais visual e vestibular, ou de processos cognitivos mais complexos como a memória visual de curto prazo e planejamento (Robinson, 1986; Wurtz, 2015).

1.3.1 Sacadas e fixações

Como a região central da retina, correspondente à região do campo visual de melhor acuidade (fóvea) possui um raio reduzido, a exploração de cenas complexas exige que os olhos se movam para diferentes regiões para formação de uma representação visual detalhada e coerente. Além disso, estímulos salientes na periferia do campo visual atraem a fóvea em sua direção de forma reflexa, facilitando a formulação de uma resposta adequada a eventos potencialmente perigosos ou que de alguma forma exijam reação rápida. Os movimentos do olhar realizados nestes dois processos, caracterizados por sua rápida aceleração e desaceleração, são denominados de sacadas. Sacadas voluntárias, ou seja, aquelas realizadas como parte dos processos de busca visual, memória visual ou ainda baseadas em planejamento mais complexo são controladas pelo colículo superior a partir de duas áreas corticais: o campo visual frontal (localizado na área pré-motora) e a área intraparietal lateral (IPL) (Leigh & Kennard, 2004; Curtis & Connolly, 2008; Blanke et al., 1999; Gaymard, Ploner, Rivaud-Péchoux, & Pierrot-Deseilligny, 1999). Tanto o campo visual

frontal quanto o colículo superior apresentam mapas motores dos movimentos sacádicos, existindo correspondência direta entre a região estimulada dessas estruturas e a direção e magnitude das sacadas observadas como resultado desta estimulação (Leigh & Kennard, 2004). Já as sacadas reflexivas, aquelas de menor latência e eliciadas por estímulos aparecendo subitamente na periferia do campo visual, parecem estar mais relacionadas com a atividade do campo visual parietal (Leigh & Kennard, 2004). Além dessas áreas, o colículo superior recebe aferências do cerebelo e dos núcleos da base (principalmente da substância negra), estruturas que têm papel modulador e inibidor sobre os movimentos sacádicos (Leigh & Kennard, 2004).

Entre um movimento sacádico e outro, o olho permanece relativamente estável, registrando a mesma imagem ao longo do tempo, em intervalos denominados de fixações. Fixações não são apenas produto de um estado de repouso passivo do sistema oculomotor, mas ocorrem devido à inibição ativa do sistema de controle sacádico (Leigh & Kennard, 2004; Lappi, 2016). Durante as fixações, movimentos que ocorrem em escala espacial e temporal muito mais curta do que as sacadas (movimentos de deriva e microsacadas) são mantidos de forma contínua, impedindo a saturação dos fotorreceptores da retina (Lappi, 2016).

A execução de sacadas requer a programação da direção e amplitude do movimento do olho, calculadas com base na posição do campo visual periférico na qual o objeto de interesse se encontra. Crianças entre 1-2 meses possuem padrão de fixação caracterizado por fixações mais longas e com menos regiões fixadas por intervalo de tempo (Regal, Ashmead, & Salapatek, 1983; Hunnius, 2007). Essas mudanças no desenvolvimento dos movimentos de sacada se refletem nos padrões globais de busca visual realizados por crianças mais novas. Nos primeiros meses de vida, crianças tendem a focar em um menor número de regiões de um estímulo, tendendo a ignorar certos aspectos da cena visual (Hunnius, 2007). Efeitos do conteúdo do estímulo sobre a probabilidade da resposta de orientação também têm sido demonstrados: crianças tendem a preferir faces e padrões similares a outros estímulos (Johnson, 1990; Slater et al., 1998), e a preferir estímulos novos aos que já são familiares (Slater et al., 1998). Com o aumento da idade, as latências para fixações a estímulos na periferia diminuem e a probabilidade de fixação aumenta, bem como e sua precisão (há menor quantidade de sacadas corretivas de curta duração após a primeira sacada na direção do estímulo). Ao escanear estímulos estáticos, crianças de 1-2 meses

examinam estímulos em longas séries de fixações próximas umas das outras (Wass, Smith, & Johnson, 2013; Bronson, 1990), enquanto crianças com 3-4 meses de idades realizam fixações com menor intervalo, sugerindo maior coordenação do sistema oculomotor.

Aslin et al. (1975) estudaram extensivamente as características de movimentos sacádicos em crianças com 1 e 2 meses de idade. Na técnica adotada pelos autores, dois adultos observavam a criança por trás de um espelho unidirecional sobre o qual era projetado para a criança um estímulo central e estímulos periféricos (anéis luminosos) a diferentes excentricidades e ângulos no campo visual. Os observadores anotavam os movimentos realizados (julgando-os apropriados ou não baseado na posição em que o estímulo aparecia), e medidas sincronizadas de eletro-oculografia eram feitas para calcular a latência dos movimentos. Os autores verificavam que estímulos até 30° de ângulo visual eram fixados, mas em sua grande proporção as sacadas realizadas eram hipométricas (não atingiam o estímulo-alvo em um único movimento, mas em uma série de movimentos sucessivos de reajuste), e tinham menor latência quando o estímulo central era removido antes da apresentação do estímulo periférico (tentativas de substituição) do que quando o estímulo central era mantido após apresentação do estímulo periférico (tentativas de adição). Apesar de a resposta de orientação na direção de um estímulo periférico aparecendo no campo visual estar presente desde o nascimento, efeitos dependentes de idade têm sido demonstrados com relação à quantidade de sacadas realizadas até que o estímulo seja efetivamente fixado, latência da primeira sacada, e excentricidade máxima que elicia a resposta de fixação (Hunnius, 2007).

Crianças muito novas têm problemas para desengajar a direção de olhar para um estímulo fixado. Alguns autores têm ressaltado que este processo é produto de imaturidades do componente inibitório, necessário para a criança dirigir o olhar do estímulo fixado para o próximo estímulo que aparece na periferia do campo visual (Butcher, Kalverboer, & Geuze, 2000; Johnson, 1990). Na faixa etária dos 1-3 meses essa dificuldade se reflete em uma menor frequência de fixação em estímulos periféricos na presença de um estímulo central e maior latência de fixação (Butcher et al., 2000), um fenômeno denominado de atenção obrigatória, dificuldade de desengajamento ou ainda de "fixação pegajosa" (Hunnius, 2007). McConnell et al. (2005) estudaram esse processo de dificuldade de desengajamento com um paradigma de desengajar-deslocar, no qual tentativas de desengajar são compostas pelo

estímulo central ainda presente na tarefa; enquanto tentativas de deslocamento removem o estímulo central e adicionam um estímulo periférico. Diferenças entre estes dois processos estão presentes em crianças de dois meses, mas não em crianças de 4 meses de idade, sugerindo maturação da capacidade inibitória sobre o sistema de controle sacádico nesta faixa etária.

As sacadas e fixações são, entretanto, apenas um dos componentes da resposta de orientação, sendo a rotação da cabeça na direção do estímulo outra dimensão relevante. A coordenação entre estes dois movimentos também é uma questão que foi objeto de pesquisa e para a qual variações durante o desenvolvimento foram encontradas. A resposta de orientação ocular geralmente precede a orientação de cabeça, que é realizada para que os músculos extraoculares não precisem sustentar o olho a grandes excentricidades por muito tempo durante as fixações (J. E. Richards & Hunter, 1997). Em adultos, estímulos com posição menor que 20 graus de excentricidade da porção central do campo visual não são acompanhados de orientação de cabeça; os estímulos com posição maior que 20 graus sim. Crianças tendem a realizar uma maior proporção de rotações de cabeça em sua resposta de orientação mesmo para estímulos a curtas excentricidades, em parte para compensar as dificuldades de desengajamento observadas no efeito de atenção obrigatória (J. E. Richards & Hunter, 1997).

1.3.2 Reflexo de acomodação e movimento de vergência

Para manter o foco em objetos próximos, é necessário aumentar a curvatura do cristalino (dioptria), o que por consequência aumenta seu poder refrativo. Para manter o foco em objetos distantes, é necessário relaxar a musculatura do cristalino, diminuindo seu poder refrativo. Essa resposta é controlada pelos músculos ciliares. Crianças nascem com duas dioptrias de hipermetropia, e são incapazes de fixar objetos muito próximos (Banks, 1980). Após o nascimento e até um mês de idade, crianças apresentam resposta de acomodação nula ou extremamente reduzida (Haynes, White, & Held, 1965; Banks, 1980). A capacidade de acomodação melhora significativamente nos dois primeiros meses de vida, e está próxima do valor adulto aos 4 meses de idade. Em função da baixa AV nos primeiros meses de vida, argumenta-se (Banks, 1980) que a impossibilidade de resolver estímulos de maior frequência espacial impossibilita a discriminação entre estimulação visual com maior ou menor grau de

nitidez (sinal que elicia a resposta de acomodação), sendo este o principal fator motivador da menor capacidade de acomodação infantil. Por sua vez, a capacidade de acomodação reduzida pode impedir que padrões espaciais com maior resolução (necessários para dirigir os processos corticais de sinaptogênese e poda sináptica) estimulem o sistema visual, uma situação que exige que tanto o sistema visual quanto o sistema oculomotor se desenvolvam de forma paralela e interagindo reciprocamente para melhora da função visual.

Os movimentos oculares de vergência também estão se desenvolvendo rapidamente nos primeiros meses de vida. Enquanto o sistema de acomodação responde ao grau de nitidez da estimulação visual, o sistema de vergência responde tanto à nitidez da imagem quanto à disparidade entre as imagens geradas nas duas retinas (Bharadwaj & Candy, 2008). Crianças entre 1-2 meses já apresentam resposta de vergência a estímulos estacionários posicionados a curta distância, apesar de esta resposta não estar completamente madura quando a demanda é para fixar objetos em movimento para mais perto ou para mais longe, o que ocorre apenas após os três meses de idade (Aslin, 1977; Hainline & Riddell, 1995).

1.3.3 Perseguição lenta

Os movimentos de perseguição lenta são movimentos oculares de rastreamento de estímulos movendo de forma contínua no campo visual, capazes de reproduzir de forma precisa a dinâmica temporal e espacial dos objetos fixados. Os sistemas neurais envolvidos no sistema de perseguição seguem um processo de maturação distinto do sistema sacádico (Luna, Velanova, & Geier, 2008). Crianças já nascem com a capacidade de fixar em objetos próximos e realizar movimentos básicos de seguimento; entretanto tais movimentos são imprecisos e compensados por sacadas e movimentos de cabeça, atrasados relativos ao estímulo, e tendem a ocorrer com maior probabilidade na direção temporal do campo visual (Johnson, 1990; Daniel & Lee, 1990). Aos três meses de idade, crianças já apresentam melhoras significativas no movimento de perseguição lenta, começam a realizar movimentos antecipatórios, e tendem a fixar com mais facilidade objetos se movendo na direção nasal. Aos seis meses de idade, as crianças já apresentam perfil de movimentos oculares de perseguição lenta próximo de um adulto, principalmente devido ao desenvolvimento de vias funcionais da área do campo visual frontal (localizada no córtex frontal) para o colículo

superior (Johnson, 1990).

Existem assimetrias nos movimentos de perseguição lenta vertical e horizontal durante os primeiros meses de desenvolvimento. Gronqvist et al. (2006) apresentaram estímulos movendo na direção vertical e horizontal, em linha reta (exigindo seguimento em uma única dimensão) e movimentos senoidais (exigindo seguimento em ambas as dimensões) para crianças de 5, 7 e 9 meses de idade. O movimento de perseguição lenta contínua horizontal permanece mais eficiente mesmo em adultos, que realizam menor número de sacadas corretivas nesta direção quando rastreando estímulos em movimentos. Foi demonstrado que este efeito é ainda maior em crianças na faixa etária estudada, com melhor desempenho quando as crianças estavam em idade mais avançada. Movimentos oculares verticais e horizontais são controlados por diferentes gânglios neuronais do tronco encefálico, sendo os movimentos verticais controlados pela formação reticular pontina, e os movimentos horizontais pela formação reticular mesencefálica, apesar de ambos os centros recebem suas aferências diretamente do colículo superior.

Pieh et al. (2012) estudaram os movimentos de perseguição lenta em crianças entre 1 e 18 meses de idade, constatando que a taxa de desenvolvimento desta função atingiu um pico entre 2-6 meses de idade. Para estímulos com velocidade maior, o seguimento mais longo detectável durante a sessão de avaliação foi menor, e houve maior frequência de sacadas durante a apresentação do estímulo. Crianças de seis meses já mostram movimentos de seguimento parecidos com o do adulto, nesta idade mostrando desenvolvimento desacelerado, que atinge paridade com a função adulta até os 18 meses de idade (Johnson, 1990; Pieh et al., 2012).

1.3.4 Reflexo vestibulo-ocular e nistagmo optocinético

Os reflexos vestibulo-oculares são caracterizados por movimentos com dinâmica similar aos de perseguição lenta que reajustam o olho na órbita para manter a imagem estável na medida em que a cabeça se movimenta (fase lenta); caso este movimento exceda a excentricidade máxima que o olho pode rotacionar na órbita, um movimento sacádico de reajuste (fase rápida) é realizado na direção oposta para retornar o olho a uma posição de repouso. A denominação deste movimento se deve ao fato de os sinais de reajuste da posição ocular se originarem não na informação visual, mas no sistema vestibular. É

conhecido que já no primeiro mês de vida o reflexo vestibulo-ocular está presente e seguindo uma forma muito parecida com a observada no adulto (Regal et al., 1983).

O nistagmo optocinético é outro tipo de reajuste reflexo da posição do olhar, que reajusta o campo visual baseado em estimulação em movimento preenchendo a maior parte do campo visual ou sua totalidade, seja por movimento do objeto ou superfície com relação ao observador ou movimento do observador com relação a estes (Daniel & Lee, 1990). Os movimentos são compostos por movimentos de perseguição lenta da direção do estímulo acompanhando sua velocidade, alternados com movimentos rápidos na direção oposta do estímulo que re-estabelecem o olho na posição de repouso, gerando um padrão de posição do olhar com forma de dente-de-serra observável na técnica de eletro-oculografia ou vídeo-oculografia. Diferentemente do reflexo vestibulo-ocular (que parece estar presente em forma parecida com a resposta adulta já no primeiro mês de vida), o reflexo optocinético só atinge características próximas da resposta adulta aos cinco meses de idade. Antes desta idade, o reflexo é detectável, mas existem assimetrias na direção no reflexo, com preferência para movimentos realizados na direção nasal (Naegele & Held, 1982).

1.4 A técnica de vídeo-oculografia

O sistema oculomotor tem sido objeto de estudo há anos, e como consequência, diferentes técnicas foram desenvolvidas para medida dos movimentos oculares ao longo do tempo. Rastreamento ocular é o termo geral para qualquer técnica que realize a medida dos movimentos oculares para fins de determinar a direção do olhar. As principais técnicas utilizadas para rastreamento ocular são (Gredebäck, Johnson, & von Hofsten, 2009; Haslwanter & Clarke, 2010): a eletro-oculografia (medida dos potenciais elétricos gerados pelos músculos controlando a posição do globo ocular), o magnetômetro de indução escleral (técnica que mede pequenas mudanças locais no campo magnético a partir do posicionamento de uma lente contendo um magnetômetro de indução sobre a córnea, que induz um sinal a partir da movimentação do olho), e a vídeo-oculografia, que se baseia no registro em vídeo do deslocamento da pupila relativo a uma ou mais reflexões de fontes de iluminação na superfície da córnea.

A técnica de vídeo-oculografia (VOG), criada no início dos anos 1970 e continuamente aperfeiçoada desde então, é a técnica dominante utilizada atualmente

para quantificação dos movimentos oculares (Gredebäck et al., 2009). Existem inúmeras variações desta técnica dependendo do contexto em que está sendo aplicada, mas todas compartilham de uma série de vantagens em relação às demais técnicas citadas. Apesar de todas as técnicas serem não-invasivas, a vídeo-oculografia não exige montagem de eletrodos ou outros aparatos de contato que podem ser desconfortáveis ao paciente, ou impraticáveis com pacientes não-colaborativos, como crianças pré-verbais. Na situação de montagem em mesa (rastreamento ocular remoto), nem mesmo a câmera precisa ser posicionada em aparato junto ao corpo do paciente, exigindo apenas que o paciente permaneça na região do campo de visão da câmera, o que pode ser realizado com ou sem restrições de posição da cabeça. O surgimento de técnicas de processamento de imagem e mapeamento do ponto do olhar que permitem o registro da posição do olhar na montagem de mesa, sem necessidade de restrição dos movimentos de cabeça, foi um importante passo para o estudo da função do sistema oculomotor e comportamento visual exploratório com a população infantil (Gredebäck et al., 2009). A técnica registra imagens de um dos olhos ou de ambos os olhos em alta taxa de amostragem (entre 60-1000Hz) para extração de características relacionadas com o ponto do olhar (isto é, uma pequena região do campo visual do participante para onde a fóvea está orientada). Geralmente, o ponto do olhar é delimitado por um plano, como a tela de um computador para rastreadores montados em mesa.

Para determinação do sinal de posição do olho nas dimensões horizontal e vertical, duas características básicas extraídas da imagem comumente utilizadas são o centro da pupila e um ou mais reflexos corneais (Duchowski, 2017). Uma fonte de iluminação na banda de comprimento de onda do infravermelho posicionada de forma coaxial com a câmera realizando o registro gera o efeito de pupila clara, tornando a região da pupila distinta da região da íris. Já uma fonte de iluminação em qualquer posição fora do eixo óptico da câmera gera o efeito de pupila escura, também tornando-a distinta da região da íris (Duchowski, 2017). Ambas as técnicas permitem que a pupila seja identificável a partir de algoritmos de segmentação binária em escala de cinza. Após extração do contorno da região da íris (isto é, das coordenadas da imagem que delimitam a região segmentada), técnicas de ajuste de mínimos quadrados para círculos ou elipses podem então ser aplicadas para determinar a coordenada do centro da pupila (Fitzgibbon, Pilu, & Fisher, 1999).

Enquanto para um participante com a cabeça fixa (por exemplo, com um suporte para queixo) a extração da posição da pupila é suficiente para determinação do ponto de olhar após calibração, para situações sem restrição (o que é tipicamente a única possibilidade de aplicação com a população infantil), além da posição da pupila, a extração dos reflexos corneais (também conhecidos como imagens de Purkinje) também é necessária. O princípio básico é o de que a posição absoluta da pupila pode variar amplamente com movimentos da cabeça, mas a posição da pupila relativa a estes reflexos muda apenas com a rotação do olho na órbita (Cornsweet & Crane, 1973; Duchowski, 2017; Gredebäck et al., 2009). Como a córnea pode ser aproximadamente tratada como um espelho esférico, reflexões de fontes de luz pontuais fixas incidirão sobre a mesma posição no plano da imagem registrada pela câmera, apesar de serem refletidas de diferentes posições na superfície da córnea na medida em que o olho muda de orientação. O vetor de posição da pupila relativo ao reflexo corneal não muda significativamente com translações da cabeça no plano perpendicular ao eixo óptico da câmera, e muda relativamente pouco com mudanças na orientação da cabeça. Apesar de a córnea refletir todo tipo de luz incidente, LEDs que emitem luz na faixa do infravermelho são os mais adequados, por emitirem luz imperceptível pelo participante e registrarem de imagens com características que favorecem a performance dos algoritmos de segmentação (Duchowski, 2017), além de possuírem perfil espectral distinto da iluminação artificial (o que permite a completa eliminação de reflexos artefactuais, como a luz da tela do computador, através do uso de filtros ópticos). Dependendo da técnica de mapeamento de ponto do olhar utilizado, mais de uma fonte de iluminação pode ser utilizada, gerando múltiplos pontos de reflexo na superfície da córnea. Montagens com duas ou quatro fontes de iluminação são comumente utilizadas (Duchowski, 2017; Guestrin & Eizenman, 2006; Yoo & Chung, 2005).

Após extração das características da imagem, uma função matemática que mapeia as características extraídas da imagem para a coordenada de ponto do olhar do participante precisa ser aplicada. Também nesta etapa existem múltiplas técnicas disponíveis, que podem ser divididas entre aquelas baseadas em modelos bidimensionais estimados via regressão ou interpolação (Hansen & Ji, 2009) ou modelos geométricos tridimensionais (Guestrin & Eizenman, 2006). No primeiro conjunto de técnicas, uma série de estímulos de calibração ocupando cada um uma pequena área do campo visual e com amplo

espaçamento entre si são apresentados ao participante enquanto amostras de posição da pupila (ou deslocamento da pupila relativo ao reflexo corneal) são registradas. Essas amostras são utilizadas para ajuste de um modelo de regressão ou interpolação que objetiva aproximar uma função não-linear de mapeamento desconhecida com um modelo matemático mais simples. Comumente são adotados modelos polinomiais de grau 2 ou 3 e produtos cruzados das coordenadas registradas (Cerroloza, Villanueva, & Cabeza, 2008). Termos de grau zero do modelo de calibração (interceptos) capturam o deslocamento da face com relação ao eixo óptico da câmera; enquanto termos de grau um, dois ou mais (fatores de escala) capturam a distância do participante à tela, a orientação relativa da face, além da anatomia de cada participante (grau de separação dos olhos, tamanho do olho e pupila, e grau de excentricidade da fóvea) (Hainline & Riddell, 1995).

As técnicas que empregam modelos geométricos tridimensionais são baseadas em modelos explícitos do processo de aquisição de imagem. Ao invés de aproximar uma função não-linear desconhecida, as restrições impostas pelo posicionamento relativo da câmera, das fontes de iluminação, combinadas com as características extraídas da imagem, são combinadas em um modelo matemático que objetiva calcular o ponto de olhar como a interseção do eixo visual do participante no espaço tridimensional com o plano do monitor (Guestrin & Eizenman, 2006; Villanueva et al., 2009). Para estimação do eixo óptico do participante (Figura 1.b), a câmera precisa ser previamente calibrada. No processo de calibração da câmera, seu comprimento focal é calculado, bem como a interseção do eixo óptico da câmera com o plano de imagem e os parâmetros extrínsecos de translação e rotação relativos ao plano do monitor. As posições dos LEDs que geram as reflexões corneais relativas à câmera no espaço tridimensional também precisam ser medidas e mantidas constantes durante a sessão experimental. A calibração da câmera e conhecimento das posições relativas das fontes de iluminação permitem o cálculo dos vetores tridimensionais de projeção da imagem. Restrições ópticas são então incluídas no modelo para determinação dos parâmetros de dois ou mais planos onde a câmera, os reflexos corneais e as fontes de iluminação estão situadas. A posição do centro da esfera corneal que gerou as reflexões dos LEDs (Figura 1.a) necessariamente se situa na linha de interseção dos planos dos reflexos, e sua posição nesta linha pode ser calculada com algoritmos de otimização aplicados sobre os parâmetros livres do modelo. O centro da

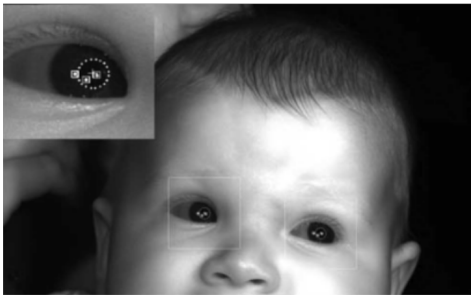
esfera corneal (relativo ao ponto nodal da câmera), combinado com o centro da pupila (calculado a partir de parâmetros anatômicos conhecidos e das características extraídas da imagem) definem o eixo óptico do participante. Pequenas variações anatômicas ainda exigem um rápido procedimento de calibração de dois pontos para determinação do grau de excentricidade entre o eixo óptico e o eixo visual do participante, mas o uso de duas câmeras não exige este procedimento (Hathibelagal, Leat, Irving, Nandakumar, & Eizenman, 2015; Model & Eizenman, 2010).

Outra técnica que também se baseia na modelagem geométrica tridimensional explora o invariante de razões cruzadas sob transformação de perspectiva (Yoo & Chung, 2005; Coutinho & Morimoto, 2013). Nesta técnica, quatro LEDs são posicionados em torno do monitor, de forma a delimitar uma região de interesse onde o ponto do olhar será situado. Quatro reflexos corneais são então extraídos da imagem registrada pela câmera (Figura 1.c). A partir das características destas reflexões registradas na imagem, é realizada a estimação da projeção de quatro pontos virtuais sobre o plano da imagem (Figura 1.d). Estes pontos virtuais, por sua vez, são gerados pela projeção das reflexões em um plano tangente à córnea e perpendicular ao eixo óptico do participante. Um modelo matemático baseado em duas transformações de perspectiva (uma entre o ponto de olhar no monitor e o plano tangente à córnea; e outra entre o plano tangente à córnea e o plano de imagem) é então aplicado. Como as razões cruzadas entre os vetores de posição da pupila são preservadas entre estas duas transformações, a posição do olhar pode ser calculada a partir de uma simples expressão baseada na igualdade entre as razões cruzadas dos vetores de cada um destes planos.

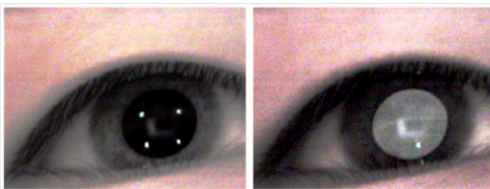
As técnicas baseadas em modelos geométricos tridimensionais exigem menor tempo de calibração, apresentando vantagem relativa em situações nas quais este procedimento é pouco aplicável ou indesejável no protocolo experimental (dependendo da montagem, registros sem qualquer procedimento de calibração com o participante podem ser realizados). Seu uso, entretanto, exige montagem mais complexa, com conhecimento dos parâmetros extrínsecos e intrínsecos da câmera, além de medida acurada do posicionamento relativo dos equipamentos (câmera e monitor). A técnica baseada em invariantes de razões cruzadas não exige estes procedimentos (Yoo & Chung, 2005), mas exige a detecção de maior número de reflexos corneais, que pode ser desafiadora

Figura 1: Técnicas de vídeo-oculografia baseadas em modelagem tridimensional. Painéis (a) e (b) reproduzidos a partir de Model et al. (2010); Painéis (c) e (d) reproduzidos a partir de Yoo & Chung (2005)

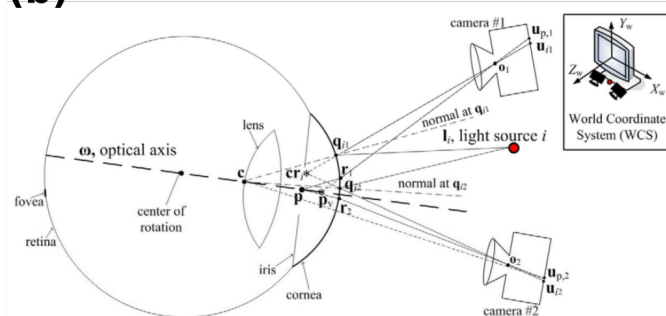
(a)



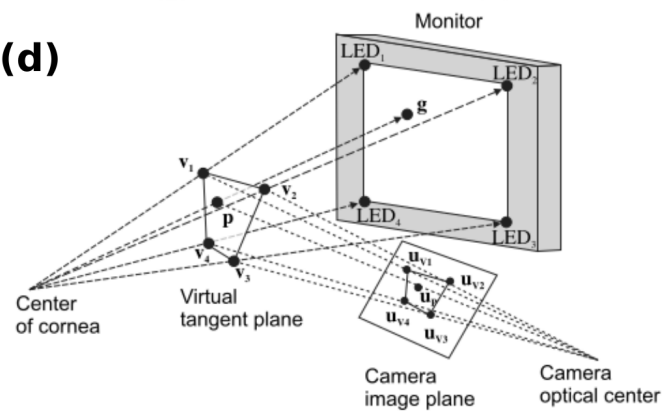
(c)



(b)



(d)



dependendo da resolução e fator de magnificação da imagem registrada, um processo que pode acabar por sacrificar o campo de visão da câmera. As técnicas baseadas em modelos bidimensionais estimados via regressão são comparativamente mais simples, mas exigem que um procedimento de calibração possa ser completado com o participante com sucesso, ou que procedimentos de calibração offline sejam aplicados manualmente após a sessão de registro (Kooiker et al., 2016). A desvantagem deste segundo conjunto de procedimentos é que a qualidade da calibração pode degradar ao longo da sessão, e pode ser pouco aplicável em situações nas quais o paciente se mova para longe da região onde a calibração originalmente foi realizada, uma limitação à qual as técnicas de modelagem geométrica não estão sujeitas.

1.5 Uso clínico da vídeo-oculografia

1.5.1 Aplicações clínicas da técnica de vídeo-oculografia em adultos

A técnica de vídeo-oculografia vem sendo utilizada na clínica oftalmológica e ortóptica para determinação objetiva do grau de alinhamento ocular e presença de alterações nos movimentos oculares, com algumas ferramentas comerciais especializadas para esta aplicação disponíveis no mercado (Narvaez Palazón, Sánchez Ventosa, Nieves Moreno, Redondo Ibáñez, & de Liaño Sánchez, 2021; Bedell & Stevenson, 2013). Para medida do alinhamento ocular, o paciente é mantido com restrição via apoio de queixo, e uma câmera com pequeno campo visual é focada em ambos os olhos, e iluminação infravermelha é utilizada para verificar o grau de alinhamento dos eixos oculares, e a eventual presença de desvios significativos (Narvaez Palazón et al., 2021).

Para fins de análise quantitativa dos movimentos oculares, parâmetros clínicos de interesse incluem a acurácia das sacadas, parâmetro de ganho nos movimentos de seguimento lento, e estabilidade das fixações, além da eventual presença de nistagmo (Bedell & Stevenson, 2013). Esta aplicação da vídeo-oculografia é uma alternativa à técnica de eletro-oculografia, que historicamente foi usada (e ainda o é) para a mesma finalidade (Bedell & Stevenson, 2013). Na avaliação diagnóstica do nistagmo, a técnica permite a quantificação exata do padrão de movimento ocular apresentado pelo paciente, o que pode ser útil no diagnóstico diferencial das diferentes condições neurológicas associadas a esta apresentação clínica (Clark et al., 2019).

O paradigma de anti-sacadas tem sido explorado como uma possibilidade de aplicação da técnica de rastreamento ocular com a população adulta (Antoniades et al., 2013; Larrazabal, Garcia Cena, & Martinez, 2019). Neste paradigma, um estímulo é apresentado na periferia do campo visual do paciente, enquanto se exige que este suprima a tendência de dirigir o olhar para o estímulo, e ao invés disso olhe para a direção oposta. A técnica tem sido apontada como um potencial indicador de sinal precoce de doenças neurodegenerativas como a demência de Alzheimer, na qual os pacientes têm dificuldades em inibir a tendência natural de realizar uma sacada na direção do estímulo. A ferramenta também tem sido estudada para caracterizar os sintomas motores da doença de Parkinson, em tarefas que permitem constatar características como sacadas hipométricas (que não atingem o alvo com precisão na primeira tentativa), maior latência e fragmentação nas sacadas realizadas (Larrazabal et al., 2019). Existe pelo menos uma publicação da especificação técnica de um sistema para detecção de movimentos oculares com foco específico em doenças neurológicas em adultos (Hernández, Hernández, Molina, Acebrón, & García Cena, 2018).

1.5.2 Utilização da vídeo-oculografia para tarefas de olhar preferencial

Não obstante o uso da vídeo-oculografia em algumas aplicações clínicas (Narvaez Palazón et al., 2021; Bedell & Stevenson, 2013), o uso da técnica com a população infantil para fins de detecção de limiares das diferentes funções visuais ainda se restringe à pesquisa. Dada a ampla utilização do paradigma de olhar preferencial tanto na pesquisa quanto na clínica, naturalmente vários trabalhos se dedicaram à questão da possibilidade de transpor esta técnica para uma situação em que a técnica de vídeo-oculografia é utilizada no lugar de um observador treinado. Borchert (2000) et al. descrevem em uma patente um sistema para medida da AV infantil utilizando rastreamento ocular, mas resultados de validação da ferramenta não estão publicados. Na descrição da invenção, os autores descrevem a apresentação um estímulo central para capturar a atenção do participante, e em seguida são apresentados dois ou mais estímulos de teste na periferia do campo visual, enquanto a posição do olhar é medida. Lengyel et al. (2003) examinaram 51 bebês saudáveis com idade entre 3-93 dias de idade, apresentando um quadrado movendo-se a 7.5°/seg para eliciar movimentos de seguimento lento. Um monitor CRT apresentava os estímulos. Os autores compararam os resultados desta técnica com o exame de olhar

preferencial com os cartões de acuidade de Teller, obtendo correlações significantes entre as duas técnicas, além de observar tendência de desenvolvimento da AV significativa dentro do curto intervalo de idade avaliado.

Kooiker et al. (2016) avaliaram 126 crianças com idade de 1-14 anos com baixa visão (utilizando os critérios da OMS de AV abaixo de 6/18 ou campo visual menor que 20 graus) com o objetivo de comparar os resultados de uma bateria de testes de avaliação visual tradicionais com soluções utilizando técnica de rastreamento ocular. As crianças foram divididas por idade em dois grupos, um com 1-6 anos de idade e outro com 7-14 anos de idade. Os autores utilizaram procedimento automatizado de calibração oferecido pelo rastreador ocular Tobii, mas utilizaram um procedimento próprio de re-calibração após a sessão quando o procedimento de calibração inicial falhou. Foram apresentadas figuras de faces se movendo pela tela para avaliar o movimento de seguimento, e os demais estímulos possuíam tamanhos de 6° e posições estáticas nos quatro cantos do monitor. Os estímulos eram apresentados por quatro segundos em cada um dos quadrantes. As sessões de avaliação eram de 7 minutos. O critério adotado pelos autores de qualidade da sessão foi de pelo menos 15% do tempo de registro válido. Os registros de rastreamento ocular foram empregados para classificar a posição do olho como adequada ou não para cada apresentação do estímulo alvo. Um movimento ocular não adequado era assim classificado quando ocorria nistagmo, quando as crianças não conseguiam sustentar a fixação aos estímulos ou se não apresentassem sacadas nas direções dos estímulos. A possibilidade de avaliação da visão de cores e contrastes também foi examinada (através da capacidade fixar estímulo de pelo menos 1,25% de contraste). A concordância entre as múltiplas metodologias de avaliação variou entre 75-90%, e a correlação com a avaliação clínica tradicional variou de 0.39 a 0.69 para cada função testada. Os autores concluíram que a avaliação por parte de um observador é comparável utilizando-se as duas técnicas, e que a técnica de rastreamento ocular apresenta a vantagem de permitir a revisão do procedimento e dos resultados. Concluiu-se que o procedimento desenvolvido pode ser utilizado para triagem rápida (o procedimento pode ser realizado em 7 minutos), mas não permite quantificação precisa de limiares de AV ou sensibilidade ao contraste.

Jones et al. (2014), utilizando sistema de rastreamento ocular comercial, elaboraram um protocolo de medida de AV, comparando os resultados com os de um teste

já estabelecido, o Keeler Infant Acuity Cards. Os valores de AV obtidos em uma amostra de 30 crianças com idades entre 2 e 12 meses foram concordantes com o teste tradicional (86% diferiram em menos de 0.3 logMAR) e com normas de acuidade previamente estabelecidas (78% dos valores estavam dentro dos limites normativos). No desenho dos autores, padrões Gabor (ondas senoidais moduladas por uma função gaussiana bidimensional) eram apresentados ao longo de uma circunferência, a um ângulo aleatório a cada apresentação, diferindo da apresentação tradicional de variar a apresentação em apenas dois pontos (esquerda e direita).

Um procedimento similar, realizado por Hathibelagal et al. (2015) descreve os resultados da técnica de olhar preferencial com uso de uma solução customizada de rastreamento ocular composta por iluminadores infravermelhos e por uma câmera monocromática de alta resolução. Das 19 crianças testadas (que possuíam de 3 a 11 meses de idade), 89% obtiveram valores de AV concordantes entre dois exames realizados (resultados com diferença menor que 0.3 logMAR), e as diferenças entre o procedimento e o teste tradicional dos Cartões de Acuidade de Teller (TAC) foram menores que 0.3 logMAR em 74% dos casos no primeiro teste e 58% dos casos no segundo teste. O procedimento dos autores apresenta uma vantagem que é a dispensa de procedimento de calibração, pois o uso de duas câmeras permite a triangulação da posição da córnea em espaço tridimensional a partir da modelagem da projeção do centro da pupila e dos reflexos de Purkinje no espaço, e a posição do olhar é estimada como a interseção dos dois vetores estimados e o monitor.

Vrabic et al. (2021) compararam os Cartões de Acuidade de Keeler (crianças mais novas) e os símbolos LEA (crianças mais velhas) com o método desenvolvido de rastreamento ocular em 36 crianças saudáveis. As idades variaram entre 5 meses e 16 anos. 97% das crianças completaram com sucesso ambos os testes. Pesquisadores utilizaram o Tobii Pro X3-120. A testagem foi realizada em sala escura. Os autores utilizaram procedimento de calibração de 5 pontos. A correlação (coeficiente de Lin) entre os dois exames foi de 0.53 (intervalo de confiança de 0.31-0.72). Os autores identificaram que no subgrupo de crianças com baixa acuidade (>0.4 logMAR) o teste de rastreamento ocular tendia a superestimar a acuidade relativa à testagem padrão; e tendia a subestimar a acuidade em crianças com alta AV (<0.4 logMAR). Apesar dessas diferenças, os autores julgaram o desempenho dos testes comparável.

Em outro estudo, Esteban-Ibanez et al. (2021) avaliaram crianças com idade entre 6 meses e 7 anos por meio de rastreamento ocular tanto na sua AV quanto na sua sensibilidade ao contraste. Os autores nomearam o procedimento de DIVE (*Device for Integrated Visual Examination*). Duas fases do estudo são relatadas: Na primeira, 60 pacientes sem problemas de visão foram testados para levantamento dos valores normativos do teste. Destes, 57 foram testados com sucesso no teste de acuidade visual, e 44 foram testadas com sucesso no teste de sensibilidade ao contraste. Na segunda fase do teste, foram testados 74 pacientes, divididos em um grupo de 28 pacientes com desenvolvimento normal e 41 pacientes com condições clínicas potencialmente relacionadas à função visual (31 eram pré-termo, 5 com baixo peso gestacional, 3 com catarata congênita e dois com nistagmo congênito). Tanto a AV quanto os valores de sensibilidade ao contraste aumentaram com a idade. A diferença do teste de grades de LEA e o teste de rastreamento ocular foi de 1.05 cpd (CI de -9.95-7.84 cpd). Os autores constataram que as diferenças entre os testes foram menores para crianças mais novas que um ano de idade.

Chang et al. (2021) testaram 16 crianças com perda visual cortical entre 12 meses e 12 anos de idade. Os autores compararam o resultado de teste visual clínico com a técnica de rastreamento ocular. Foi encontrada alta correlação entre os resultados de AV medida por ambas as técnicas ($r=0.82$), e alta fidedignidade teste-reteste ($r=0.96$).

Em resumo, todos os trabalhos publicados até o momento dedicados a realizar a técnica de olhar preferencial utilizando vídeo-oculografia demonstram correlações significativas dos valores medidos com as técnicas tradicionais, mas o grau de concordância difere amplamente de estudo para estudo, o que é esperado dada a grande variação metodológica (sistema de rastreamento ocular, estímulos apresentados e esquema de apresentação) e variação da população estudada (idade e presença ou ausência de diagnóstico de condição visual). Alguns trabalhos utilizam um intervalo de idade muito amplo, como o de Vrabic et al. (2021), o de Chang et al. (2021) e o de Kooiker et al. (2016). A testagem com crianças pré-verbais difere da testagem de crianças verbais, pois no primeiro caso as crianças não podem receber e agir de acordo com instruções verbais passadas pelo aplicador. Medidas globais de testabilidade reportadas por estes trabalhos, que agregam tanto crianças pré-verbais quanto crianças e adolescentes verbais, não são particularmente informativas quando à aplicabilidade da técnica na população pré-verbal e não-verbal,

pois o procedimento realizado é inevitavelmente diferente com crianças mais velhas que podem atender a instruções verbais do avaliador. A Tabela 1 compara alguns aspectos metodológicos dos diferentes trabalhos citados, com os objetivos de ilustrar a ampla variação de metodologia até então encontrada na literatura, e o de situar a metodologia que virá a ser descrita no presente trabalho dentro do conjunto de trabalhos prévios.

2 Justificativa

Dada a ubiquidade da técnica de olhar preferencial em contexto de avaliação clínica da função visual de crianças pré-verbais, consideramos a possibilidade de utilizar a tecnologia de rastreamento ocular como uma alternativa ao método tradicional dependente do julgamento de um observador treinado. Este procedimento supera várias das limitações da técnica de PL tradicional. A primeira delas é maior reprodutibilidade da técnica, em função da solução automatizada permitir registrar de forma objetiva as trajetórias de movimentos oculares do paciente durante a apresentação de cada estímulo. Estes registros podem ser mantidos em bases de dados para revisão dos critérios de análise, verificação da qualidade dos registros, e comparação dos resultados entre diferentes locais de realização dos exames. A apresentação automatizada de estímulos apresenta outras vantagens à parte, que são maior controle sobre parâmetros da aplicação (luminância de fundo, posição do estímulo no campo visual da criança e distância da criança para o estímulo) e a maior flexibilidade para selecionar a tarefa e os parâmetros de estimulação mais adequados para cada quadro clínico. Este segundo ponto permite ainda redução do custo e maior agilidade do processo de validação, por não depender de processo de manufatura de novos equipamentos e aparatos, bastando a distribuição de programas de computador para apresentação dos estímulos. Por se tratar de solução baseada em software e em dispositivos periféricos acessíveis, a solução é portátil e a montagem do equipamento não requer a presença de equipe especializada, podendo ser feita por assistência remota. Essas vantagens permitem maior agilidade e menor custo no processo de validação e a eventual distribuição comercial da solução para clínicas e hospitais.

Apesar de ser uma ferramenta amplamente utilizada em áreas como pesquisa de processos psicológicos básicos, design de interfaces gráficas e marketing (Duchowski, 2017), além de algumas aplicações na própria clínica oftalmológica (Narvaez Palazón et al., 2021; Bedell & Stevenson, 2013), a vídeo-oculografia ainda não encontrou espaço como ferramenta para aplicação da técnica de olhar preferencial, com o objetivo de determinar limiares psicofísicos nas populações pré-verbais e não-verbais. Trabalhos recentes chegaram a automatizar o procedimento de olhar preferencial (Hathibelagal et al., 2015; Jones et al., 2014), demonstrando boa concordância com as técnicas tradicionais, mas estes trabalhos ainda não geraram soluções aplicáveis no uso clínico cotidiano. A

Tabela 1: Resumo dos estudos dedicados à aplicação da vídeo-oculografia para estudo das funções visuais a partir do olhar preferencial.

Trabalho	Monitor	Iluminação ambiente	Estímulos de teste	Sistema de rastreamento ocular	Participantes	
Esteban-Ibanez et al. (2021)	12" cd/m2	120	Sala escurecida com iluminação indireta.	4 posições (cantos) com duração variável de até 3 segundos. Estímulo lúdico contingente à detecção. Procedimento adaptativo.	Sistema próprio, 60Hz, com calibração de 5 pontos	6 meses - 7 anos (sem diagnóstico clínico)
Hathibelagal et al. (2015)	21" 67 cd/m2		Sala escurecida com iluminação indireta	4 posições (cantos) via Procedimento adaptativo.	Sistema próprio (dispensa calibração)	3.2 - 11 meses (sem diagnóstico clínico)
Kooiker et al. (2016)	24" cd/m2	210	Iluminação ambiente indireta.	4 posições (cantos) variando em contraste via procedimento adaptativo	Sistema comercial (Tobii T60XL). Calibração de 5 pontos.	1 ano - 14 anos (diagnóstico baixa visão)
Jones et al. (2014)	30"		Não especificado	Procedimento adaptativo com estímulo lúdico contingente à detecção. Padrões Gabor distribuídos circularmente a 8° do centro do monitor.	Sistema comercial (Tobii X120) com calibração customizada de 2 pontos	2.6-12.7 meses (sem diagnóstico clínico)
Chang et al. (2021)	24", cd/m2	100	Sala escura, com iluminação apenas no monitor.	Esquerda-direita apresentados por dois segundos via procedimento fixo.	Sistema comercial (SR Research EyeLink 1000 Plus)	12 meses-12 anos (diagnóstico de perda visual cortical)
Vrabic et al. (2021)	15.6"		Sala escura, com iluminação apenas do monitor.	Esquerda-direita (padrão gradeado de alto contraste)	Sistema comercial (Tobii Pro X3-120)	5 meses-16 anos (sem diagnóstico clínico)

transferência de tecnologias de avaliação visual requer um rigoroso processo de validação com uma população-alvo, cujo custo em termos financeiros e de tempo precisa se justificar sob o ponto de vista de garantir que uma nova técnica, mesmo que aplicável em contexto de pesquisa, atenda aos requisitos de usabilidade em contextos clínicos. Considerações como o tempo reduzido do qual os profissionais que vão realizar a avaliação dispõem, o nível de investimento em equipamento e equipe necessários, a portabilidade e praticidade do teste precisam ser consideradas previamente a investimentos em estudos de validação.

Três limitações para o uso clínico amplo da vídeo-oculografia na clínica (mais especificamente, na avaliação neurológica) foram identificadas por Clark (2019): (1) Indisponibilidade de equipe treinada para aplicar e interpretar resultados; (2) Escassez de dados normativos que possam servir de referência para determinar quadros clínicos; e (3) Falta de padronização de protocolos que possam ser reproduzidos. A revisão bibliográfica conduzida na sessão anterior sugere situação similar na área de pesquisa em psicofísica clínica: Apesar de existirem alguns trabalhos que demonstram a possibilidade de uso da técnica, seu uso ainda não é tão abrangente a ponto de convergir para procedimentos que sejam universalmente aceitos como protocolos clínicos. As dificuldades encontradas para a transposição da técnica para a clínica não parecem estar relacionadas às limitações da técnica em si, mas sim à carência de pesquisas translacionais com foco clínico.

A criação de bancos de dados normativos que sejam úteis do ponto de vista clínico, por sua vez, pressupõe algum grau mínimo de padronização de procedimentos. No caso dos exames de acuidade visual e sensibilidade ao contraste tradicionais, esta padronização é usualmente garantida pelo próprio processo de manufatura dos cartões, pranchas ou outros artefatos nos quais o teste é baseado, que é centralizado em um ou um conjunto pequeno de fabricantes responsáveis pela produção, controle de qualidade e comercialização dos testes. Para exames baseados na técnica de vídeo-oculografia, uma padronização de procedimentos tal como estabelecido para os testes tradicionais ainda não existe, como pode ser atestado pela ampla variedade metodológica dos trabalhos prévios. Soluções baseadas em software, entretanto, têm uma vantagem justamente por não depender de um processo de manufatura centralizado, que é a possibilidade de criação de padrões compartilhados por múltiplos agentes comerciais interessados em desenvolver e comercializar ferramentas baseadas na técnica, tal como é feito na área de eletrofisiologia visual, cujos protocolos

geralmente são baseados nos padrões estabelecidos pela *International Society for Clinical Electrophysiology of Vision* (ISCEV) (Robson et al., 2018).

A existência de um padrão para aplicação clínica da técnica de olhar preferencial por vídeo-oculografia poderia especificar, por exemplo, condições gerais de iluminação ambiente, especificações técnicas do monitor de apresentação dos estímulos (luminância, dimensão física, resolução), do hardware de aquisição de imagem (exposição, ganho, profundidade de foco, fator de magnificação e tamanho do campo de visão), bem como dos desempenhos esperados dos algoritmos de processamento de imagem, de forma a permitir que essas condições possam ser reproduzidas por múltiplos agentes (comerciais ou não) interessados no desenvolvimento da técnica. Quando se trata de população pré-verbal, uma série de elementos precisam ser estabelecidos, em adição aos aspectos técnicos citados. Por exemplo, a técnica de vídeo-oculografia é dependente de procedimento de calibração com o participante, mas procedimentos de calibração de propósito geral são falíveis com população pré-verbal (Harris, Hainline, & Abramov, 1981), sugerindo que esforços também podem ser desenvolvidos no sentido de estabelecer qual das técnicas de vídeo-oculografia é a mais adequada para o maior número de participantes. Além disso, o manejo de pacientes em diferentes situações (baixa colaboração e adesão ao teste) pode também ser especificada, gerando melhores estimativas de testabilidade.

3 Objetivos

3.1 Objetivo Geral

Construir um teste de avaliação de acuidade visual por vídeo-oculografia para aplicação clínica, baseado em um sistema de rastreamento ocular de baixo custo.

3.2 Objetivos específicos

(1) Desenvolver uma solução automatizada da técnica de olhar preferencial (*Automated Preferential Looking* - APL) aplicável a crianças de 2 a 36 meses de idade. Tal solução consiste em um módulo de apresentação de estímulos e de um módulo de detecção e análise de movimentos oculares em tempo real;

(2) Determinar a adequação dessa solução para uma aplicação clinicamente relevante, que é a avaliação de acuidade visual de resolução (comparável à realizada pelo método tradicional dos Cartões de Acuidade de Teller - TAC-II).

(3) Determinar se os valores de AV calculados com esta aplicação são consistentes com os valores estimados a partir do método clássico dos Cartões de Acuidade de Teller II.

(4) Verificar se os valores de AV registrados com a atual técnica são consistentes com as curvas de desenvolvimento estabelecidas com o exame TAC-II;

(5) Desenvolver uma aplicação para plataforma Desktop que permita a fácil aplicação do exame, que esteja em condições de ser distribuída para fins futuros de validação em larga escala, licenciamento para outros usos, ou uso direto em contexto clínico;

4 Metodologia

4.1 População

Para determinar a validade, reprodutibilidade e usabilidade do teste, foram recrutadas crianças saudáveis em uma faixa etária próxima da faixa etária para a qual o exame dos Cartões de Acuidade de Teller (TAC-II) foi validado, compreendendo idades entre 2 e 36 meses de idade. O tamanho amostral adotado é o de 60 participantes, calculado com o objetivo principal de determinar a magnitude de potenciais diferenças sistemáticas entre o método APL e o método TAC-II. As 60 crianças recrutadas serão distribuídas de forma homogênea dentro do intervalo de idades para o qual o exame TAC-II foi validado, a fim de capturar a dinâmica normal de desenvolvimento da AV (Salomão & Ventura, 1995). Pacientes com diagnóstico de qualquer condição neurológica ou oftalmológica com potencial prejuízo visual ou do desenvolvimento psicomotor (determinado por avaliação médica e psicológica realizada no próprio ambulatório onde a pesquisa foi realizada) foram excluídas da amostra, bem como pacientes que apresentassem AV significativamente abaixo dos limites normativos (<1 oitava do limite inferior). Pacientes prematuros foram incluídos na amostra, desde que não apresentassem prejuízo visual de acordo com o mesmo critério.

Foi estabelecida colaboração com o Departamento de Pediatria do Hospital da Faculdade de Medicina de Jundiaí (HU-FMJ) para realizar o recrutamento dos participantes do presente estudo. As sessões para coleta de dados foram realizadas neste mesmo local, aproveitando a ocasião na qual as mães realizam consultas de acompanhamento pediátrico, a fim de garantir a adesão e minimizar custos de transporte das mães e crianças. A criança permanecia junto de seus pais ou de um responsável autorizado durante todo o procedimento, cabendo ao avaliador apenas orientar sobre como posicionar a criança para acompanhar a tarefa experimental. O presente procedimento foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Psicologia, sob o número de processo 42380020.9.0000.5561 (CAAE), e também obteve autorização da diretoria do HU-FMJ para sua realização. Todos os responsáveis que participaram da tarefa com seus filhos assinaram um termo de consentimento livre e esclarecido (TCLE), que detalhava os objetivos e procedimentos do estudo (Anexo A).

4.2 Materiais

Um monitor Dell Ultrasharp U2419H de 23,8" (53,83 cm x 34,84 cm) foi utilizado para apresentação da tarefa experimental. A dimensão horizontal do monitor (53,83 cm) é próxima da largura dos Cartões de Acuidade de Teller (55,5 cm), de forma que os estímulos apareciam aproximadamente na mesma excentricidade horizontal do que nos cartões, em tentativas nas quais o paciente estivesse fixando no centro do monitor antes da apresentação. O tempo de resposta do monitor é de 8 ms de acordo com as especificações técnicas do fabricante (metade do intervalo entre quadros da câmera utilizada para rastreamento ocular). A distância entre *pixels* do monitor é de 0,2745 mm, de forma que o padrão gradeado com maior frequência espacial representável pelo monitor é de 19,51 ciclos/cm (assumindo que dois pixels são utilizados para representar o menor ciclo possível). Uma distância de apresentação de 60 cm foi utilizada para todas as crianças, de forma que todos os padrões de teste ocupassem 14,76 graus de ângulo visual a esta distância.

A aquisição de imagem foi realizada com uma câmera Firefly S2, (FLIR Imaging), equipada com sensor CMOS IMX296 (dimensão 1/2.9"). As imagens foram adquiridas em escala monocromática, com uma taxa de amostragem de 60 Hz e resolução de imagem de 1440 por 900 pixels. Uma lente de 4,0 mm foi utilizada para todas as sessões. Um LED infravermelho (pico de comprimento de onda 850 nm) foi posicionado ao lado da câmera a fim de gerar o reflexo corneal e facilitar o processamento de imagem baseado no efeito de pupila escura. O modelo de câmera utilizado foi escolhido pela adesão ao protocolo GenICam (interface de programação compartilhado por várias câmeras no mercado), e fácil instalação via interface USB (utilizando biblioteca aberta *libusb*), o que permite certo grau de independência da aplicação com relação ao hardware utilizado. Qualquer câmera USB que atenda ao protocolo pode princípio pode ser utilizada pelo instrumento, desde que cumpra requisitos mínimos de resolução espacial e temporal equiparáveis ao do modelo utilizado. Tal arranjo permite levantar um conjunto de especificações técnicas da aquisição de imagem que precisam ser atendidas e possam ser verificáveis por análise das imagens digitais geradas (como iluminação e contraste médios) para que a solução que está sendo desenvolvida seja aplicável independente da escolher particular de hardware.

O uso de câmeras USB de propósito geral para rastreamento ocular é uma tendência recente em sistemas de vídeo-oculografia utilizados em pesquisa, motivado

pelo menor custo e maior flexibilidade para implementação de algoritmos em etapas de prototipagem (Matsuda, Nagami, Sugase, Takemura, & Kawano, 2017; Hosp et al., 2020; Ivanchenko, Rifai, Hafed, & Schaeffel, 2021). As principais motivações para escolha desta estratégia no desenvolvimento do sistema utilizado para o presente projeto foram: (1) Permitir o desenvolvimento da presente tarefa de medida da acuidade visual em uma plataforma agnóstica com relação ao hardware, de forma que a tarefa pudesse ser reproduzida com outras configurações de câmera e iluminação, a fim de facilitar sua eventual reprodução por outros laboratórios, e (2) permitir o registro concomitante em vídeo da região do olho, de forma anônima (sem registrar a região da face do participante), para fins de recalibração após a sessão e para fins de verificar o desempenho das regras de decisão de fixação do software (as soluções de rastreamento ocular mais comuns do mercado hoje não permitem registro em vídeo).

4.3 Procedimento

4.3.1 Cartões de Acuidade de Teller (TAC-II)

Para medida dos valores de acuidade de referência, um conjunto dos Cartões de Acuidade de Teller TAC-II (Teller et al., 1986) foi utilizado. O conjunto completo é composto por 16 cartões de 25,5 cm x 55,5 cm, de cor de fundo cinza, que contém em um dos lados uma grade preta e branca de alto contraste na forma de listras verticais. A luminância do fundo cinza corresponde à média das luminâncias do preto e do branco. Os cartões contam com uma abertura pequena no centro, para que o avaliador possa verificar a direção do olhar da criança durante o teste. Os cartões possuem grades de diferentes frequências espaciais, a primeira delas de 0,23 cpcm (ciclos por centímetro), e a última 26,0 cpcm. A frequência espacial aumenta aproximadamente 50% de um cartão para o próximo. Os valores normativos do teste utilizados para comparação (Salomão & Ventura, 1995) foram ajustados para baixo em meia oitava (0,15 logMAR), seguindo a recomendação de Clifford et al. (2005), que identificaram diferenças sistemáticas entre os valores de AV medidos pelo TAC original e o TAC-II (versão do exame utilizada no presente estudo).

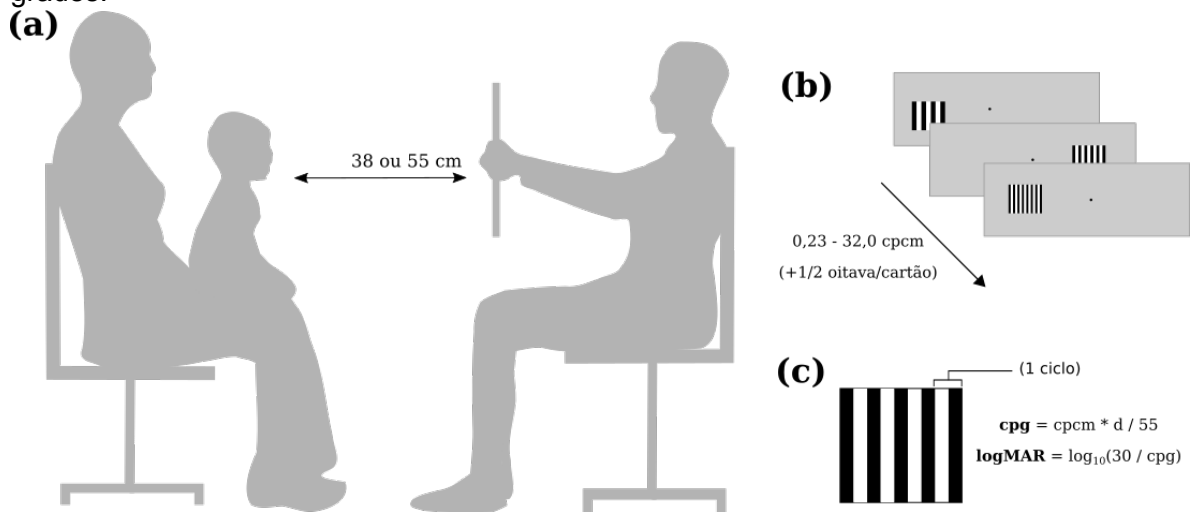
Durante o momento da avaliação, as crianças permaneceram sentadas no colo da mãe e de frente para o avaliador, a uma distância de 38 cm para crianças menores de 6 meses de idade, e a uma distância de 55 cm para crianças com idade maior ou

igual a 6 meses (Figura 2). Em seguida, o avaliador começava a apresentar os cartões, começando de uma frequência espacial mais baixa, e incrementando o valor de frequência espacial a cada apresentação. O cartão inicial escolhido ao início da avaliação possuía frequência espacial variando de 0,86 cpcm até 3,2 cpcm de acordo com a idade da criança. O examinador, cego para o lado ao qual a grade está posicionada, realizava uma série de apresentações do mesmo cartão até chegar a um julgamento sobre qual lado do cartão o padrão está localizado, a partir de pistas no comportamento do bebê, como direção do olhar e orientação da cabeça. Em ocasiões nas quais o bebê podia entender instruções verbais básicas (depois de 24 meses de idade), uma estratégia de olhar preferencial operante foi adotada (D. L. Mayer & Dobson, 1982), e era pedido às crianças para que apontassem para o padrão de grade. Quando o examinador estava pronto para realizar um julgamento sobre o lado da grade, ele sinalizava o lado percebido para um segundo examinador posicionado atrás da criança, que confirmava se o lado sinalizado estava correto ou não. Caso estivesse, o primeiro avaliador aumentava a frequência espacial da apresentação seguinte, e caso não estivesse, o avaliador diminuía a frequência espacial, em um procedimento psicofísico de escada. O intervalo regular entre as apresentações durante o procedimento é de um cartão (meia oitava). Se no começo de um exame o examinador suspeitasse que qualquer cartão poderia estar muito acima do limiar da criança, optava-se por diminuir a frequência espacial em uma oitava inteira (dois cartões). O critério para finalizar um exame era o término de, pelo menos, três reversões, isto é, três pares de cartões onde o julgamento passou de visto para não visto ou vice-versa. O limiar de AV era então estimado como a média geométrica das três últimas reversões, e então convertido para a escala de logMAR.

4.3.2 Tarefa de olhar preferencial automatizado (APL)

A avaliação com o rastreamento ocular ocorreu no mesmo dia da avaliação com o TAC-II. A situação de aplicação da tarefa experimental está representada na Figura 3. O paciente, posicionado confortavelmente junto a um adulto responsável por ele, assistia à apresentação sequencial dos estímulos em um display posicionado a 60 cm de distância. Durante a apresentação dos estímulos, uma câmera, posicionada a 30 cm do rosto do paciente, registrou seus movimentos oculares em tempo real. Para manter equivalência com o exame TAC-II, a tarefa experimental consistiu em tentativas de teste, nas quais um padrão

Figura 2: (a) O avaliador se senta de frente para a criança mostrando os cartões a uma distância de 38 cm (para crianças mais novas que 6 meses de idade) ou 55 cm (para crianças mais velhas que 6 meses de idade). (b) A frequência espacial da grade impressa na frente dos cartões varia de 0,23 ciclos por centímetro (cpcm) até 26 cpcm entre os 16 cartões do conjunto, que são espaçados em $\frac{1}{2}$ oitava entre si. (c) O cálculo do valor de acuidade (em logMAR) considera a distância à qual a criança estava do cartão e a frequência espacial das grades.

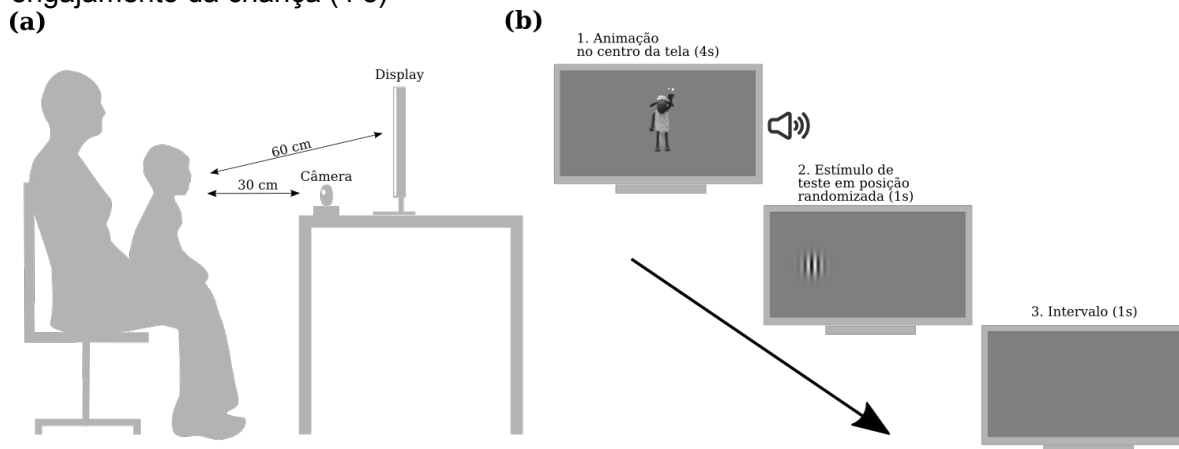


Gabor vertical monocromático de alto contraste e de frequência espacial variável aparece contra um fundo cinza, do lado esquerdo ou direito da tela, aleatoriamente. A criança, após assistir a uma breve animação no centro do monitor, tende a dirigir o olhar para o lado em que o padrão é apresentado, caso este esteja abaixo de seu limiar de detecção, de forma análoga ao exame dos Cartões de Acuidade de Teller, ou tende a apresentar comportamento visual exploratório não-relacionado à posição do estímulo, caso este esteja acima do seu limiar de detecção.

Durante todo o tempo de avaliação, a criança permanecia sentada no colo de um adulto responsável (geralmente a mãe ou o pai, ou algum outro familiar). Ao adulto, foi comunicada apenas a instrução de se sentar de frente para o computador e manter a criança em seu colo durante toda a tarefa. Às crianças mais velhas (no grupo de idade de 24 meses de idade), foi comunicado que ela iria assistir a um desenho animado. Ao início de cada sessão, a criança assistia a uma curta animação do desenho *Shaun the Sheep* (com 1 minuto de duração) com música e personagens coloridos, para que ela se habitue à situação de avaliação e o avaliador pudesse ajustar a montagem da câmera à altura e posição da criança. Após esse período, a tarefa experimental começava. A criança permanecia todo o tempo livre para se movimentar e comunicar com seu familiar ou responsável e com o avaliador. Caso a criança se distraísse com algum outro objeto ou pessoa na sala de avaliação, ou mostrasse pouco interesse na tarefa, o avaliador estimulava a criança a observar a tarefa, apontando para a tela do monitor, ou chamando atenção para a animação que frequentemente aparecia entre os estímulos de teste. O avaliador se mantinha, via de regra, fora do campo visual da criança durante a apresentação dos estímulos, mas eventualmente se colocava no campo de visão da criança para se comunicar com ela, com o adulto responsável, verificar o enquadramento da câmera e a situação da tarefa. O adulto responsável segurando a criança era livre para se comunicar verbalmente com a criança, mas era instruído a não apontar para a tela.

Durante todo o exame, eram apresentadas animações para manter a atenção da criança na tarefa. A testagem psicofísica geralmente impõe altas demandas atencionais por parte da criança, o que a leva facilmente à irritabilidade e falta de cooperação. O objetivo destas animações era o de minimizar este problema, apresentando para a criança um breve vídeo original, animado e com som. Ao espaçar tais tentativas entre as tentativas de teste

Figura 3: Uma distância padronizada de 60 cm do monitor é utilizada para todas as crianças. Sequências de três estímulos de teste são apresentados do lado esquerdo ou direito da tela de forma aleatória (1 s), intercalados por uma dentre quatro animações de um personagem de desenho infantil (*Shaun the Sheep*), acompanhados de uma música, para manter o engajamento da criança (4 s)



de validação, espera-se manter a criança engajada na tarefa pelo tempo necessário para estimação do limiar. Caso a criança apresentasse estado de sonolência e/ou baixo alerta, não colaborasse com a tarefa por tempo suficiente (recusando-se a participar ao apresentar comportamentos como chorar ou tentar de sair do colo do adulto), ou apresentasse movimentação excessiva de modo a prejudicar o enquadramento da região da face no campo de visão da câmera, a tarefa era interrompida e os resultados da criança eliminados do estudo, de acordo com protocolo comunicado no termo de consentimento livre e esclarecido e aceito pelos pais. A razão da não-participação da criança era documentada, para caracterizar e a amostra de crianças que não puderam ser testadas, mas seus resultados foram excluídos da análise comparativa posterior.

Um *script* utilizando a biblioteca matemática padrão da linguagem Python foi utilizado para sintetizar os estímulos de teste. Para este fim, o espaçamento entre pixels do monitor de acordo com as especificações técnicas do fabricante (0,2745mm) foi multiplicado pela quantidade de *pixels* utilizada por cada estímulo (576 *pixels*) para determinar o tamanho de cada padrão em centímetros. Para o cálculo do tamanho do estímulo em graus no campo visual, a fórmula da tangente inversa da razão entre a largura do estímulo pela distância de visualização foi aplicada. Em seguida, uma série de valores de *pixels* por ciclo foi gerada de acordo com a progressão de frequência espacial desejada, todas respeitando aos critérios

de número par de pixels por ciclo (para evitar descontinuidades entre um ciclo e outro), e número par de ciclos ao longo do padrão (para evitar descontinuidades na borda do estímulo). Por fim, os valores dos padrões em ciclos por centímetro foram determinados pela divisão do número de ciclos de cada padrão pela sua largura em centímetros. Os arquivos de imagem gerados a partir do *script* foram salvos em formato rasterizado sem perdas e então carregados e renderizados no monitor utilizando uma placa gráfica Intel HD Graphics (OpenGL 4.5). A tarefa foi apresentada utilizando aplicação desenvolvida com base na biblioteca GTK-4. A aplicação foi desenvolvida de modo que *scripts* na linguagem de programação Lua (Ierusalimschy, De Figueiredo, & Filho, 1996) possam carregados para controlar a apresentação de estímulos e sincronização com o sinal de rastreamento ocular (ver Apêndice 2). Isto permite, por exemplo, que a tarefa desenvolvida no presente trabalho seja distribuída na forma de arquivo de texto, que pode ser reproduzido e adaptado para outras tarefas análogas sem necessitar recompilação da aplicação e de suas dependências nativas.

Os estímulos foram apresentados com correção gama, calculada de acordo com protocolo estabelecido para apresentação de tarefas psicofísicas que requerem especificação precisa do contraste (Pelli & Zhang, 1991). A luminância emitida pelo monitor foi medida a partir apresentação da sequência completa de fundos homogêneos monocromáticos (em escala de 8 bits), a fim de medir a luminância do monitor para todos os níveis de cinza representáveis e para verificar variações de iluminação em diferentes regiões do monitor (Pelli & Zhang, 1991). As medições de níveis de luminância foram realizadas com um luminômetro Konica Minolta LS-100. As rotinas de correção gama implementadas na biblioteca psicofísica PsychoPy (Peirce, 2007) foram utilizadas para determinar os níveis de cinza utilizados para renderizar os estímulos de teste com a correção aplicada.

Uma sequência fixa de estímulos foi apresentada para cada participante. A frequência espacial inicialmente apresentada foi de 0,13 ciclos/grau, e a frequência final foi de 19,18 ciclos/grau. Cada valor de frequência espacial da sequência apresentada representa um incremento de aproximadamente 50% do valor de frequência espacial do estímulo anterior. Incrementos de precisamente 50% a cada nível de frequência espacial não são possíveis, pois a representação adequada de cada ciclo do padrão exige um número par de *pixels* para representar cada padrão, e um número par de padrões precisava ser

apresentado para evitar descontinuidades dos ciclos nas bordas, de forma que os valores finais de frequência espacial são limitados pelo espaçamento entre *pixels* do monitor. Os estímulos foram apresentados em sequências de três estímulos de teste por vez, com duração de um segundo cada, apresentados de forma aleatória na posição esquerda ou direita. Cada sequência destas era intercalada por um estímulo de animação. Essa apresentação era repetida pelo menos quatro vezes; caso a criança não tivesse prestado atenção na tarefa durante a apresentação da tripla de estímulos em nenhuma dessas ocasiões, a apresentação se repetia antes que uma nova tripla com valores maiores de frequência espacial fosse apresentada. Os estímulos de teste tinham duração de 1 segundo cada, e eram espaçados por intervalos de 1 segundo no qual nenhum estímulo era apresentado. Tentativas "lúdicas" eram utilizadas para prender a atenção da criança à tarefa, na forma de uma animação acompanhada de uma música, apresentada sempre no meio da tela por uma duração de 4 segundos. Quatro variações da animação eram utilizadas com diferentes sons para evitar efeitos de habituação durante a execução da tarefa. Estes parâmetros foram decididos com base em tentativas piloto iniciais; nas quais se constatou que a apresentação de estímulos de teste em sequências mais longas (ou com apresentações mais longas; ou com maior número de apresentação de estímulos de teste) falhavam em manter a atenção das crianças na tarefa. O valor de acuidade final medido foi composto pelas três transições de um estímulo rotulado como "visto" para outro marcado como "não visto" com maior frequência espacial.

Para fins de determinar o mapeamento do vetor P-CR para coordenadas de direção do olhar, utilizamos o estímulo de animação, apresentado no centro da tela, e dois dos estímulos de teste, de baixa frequência espacial, do lado esquerdo e direito da tela, como pontos de referência (sendo, portanto, um procedimento de calibração utilizando três pontos de posições distintas do monitor). Os pontos de referência utilizados para calibração foram validados após a sessão a partir de análise do registro de vídeo, e a partir da verificação da ordem relativa entre as coordenadas de fixação esperadas, sendo mantidos apenas estímulos efetivamente fixados pelo participante. Uma interpolação cúbica entre os pontos foi realizada para determinar a posição do olhar nos demais pontos. Para fins de calibração do ponto do olhar durante as sessões experimentais, fixações na sequência inicial de estímulos de baixa frequência espacial e no estímulo de animação foram utilizadas.

Apenas estímulos efetivamente fixados, tal como pode ser aferido pela análise qualitativa sobre o segmento de vídeo gravado, foram incluídos na fixação. Um mínimo de seis pontos de calibração foi utilizado, contemplando cada uma das três áreas de apresentação e cada um dos olhos (registro binocular). Regiões de interesse circulares em volta do estímulo (14 graus) foram delimitadas para contagem do tempo relativo de fixação para cada apresentação.

Durante o desenvolvimento do presente projeto, produzimos uma solução de software completa para rastreamento ocular (Apêndices A e B). As principais motivação para este desenvolvimento foram: (1) Permitir o desenvolvimento da presente tarefa de medida da AV em uma plataforma agnóstica com relação ao hardware, de forma que a tarefa pudesse ser reproduzida com outras configurações de câmera e iluminação, a fim de facilitar sua eventual reprodução por outros laboratórios, e (2) permitir o registro concomitante em vídeo da região do olho, de forma anônima (sem registrar a região da face do participante), para fins de recalibração após a sessão e para fins de verificar o desempenho das regras de decisão de fixação do software (as soluções de rastreamento ocular mais comuns do mercado hoje não permitem registro em vídeo).

Para verificar a capacidade de detecção do algoritmo de análise da posição do olhar, coube ao avaliador examinar o registro de vídeo da sessão, acompanhado do sinal bruto ou calibrado de vídeo-oculografia e marcar cada segmento de vídeo a partir da decisão binária de "visto" ou "não visto", baseado em características como presença de sacadas na direção do estímulo de teste, ou fixação sustentada na direção do estímulo de teste, de modo semelhante ao que é feito com os cartões de acuidade. A principal diferença da técnica aqui (em relação à técnica de olhar preferencial tradicional) é que o avaliador tinha como dados tanto o registro em vídeo da região do olho, quanto o gráfico do sinal bruto de posição do olhar associado a cada apresentação do estímulo (mas não à medida final de tempo relativo de fixação, utilizado na análise automatizada). O avaliador tinha acesso apenas à região do olho sob magnificação, gravada em vídeo monocromático, o que não permitia identificação de outras características que podem ser identificadas na técnica tradicional, como orientação da face, do corpo, ou comportamentos de apontar ou qualquer vocalização da criança. Um valor de AV foi calculado também utilizando esta metodologia, que será denominada VPL (*video preferential looking*) para fins de comparação com a avaliação automatizada. Também aqui,

Figura 4: Montagem dos equipamentos para avaliação no ambulatório de pediatria do HU-Jundiaí



a AV final foi calculada como a média das três últimas transições de um estímulo marcado como visto para um estímulo marcado como não visto.

4.4 Análise de dados

Após extração do sinal de posição do olhar, foram calculadas para cada apresentação o tempo relativo de fixação, ou seja, a proporção de amostras na qual o sinal era registrado dentro de uma região de interesse de circular de 14° envolvendo o estímulo de teste. A detecção do estímulo foi determinada a partir de um critério mínimo de tempo relativo de fixação, aplicado de forma uniforme para todos os participantes. As três últimas transições de estímulos rotuladas como vistas para não vistas foram utilizadas como critério para determinar o valor de AV. Além do sinal de rastreamento ocular, foi registrado o engajamento do paciente com a tarefa, isto é, a proporção de tempo do exame em que a criança estava posicionada na área rastreável e orientada para a tela do computador de

modo que o sinal de posição do olhar pudesse ser registrado. O cálculo desta métrica envolve contar quantas amostras do sinal de posição do olhar foram geradas pelo rastreador, independente da coordenada registrada, e dividir pelo total de amostras que poderiam ser registradas dentro do intervalo de tempo em questão, de acordo com a taxa de amostragem da câmera. São descontados do engajamento total, portanto, intervalos em que a criança realiza movimentos bruscos, piscadas, permanece de olhos fechados, ou orientada para o avaliador, para a mãe ou para outros objetos da sala.

Para determinar a comparabilidade do classificador automático de detecção do estímulo (baseado no tempo relativo de fixação) com os julgamentos de olhar preferencial realizados por um observador treinado de forma independente do critério de duração da fixação adotado, realizamos a análise de olhar preferencial tradicional também sobre os registros de vídeo da sessão de rastreamento ocular, que denominaremos de olhar preferencial por vídeo (*vídeo preferential looking* - VPL). Os registros continham apenas pares de imagens de uma região de interesse retangular em torno dos olhos da criança, não permitindo identificar outras características do participante ou de seu desempenho durante a tarefa que não a direção do olhar. Estes julgamentos foram realizados sem conhecimento de qualquer resultado produzido pela análise automatizada. Tomando os julgamentos da detecção do estímulo como linha de base, foi calculada a matriz de confusão (positivos e negativos falsos; positivos e negativos verdadeiros) para cada critério possível de duração da fixação na região de interesse. Um positivo verdadeiro, por exemplo, era registrado sempre que o observador treinado e o algoritmo registravam uma detecção, e um falso positivo quando o algoritmo automatizado registrava uma detecção mas o observador não. A partir desta matriz, foi realizada análise de curva ROC (Stanislaw & Todorov, 1999), utilizando a estatística de área sob a curva (A') como principal métrica de detectabilidade insensível à variável decisória (duração da fixação). A curva ROC foi construída com base na amostragem de 100 pontos igualmente espaçados dentro do intervalo de possíveis limiares de duração da fixação (de 0 a 1500 ms seguidos à apresentação do estímulo de teste). Um interpolante linear foi utilizado para determinar o valor da função nos intervalos entre os pontos. A área sob a curva (A') foi calculada com base na quadratura numérica da função gerada.

Para comparação das medidas de AV realizadas com o TAC-II e com

o procedimento de olhar preferencial por vídeo-oculografia automatizado (APL) em desenvolvimento, foi calculado o intervalo de confiança sobre as diferenças dos valores de AV medidos nos dois testes para cada participante, bem como aplicado o teste T de uma amostra sobre as diferenças, adotando a hipótese nula de que a diferença média entre as medidas nos dois testes é igual a zero. Os valores da avaliação automatizada foram também comparados com os valores de AV medidos pelo TAC-II por meio de análise de regressão linear simples, sendo o valor do teste tradicional e a idade da criança os dois regressores de interesse principais. A hipótese adotada para o modelo de regressão é a de existência de coeficiente de regressão positivo significativamente maior que zero em ambos os casos. Todos os testes estatísticos de significância de hipótese nula adotaram o valor crítico de 5%. As bibliotecas para análise estatística *SciPy* (Virtanen et al., 2020) e *Statsmodels* (Seabold & Perktold, 2010) foram utilizadas para estimação dos parâmetros de interesse.

5 Resultados

Um total de 60 pacientes foram recrutados para o exame com o procedimento TAC-II e com o procedimento análogo utilizando a vídeo-oculografia (APL). Destes, 38 pacientes terminaram com sucesso o procedimento de vídeo-oculografia (testabilidade de 63%). Dentre os pacientes testados com sucesso, 12 (32%) eram do sexo masculino, enquanto 26 (68%) eram do sexo feminino. Todos os pacientes finalizaram o TAC-II com sucesso. A idade média dos participantes era de 16 meses (mínimo: 2 meses; máximo: 36 meses). Quatro dos pacientes testados eram prematuros (N=2 com idade gestacional de 34 semanas; N=2 com idade gestacional de 36 semanas).

As sessões de avaliação duraram em média 6,05 minutos (DP = 2,13 minutos). O engajamento médio (proporção de tempo em que o paciente permaneceu orientado para o monitor, de forma a registrar sinal de posição do olhar) foi de 52,6% (DP = 17,9%). O engajamento foi significativamente maior durante a apresentação das animações (M=60,3%, DP=19,9%) quando comparado com os intervalos de apresentação dos estímulos de teste (M=47,4%, DP=18,3%; $t = 6,60$, $p < 0,0001$).

Os valores de AV medidos com a técnica de vídeo-oculografia (VOG), junto aos valores do teste tradicional TAC-II, estão apresentados na Tabela 2 para fins de comparação. Os valores de acuidade medidos foram representados graficamente contra os limites de tolerância corrigidos do TAC-II (painel inferior esquerdo da Figura 5), tal como levantados em estudo normativo no Brasil (Salomão & Ventura, 1995). Dentre os 38 participantes testados com sucesso, 28 (73%) apresentaram valores de AV medida pelo teste APL dentro dos limites normativos, sendo que os 10 participantes restantes (27%) apresentaram valores de AV inferiores aos limites normativos. Todos os participantes apresentaram AV medida pelo teste TAC-II dentro dos limites normativos.

O resultado do exame de vídeo-oculografia se mostrou positivamente relacionado com o resultado dos Cartões de Acuidade de Teller ($\beta = 0,25$; $p = 0,02$), conforme pode ser observado no painel superior esquerdo da Figura 5. Uma relação positiva entre a idade do participante e o valor de acuidade medido pelo teste APL também foi constatada ($\beta = 0,018$ decremento na medida em logMAR por mês de idade; $p = 0,0004$), relação que está representada graficamente no painel superior direito da Figura 5. Uma relação positiva similar entre idade e acuidade também foi identificada para o teste TAC-II ($\beta = 0,017$

decremento na medida em logMAR por mês de idade; $p = 0,0002$).

A diferença média do valor de acuidade medido pelos dois testes, entretanto, não foi nula, mas de 0,38 logMAR (IC = 0,27 - 0,48 logMAR; $t=7,11$; $p < 0,0001$). Como diferenças positivas em logMAR representam menor acuidade, esta diferença significa que o teste de VOG em média subestimou a acuidade de cada paciente por este valor, comparado com o valor medido pelos Cartões de Acuidade de Teller. O painel esquerdo da Figura 6 mostra como as diferenças de teste estão concentradas abaixo do que seria esperado na hipótese de equivalência dos testes, enquanto o painel direito mostra o histograma das diferenças dos valores de AV medidos pelas duas técnicas. A magnitude média da diferença encontrada (0,38 logMAR) representa pouco mais que uma oitava de diferença (equivalente a dois cartões do teste TAC-II).

Outros 22 pacientes também foram avaliados com ambas as técnicas, mas seus resultados precisaram ser removidos na etapa de análise, de acordo com os seguintes critérios: paciente estava sonolento ou não manteve atenção na tarefa por estar em baixo estado de alerta ou ter dormido durante a aplicação, não apresentando fixações consistentes nem nos padrões de baixa frequência espacial e nem na animação com som (N=5); criança demonstrou irritação durante a tarefa, chorando e se recusando a participar por tempo suficiente para apresentação de toda a série de frequências espaciais (N=5); criança se mostrou inquieta e pouco colaborativa durante a tarefa, se movimentando (querendo sair do colo do adulto, permanecendo orientada para este, ou não olhando para o monitor) de forma a prejudicar o enquadramento da região do rosto no campo visual da câmera por tempo suficiente (n=10); paciente se recusou a participar da tarefa logo no início (N=2). A idade média dos pacientes que não completaram a avaliação com sucesso foi de 8 meses (mínimo = 2 meses, máximo = 22 meses, $dp = 6.5$ meses), idade significativamente menor do que o grupo de crianças que completaram a avaliação com sucesso ($T=2,75$, $p=0.008$). Sete das crianças (31%) tinham idade inferior a 3 meses. Neste grupo, 12 crianças (55%) eram do sexo feminino, e 10 (45%) do sexo masculino (proporção que não diferiu significativamente da proporção de crianças que terminaram o teste com sucesso, $p = 0,177$).

Nos casos em que a criança se recusou a participar da tarefa por qualquer razão, a sessão era imediatamente interrompida e não era retomada, de acordo com o protocolo de pesquisa comunicado e aceito pela(o) responsável. Ressaltamos que a baixa testabilidade

Tabela 2: Valores de acuidade medidos pela técnica de vídeo oculografia (VOG) e Cartões de Acuidade de Teller

Paciente	Sexo	Idade (meses)	Acuidade (APL)	Acuidade (Teller)	Diferença (Teller-APL; logMAR)
A.M.	F	10	1,18	0,49	0,69
A.P.	F	2	1,15	0,70	0,45
A.S.	M	4	0,86	1,26	-0,40
A.B.	F	8	0,94	0,30	0,64
B.L.	M	36	0,83	0,30	0,54
B.M.	F	12	0,89	0,44	0,44
C.S.	M	20	0,98	0,15	0,84
C.C.	F	20	0,73	0,75	-0,01
C.L.	F	12	0,96	0,44	0,52
E.S.	F	22	0,99	0,39	0,60
F.J.	M	36	0,79	0,25	0,54
G.L.	M	2	1,12	1,31	-0,19
G.S.	M	20	1,25	0,44	0,80
H.S.	F	24	0,51	0,15	0,36
I.B.	F	26	0,96	0,10	0,87
J.F.	F	12	1,01	0,59	0,42
L.O.	F	18	0,79	0,30	0,49
L.V.	F	10	1,09	0,49	0,60
L.Z.	F	8	0,94	0,34	0,59
L.N.	M	18	0,81	0,25	0,57
M.P.	F	14	0,79	0,15	0,65
M.F.	F	20	0,73	0,34	0,39
M.C.	F	22	1,07	0,30	0,78
M.E.	F	2	1,22	0,30	0,93
M.V.	F	16	1,09	0,59	0,50
M.S.	F	6	0,60	0,49	0,11
O.C.	M	16	0,69	0,44	0,24
R.V.	F	36	0,56	-0,00	0,56
R.M.	M	3	1,19	1,26	-0,07
S.L.	M	4	1,01	0,75	0,27
S.G.	M	16	0,53	0,30	0,23
S.M.	M	24	0,42	0,64	-0,22
S.F.	F	6	0,94	0,49	0,44
T.S.	F	8	0,96	0,30	0,67
V.F.	F	10	1,09	0,30	0,80
V.G.	F	30	0,79	0,20	0,59
V.S.	F	36	0,53	0,25	0,28
H.S.	M	28	0,79	0,30	0,49

Figura 5: Resultados da aplicação do teste proposto. Painel superior esquerdo: AV medida pelo teste proposto de vídeo-oculografia (APL) relacionada o teste tradicional de acuidade visual (Teller). Painel superior direito: Relação da AV medida pelo teste proposto com a idade do participante. Painel inferior esquerdo: Comparação dos valores de acuidade visual medidos com o teste computadorizado com os valores normativos de acuidade do TAC-II. Painel inferior direito: Valores de acuidade visual medidos com o teste TAC-II comparados com os mesmos valores normativos.

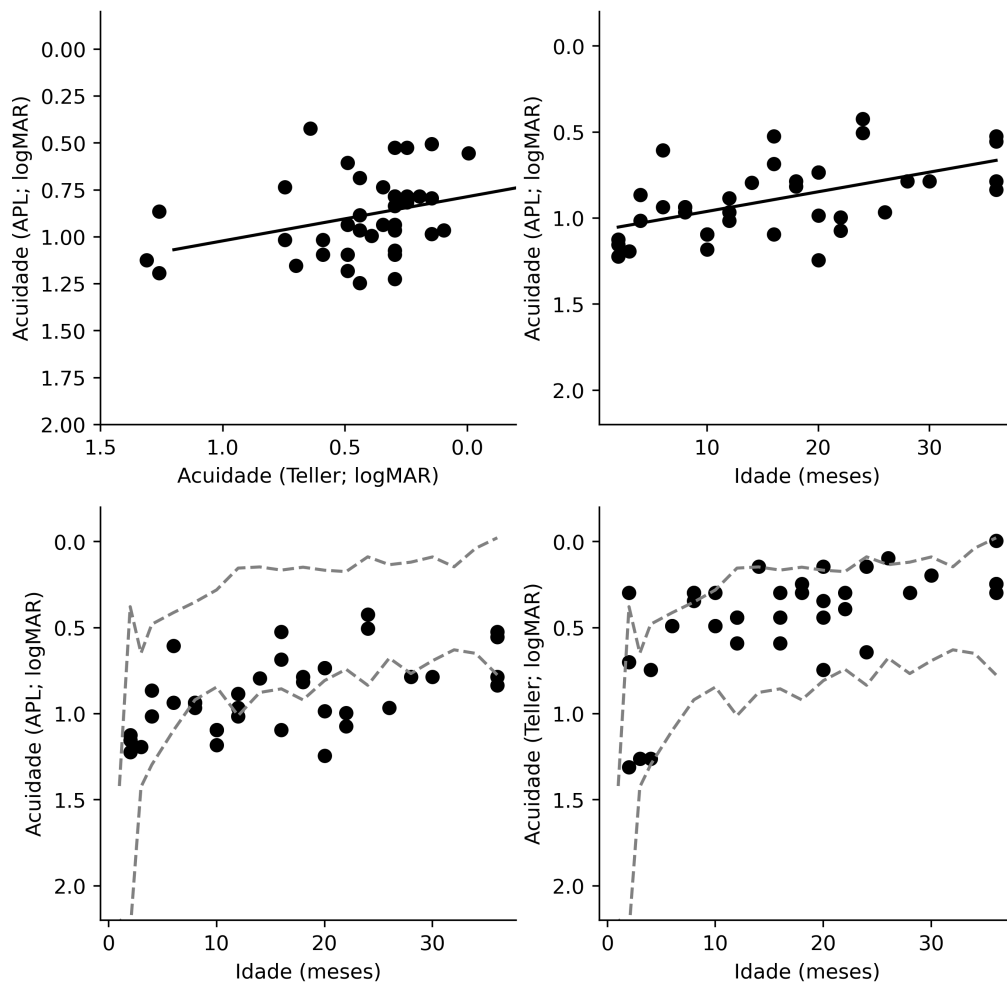
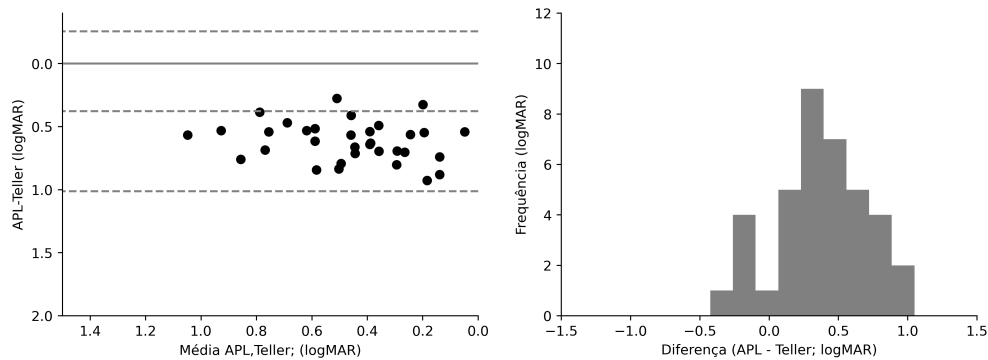


Figura 6: Painel Esquerdo: Representação gráfica de de Bland-Altman representando as diferenças do exame APL e Teller. As linhas tracejadas representam o limite de dois desvio-padrão em torno da média das diferenças (em logMAR). Painel direito: Histograma das diferenças de AV medida pelos dois testes.

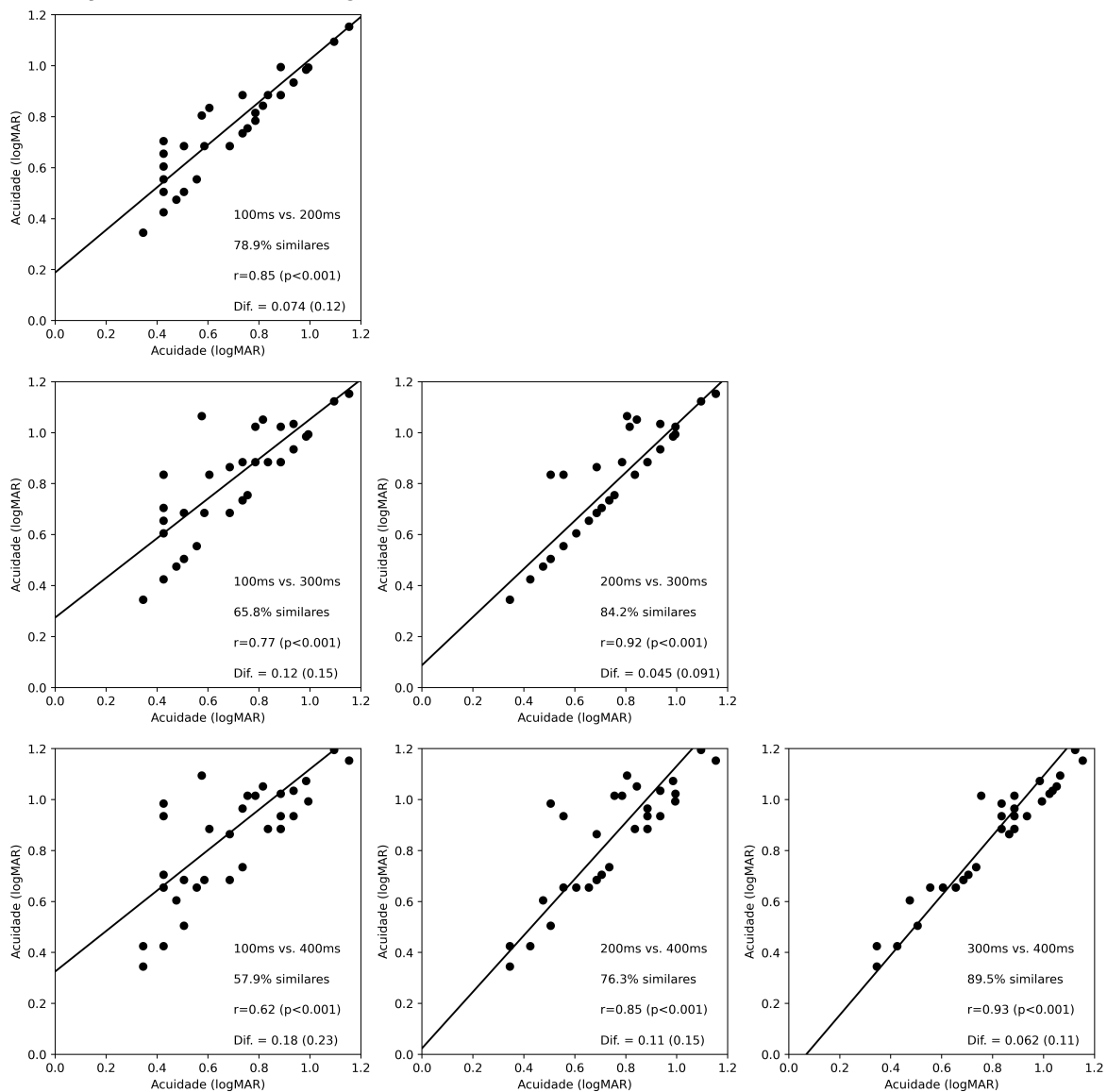


(63%) pode não ser representativa de situações ideais de testagem, pois as avaliações foram realizadas no período da manhã (muitas vezes durante o período em que as crianças mais novas rotineiramente estavam dormindo), e após as crianças já terem passado por extensa avaliação médica e cognitiva, portanto a fadiga pode ter sido um fator contribuinte para a baixa adesão destes pacientes.

A fim de explorar o impacto dos parâmetros da análise automatizada de detecção do estímulo, testamos múltiplos critérios de duração da fixação dentro da região de interesse do estímulo de teste, aplicado de forma uniforme a todos os participantes. A análise automatizada registra uma detecção do estímulo pela criança sempre que são registradas fixações na região de interesse (estímulo de teste) por período maior ou igual a uma determinada proporção do tempo de apresentação. O impacto do critério de detecção automática sobre as estimativas de AV está representado na Figura 7, na qual são mostrados os valores de VA calculados com base em três pontos de corte arbitrários em 100, 200 e 300 milissegundos. Observa-se por meio desta figura que critérios similares levam a alta similaridade entre os valores medidos para a maioria dos participantes, como pode ser constatado pelo coeficiente de correlação calculado entre as medidas de AV para cada par de critérios (região inferior direita de cada gráfico).

A capacidade de detecção do estímulo através da duração relativa da fixação foi avaliada a partir da curva ROC (Figura 8), em uma análise que tomou o limite inferior da duração da fixação como variável decisória. A curva ROC foi calculada com base na

Figura 7: A coluna esquerda compara o critério de duração ≥ 100 ms com os critérios de ≥ 300 ms e ≥ 400 ms. A coluna central compara o critério de duração de ≥ 200 ms com os critérios de ≥ 300 ms e ≥ 400 ms. A coluna direita compara o critério de duração de ≥ 300 ms o critério de duração de ≥ 400 ms. O critério de similaridade representado no rótulo de cada gráfico significa uma comparação em que o par de critérios comparados resultou em diferença menor que 0,15 logMAR.



matriz de confusão gerada a partir das respostas do algoritmo e da avaliação a partir do registro de vídeo por um observador treinado na técnica de olhar preferencial. Observamos a partir desta análise um valor $A' = 0,84$ (superior, portanto, ao critério do julgamento ao acaso em 0,5 e próximo do valor máximo de 1,0), o que sugere a proximidade do critério automatizado com os julgamentos de um observador treinado na análise de olhar preferencial. Procedemos, nas análises seguintes, adotando o critério de duração da fixação correspondente ao ponto da curva ROC com a maior acurácia (ponto com maior taxa relativa de positivos verdadeiros e negativos verdadeiros), que mostrou ser o de tentativas em que o paciente fixou em pelo menos 315 milissegundos na região de interesse, ponto destacado na Figura 8 (correspondente a uma acurácia máxima de 85%). Este critério foi aplicado uniformemente para todos os participantes nas análises posteriores para determinar os valores nominais de acuidade visual por análise automatizada (APL).

A Figura 9 compara os valores de acuidade calculados pela análise automatizada (APL, *automated preferential looking*) com os valores de acuidade calculados com base no julgamento de um observador treinado do registro de vídeo (VPL, *video preferential looking*). Para esta análise, foi adotado o critério ótimo previamente citado na análise automatizada (melhor acurácia, em 315 milissegundos). A análise mostra pouca diferença entre as duas técnicas (média=0,17 logMAR, DP=0,17 logMAR), com valores de acuidade menor que meia oitava (equivalente menos de um cartão de diferença) para 55% dos participantes, e valores menores que uma oitava (equivalente a menos de dois cartões de diferença) para 82% dos participantes. No critério decisório ótimo de duração da fixação adotado, a análise automatizada não difere de um observador treinado na maioria dos casos. Tal análise nos sugere dois pontos: (1) que para a maioria dos casos, a análise automatizada não difere significativamente do que um observador treinado; e (2) que as diferenças encontradas provavelmente não são em função da adoção de um algoritmo automatizado e de um critério uniforme de duração das fixações, mas sim diferenças de procedimento (duração e características do estímulo) e/ou diferentes graus de engajamento das crianças com a tarefa automatizada e com o teste tradicional, que envolve a interação com um avaliador humano.

Figura 8: No gráfico, destaca-se a região da curva na qual observamos maior acurácia global para os participantes da amostra, correspondente ao critério de detecção ≥ 315 ms, critério que foi então aplicado uniformemente para determinar o valor de acuidade de forma automatizada (APL).

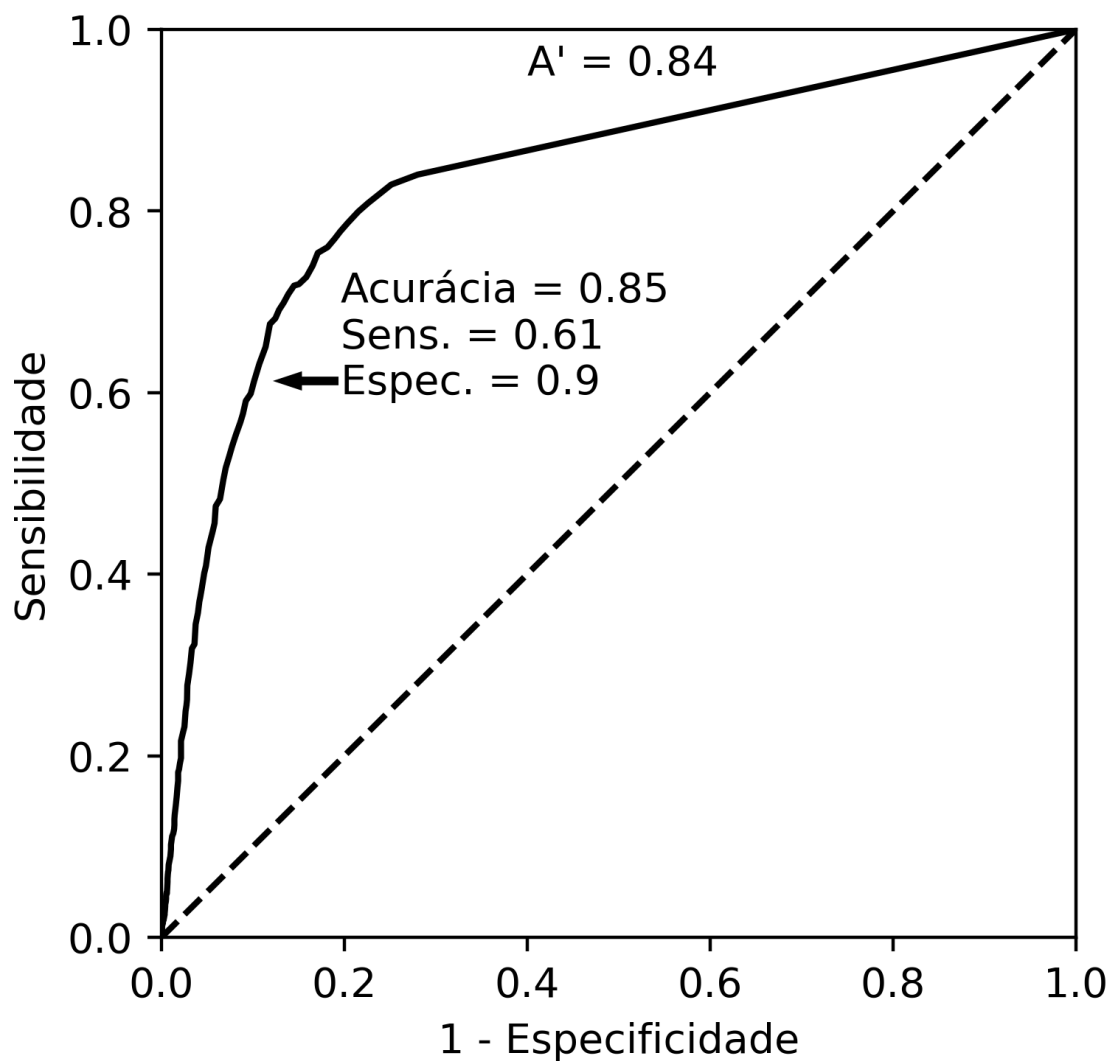
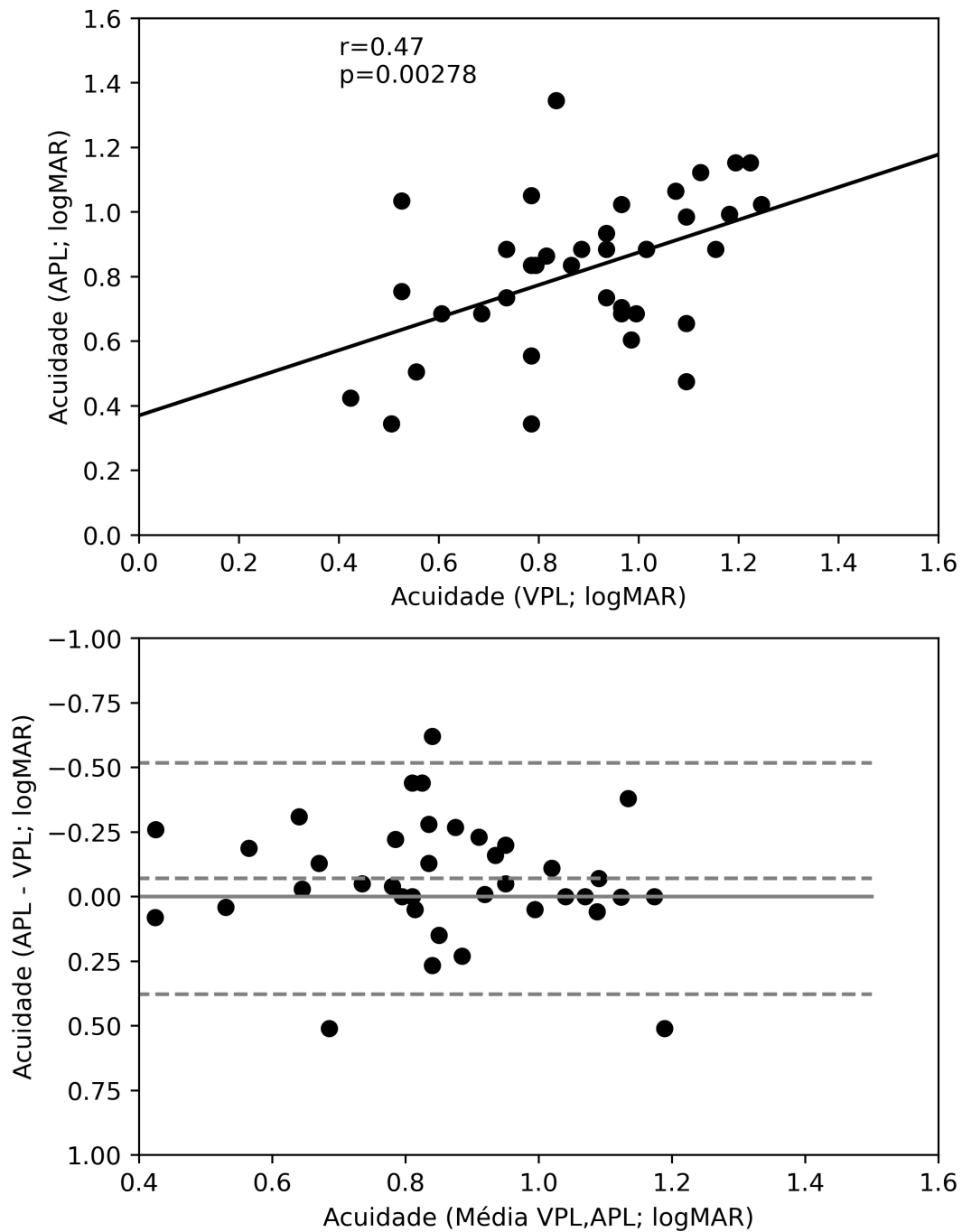


Figura 9: A análise automatizada utilizou o critério de detecção com maior acurácia (≥ 315 ms). Painel superior: Comparação por análise de regressão. Painel inferior: representação gráfica de Bland-Altman das diferenças.



6 Discussão

6.1 Aplicabilidade do programa para testagem da acuidade visual

No presente trabalho, descrevemos o desenvolvimento e aplicação de um programa de computador para testagem rápida da AV por meio de vídeo-oculografia. A maior parte das avaliações realizadas pôde ser completada dentro de um intervalo de 10 minutos, sendo, portanto, um teste de duração comparável com o teste TAC-II. Apesar de problemas com a testabilidade (notadamente pior em crianças mais novas) e das diferenças dos valores medidos pelos dois procedimentos (discutida abaixo), pudemos mostrar que para uma proporção razoável de crianças, que mostram bom grau de engajamento com a tarefa (em especial aquelas mais velhas), o procedimento pode determinar quais estímulos de teste foram detectados ou não pelo participante.

Consideramos que uma contribuição original do presente trabalho foi análise comparativa do desempenho de um algoritmo de detecção de fixações com os julgamentos de um observador por meio do registro em vídeo concomitante ao sinal bruto de rastreamento ocular. Enquanto os trabalhos prévios que se dedicaram a transpor a técnica de olhar preferencial para o meio informatizado utilizaram critérios próprios de identificação da detecção, geralmente baseados em pilotos ou heurísticas específicas (Hathibelagal et al., 2015; Vrabic et al., 2021), adotamos uma análise quantitativa que demonstrou um grau razoável de concordância entre as duas técnicas. Encorajamos que trabalhos futuros adotem abordagem similar, ao permitir uma comparação no nível de análise de tentativas individuais, não se restringindo à comparação dos valores de AV final medidos. Tal técnica pode ser expandida de múltiplas formas: pode-se realizar uma análise comparativa de diferentes critérios de duração da fixação, uso de outras características extraídas do sinal de rastreamento ocular (como sacadas) ou ainda comparações utilizando múltiplos observadores. Julgamos que tal análise se mostrará importante na medida em que a linha de pesquisa em olhar preferencial automatizado evolua para a etapa de normatização, pois a comparação apenas dos resultados finais da avaliação não permite atribuir eventuais diferenças a um conjunto de determinantes relacionados a diferenças na tarefa em si (engajamento da criança, diferenças no regime de apresentação) ou ao algoritmo de análise automatizada.

6.2 Comparação entre os valores de acuidade medidos pelas duas técnicas

O exame proposto resultou em diferenças nos valores de acuidades medidos quando comparado ao teste TAC-II. A análise comparativa entre o valor de AV calculado pela análise automatizada (APL) e os julgamentos de um observador treinado a partir do registro de vídeo (VPL) mostrou diferenças menores que a análise comparativa entre o valor da análise automatizada e o teste tradicional. Os valores de APL e VPL diferiram em 0,17 logMAR em média, com 82% dos pacientes variando em menos que 1 oitava. Já os valores APL e Teller diferiram em 0,37 logMAR em média, com apenas 23% dos pacientes apresentando diferenças menores que uma oitava. Tal resultado sugere que as diferenças não se devem, na sua maior parte, ao caráter automatizado da tarefa (uso de algoritmo padronizado de classificação de detecção), mas sim à diferenças na dinâmica de aplicação da tarefa (uso do computador, menor grau de interação com avaliador, diferenças no conteúdo apresentado, etc.). Trabalhos prévios também encontraram valores de AV calculados por rastreamento ocular que foram inferiores aos valores normativos de um teste tradicional. No trabalho de Jones et al. (2014), por exemplo, 22% dos participantes tiveram valores de AV abaixo dos valores normativos do teste original (em comparação a 27% dos pacientes no presente trabalho).

O tempo que o estímulo permanece na tela é inferior ao tempo de apresentação típico durante o teste dos Cartões de Acuidade de Teller, que é flexível e controlável pelo observador a depender do nível de atenção da criança. Apesar de o tempo de apresentação ser o mesmo para todas as frequências espaciais, é possível que crianças precisem de um tempo maior para fixar estímulos próximos do seu limiar. Aumentar o tempo de apresentação de forma contingente às frequências apresentadas, entretanto, também traz problemas: um maior intervalo para fixação de estímulos de alta frequência espacial pode gerar maior quantidade de falsos positivos (fixações na direção do estímulo ocorrendo ao acaso), o que também pode dificultar a comparabilidade das duas técnicas. Notamos também a diferença dos estímulos utilizados nas duas tarefas. A máscara Gabor pode ter contribuído para tornar o estímulo menos conspícuo, já que o estímulo apresenta contraste comparável com os Cartões de Acuidade de Teller apenas no seu centro (4° centrais), decaindo de contraste na medida em que se afasta do centro, de forma que seu contraste médio é menor do que um estímulo circular de mesmo tamanho sem a máscara de contraste. A principal motivação

para utilização da máscara foi eliminar quaisquer artefatos de borda (que aparecem quando da apresentação do estímulo em tela digital) que permitisse identificar o estímulo, produzindo uma transição suave do padrão para o fundo. O uso de máscaras de contraste de menor borda podem melhorar a detectabilidade dos estímulos.

É possível levantar outras hipóteses sobre fatores que podem explicar esta diferença. Pode ser que haja habituação por parte da criança aos estímulos de teste (crianças tenderão a focar menos os estímulos de alta frequência espacial, mais próximos do final da tarefa, por já terem sido expostas repetidamente a eles). A habituação a um estímulo apresentado de forma repetida é amplamente conhecida na literatura, e é inclusive um paradigma utilizado para estudar o processamento cognitivo em bebês, em função de sua preferência sistemática por estímulos novos em detrimento de estímulos conhecidos (Tafreshi et al., 2014). Não podemos excluir a hipótese de que esse processo se dê de forma diferencial entre os dois testes, pois no teste automatizado, não há presença de um avaliador constantemente conversando e apresentando novos estímulos à criança ao longo da apresentação. Aumentar a ludicidade da tarefa, apresentando estímulos mais variados, com maior frequência e duração, ou torná-los contingentes à detecção do estímulo pode ser uma melhoria a se considerar. Apesar de aumentarem o tempo total de aplicação da tarefa, podem melhorar o engajamento da criança, aumentando o aproveitamento.

Não podemos também afastar o fator fadiga (crianças tenderão a engajar menos com a tarefa com o passar do tempo independentemente de quais estímulos sejam apresentados); fotosensibilidade (a exposição contínua à luz do monitor pode ter gerado baixo engajamento após alguns minutos). Se algum destes fatores (ou combinações deles) contribuíram tanto para a baixa testabilidade constatada quanto para a subestimação dos valores de AV é uma questão ainda em aberto.

Apesar destas diferenças, foi encontrada relação positiva entre a idade dos participantes e o valor de AV medido, o que sugere que o teste consegue capturar a dinâmica de desenvolvimento da AV. Também foi encontrada relação positiva entre os valores de AV medidos pelo TAC-II e pelo teste em desenvolvimento. Dado o pequeno número de pacientes testados, e o fato de não termos incluído crianças com diagnóstico de perdas na AV, a capacidade do teste em capturar diferenças interindividuais de AV clinicamente significativas ainda está por ser estabelecida. Nossos resultados apontam

para a necessidade de refinamentos metodológicos que possam melhorar o engajamento da criança com a tarefa antes de uma validação em larga escala, para uma melhor paridade com o teste tradicional.

6.3 Testabilidade

Foi constatada baixa testabilidade (63%) na amostra de crianças examinadas (22 de 60 participantes não puderam ser testados) pelas razões de baixa cooperação, inquietação durante a tarefa ou reduzido estado de alerta e atenção ao procedimento. Todos os pacientes testados completaram o teste TAC-II com sucesso. A tendência de pacientes que não colaboraram por estarem muito ativos, sem terem demonstrado irritação (se movimentando, desejando se comunicar com o adulto ou avaliador, e mantendo pouca atenção na tarefa) foi observada em pacientes com idade mais avançada (acima dos 6 meses). Já os pacientes que não colaboraram por demonstrar irritação, choro e desejo de interromper a tarefa estão distribuídos por todo o intervalo de idades testado. Notamos a alta proporção de crianças com idades entre 2 e 3 meses que não puderam ser testadas (que compõem 31% deste grupo). Caso apenas crianças com idade superior a 3 meses fossem consideradas para o estudo, a testabilidade do exame aumentaria para 75%, o que sugere que a técnica com rastreamento ocular precise ser reservada para idades um pouco mais avançadas. Apesar de ser tecnicamente possível, a avaliação com rastreamento ocular com crianças com 2 e 3 meses é mais desafiadora, principalmente em ambiente ambulatorial. As crianças estavam realizando a avaliação fora de ambiente familiar, e em horário que usualmente poderiam estar dormindo, o que naturalmente pode ter causado irritabilidade durante aplicação do exame.

Neste quesito, julgamos que possa existir uma vantagem da técnica tradicional: por exigir interação direta com o avaliador, a criança pode aderir melhor à tarefa proposta, pois a apresentação dos estímulos é continuamente interrompida para comunicação entre o avaliador e a criança. O baixo engajamento com a tarefa mostrada no computador não ocorreu para todos os participantes: boa proporção das crianças se manteve engajada na tarefa durante toda a duração, principalmente durante a apresentação da animação, que atraía a atenção das crianças na maior parte dos casos. Existe aqui, portanto, um conjunto de fontes de variação importantes e que podem prejudicar a comparação das duas técnicas de

avaliação, que são o nível de alerta e motivação da criança e seu engajamento com a tarefa. Para o desenvolvimento futuro da tarefa, será interessante melhorar a garantia de manter a atenção da criança na tarefa, utilizando uma extensa variedade de estímulos animados, e interrompendo a tarefa mais frequentemente para comunicação com a criança, de forma a garantir seu melhor engajamento. Outro ponto que pode ser melhorado é garantir que o nível de luminância do monitor esteja confortável para a criança se manter focada durante toda a tarefa.

Um fator adicional que pode ter influenciado os resultados (e este não é um fator particular ao rastreamento ocular, mas compartilhado com métodos tradicionais) é a reduzida capacidade das crianças mais novas de desengajar de um estímulo após a fixação, conhecida por atenção obrigatória na literatura (Hunnius, Geuze, & van Geert, 2006). Após realizar uma fixação, crianças muito novas tendem a persistir na região fixada, mesmo após retirada do estímulo. No teste tradicional, o avaliador mostra cada estímulo por uma duração variável, a depender de sua avaliação sobre como a criança está explorando seu campo visual. No presente trabalho, uma duração fixa de estímulo foi utilizada, a fim de minimizar o potencial aumento de falsos positivos que pode ocorrer com uma apresentação mais longa, e a fim de maximizar a quantidade de estímulos apresentados durante uma sessão. Aumentar o tempo de apresentação de cada estímulo, por sua vez, pode comprometer a aplicabilidade com crianças mais velhas, que podem habituar-se e ficar entediadas com a tarefa caso ela tome muito tempo. O efeito de atenção obrigatória pode ter prejudicado o desempenho de crianças mais novas, que fixavam no estímulo animado altamente saliente, e precisavam realizar uma sacada na direção do estímulo de teste logo após este (com um intervalo de um segundo entre cada estímulo). A avaliação do uso de esquemas de apresentação mais adaptados para cada idade pode ser um objeto futuro de pesquisa para refinamento da técnica.

Limiares psicofísicos medidos em crianças em tarefas que dependem de sua cooperação e atenção voluntária tem sido tomadas como medidas conservadoras, que refletem um limite inferior da sua capacidade, mas que podem não refletir sua capacidade máxima em função de aspectos motivacionais e atencionais (Aslin & Smith, 1988). Neste sentido, propomos como estratégia de pesquisa que pode melhorar a testabilidade do procedimento a utilização complementar de aspectos não-voluntários, como o reflexo

optocinético (nistagmo), que também pode ser medido por meio de técnica de rastreamento ocular. Com o desenvolvimento da técnica, múltiplos procedimentos podem ser aplicados em uma mesma criança, e na hipótese de o valor de AV ser subestimado por conta de um paciente não-colaborativo no paradigma utilizando comportamento voluntário, a medida utilizando o desempenho reflexo possa ser utilizada de forma substitutiva, caso esta se mostre mais alta.

6.4 Limitações

O estudo foi realizado usando um sistema com relativamente baixa taxa de amostragem (60Hz), comparável aos rastreadores mais simples comercialmente disponíveis. Ressaltamos que tais sistemas são adequados para pesquisas cujas métricas são baseadas em tempo relativo de fixação, sendo que sistemas com taxas de amostragem até mais baixas, na faixa dos 25Hz, foram utilizados com sucesso para pesquisa em olhar preferencial (Holmqvist et al., 2011). Sistemas com maiores taxas de amostragem, entretanto, dispõem de medidas temporais mais precisas. Notamos ainda que o software utilizado, assim como qualquer sistema de rastreamento ocular, não está completamente livre de artefatos de detecção da pupila ou do reflexo corneal. Não acreditamos que artefatos tenham impactado severamente nossos resultados por dois motivos: Todas as sessões foram realizadas em salas sem janelas ou qualquer forma de iluminação natural, o que minimiza a presença de reflexos corneais artefactuais que possam ser erroneamente interpretadas pelo rastreador como pontos de referência para cálculo do vetor P-CR. A utilização de um algoritmo de segmentação semântica para delimitar a região da pupila (Apêndice A) também minimiza as chances de artefatos relacionados ao processamento de imagem.

Realizamos um procedimento de calibração relativamente simplificado, utilizando apenas três pontos e interpolação cúbica entre eles, cujas fixações foram validadas posteriormente a partir da análise do registro de vídeo. Apesar de o desempenho global do algoritmo classificador ($A' = 0,84$) não sugerir a existência de viés de mapeamento das coordenadas de posição do olhar introduzido pela calibração, não podemos excluir a presença de vieses em escala menor do que a das regiões de interesse adotadas, que possam eventualmente impactar as estimativas de tempo relativo de fixação, nas situações

em que o paciente estivesse fixando regiões limítrofes do estímulo. Em pacientes adultos, é possível minimizar o viés do procedimento de calibração através de mapeamentos mais completos, mostrando quantidades maiores de pontos, com grades de 9 ou até 16 pontos de referência. Em bebês, tais procedimentos não são aplicáveis, pois dificilmente pacientes toleram e colaboram com procedimentos mais longos e que exigem maior cooperação. Além disso, não foi realizada validação dos pontos de mapeamento da calibração a partir de modelagem geométrica (Hathibelagal et al., 2015), disponível em rastreadores com hardware mais avançado e/ou que realizam a aquisição com restrição de movimento do participante. Julgamos que um procedimento de calibração mais preciso, mas que ainda seja aplicável com crianças em todas as faixas de idade, possa melhorar a ferramenta de avaliação proposta, sendo este um dos pontos que podemos desenvolver para trabalhos futuros. Notamos que a calibração de rastreadores oculares sem a adoção de tais modelos torna as medidas de coordenadas de posição do olhar sensíveis a deslocamentos do participante na direção da câmera, e como bebês se movimentam e se reposicionam muito durante a sessão, pode ser que variações na distância entre o participante e a câmera possam ter afetado a função de mapeamento ocular em momentos específicos da aplicação. Apesar de movimentos serem levemente restringidos pelo fato de ele estar no colo do adulto, nenhuma restrição rígida foi imposta à criança, que podia se movimentar para a frente durante a aplicação da tarefa, o que pode ter reduzido a acurácia da função de mapeamento da posição do olhar nestas situações. Realizar o registro com algum tipo de restrição física (bebê conforto, por exemplo) pode, portanto, melhorar a acurácia das medidas realizadas.

Ressaltamos que o procedimento adotado com crianças não-cooperativas realizado durante a pesquisa não foi o ideal, em função do tempo restrito que cada criança tinha para realizar as avaliações (as crianças estavam em ambiente ambulatorial, realizando consultas médicas e avaliações cognitivas no mesmo dia, e não tinham mais que 30 minutos para realizar ambos os procedimentos de avaliação visual). Na aplicação do TAC-II, a orientação comunicada no manual de aplicação do teste é a de que, para pacientes pouco colaborativos, seja feita a interrupção da avaliação, tentativa de recuperar a atenção e colaboração da criança interagindo com ela fora do procedimento de testagem, para só então retomar a avaliação, o que não era possível dado tais restrições de tempo. Além disso, a aplicação do exame era feita na grande maioria dos casos após a criança ter passado por avaliação

médica e cognitiva, portanto o fator de fadiga da criança (que já poderia estar cansada, com sono ou fome após extenso período de exame médico e testagem) pode ter prejudicado sua colaboração para realização de mais um teste que exigia sua atenção. Portanto, não podemos considerar que a taxa de testabilidade estimada a partir da amostra atual seja particularmente representativa de situações clínicas onde a criança esteja realizando apenas avaliação visual, e tenha mais tempo para realizar outras atividades fora do período de avaliação que melhorem sua cooperação. A aplicação do teste em outros contextos pode, portanto, oferecer estimativas de testabilidade mais representativas.

Outros problemas com testabilidade podem estar relacionados com condições não-ideais de montagem do equipamento. Muitos estudos de vídeo-oculografia com crianças novas utilizam cabines ou painéis laterais de cor neutra em torno do adulto segurando a criança e da criança para minimizar a presença de distratores provenientes de estímulos do ambiente, tornando a tarefa o elemento mais saliente do ambiente (Pel, Manders, & van der Steen, 2010; Sasson & Elison, 2012), condições que não puderam ser replicadas no consultório clínico onde as crianças foram testadas, mas que podem ser arranjadas facilmente em ambiente dedicado à avaliação visual clínica.

6.5 Relação dos resultados com o critério de detecção do estímulo

Dada a ausência de critérios padronizados de detecção de estímulo baseado em vídeo-oculografia na literatura, buscamos em nossa análise adotar um critério comparável ao julgamento qualitativo realizado por um observador treinado, tal como realizado na técnica de olhar preferencial tradicional. Para este fim, construímos uma matriz de confusão comparando os resultados do algoritmo com os do julgamento de um observador treinado, encontrando uma acurácia máxima de 85% (sensibilidade de 61% e especificidade de 90%) ao adotar o critério de tempo relativo de fixação maior ou igual a 315 milissegundos na região de interesse. A estatística A' (calculada no presente trabalho em 0,84) representa um critério de detectabilidade independente do critério de duração da fixação utilizado, e pode servir como um índice para medida do impacto de futuras manipulações experimentais (tamanho do estímulo e da região de interesse, tempo de apresentação, idade e características clínicas do participante), bem como para a comparação de resultados de estudos similares. Apontamos que tal abordagem é um caminho interessante para trabalhos futuros de

validação, na medida em que permite, em certa medida, explicitar o grau do impacto dos algoritmos utilizados (e dos respectivos parâmetros) nos valores de AV medidos. Tal análise pode ser ainda enriquecida através do emprego de múltiplos observadores.

A avaliação automatizada requer uma série de decisões aparentemente arbitrárias na etapa de análise dos resultados, que irão influir nos valores de AV calculados. Essas escolhas são ainda mais importantes caso procedimentos psicofísicos adaptativos sejam aplicados, pois os critérios não podem ser revistos na medida que a escolha de quais estímulos serão apresentados é realizada em tempo real. Vrabic et al. (2021) reportam os valores para dois limiares de corte de tempo de fixação em cada um dos estímulos de interesse. Outros trabalhos reportam os critérios de escolha, mas não discutem as consequências dessas escolhas. Estas variações se aplicam também aos exames tradicionais de AV, pois cada avaliador realiza uma avaliação subjetiva sobre uma série de fixações, com a diferença que estes critérios nem sempre são formalizados, ou são formalizados com diferentes graus de precisão. Por exemplo, enquanto o manual do TAC-II não impõe limites na quantidade de apresentações, o estudo de validação de Salomão & Ventura (1995) utiliza pelo menos três reversões.

É sempre possível que um avaliador treinado julgue o registro de rastreamento ocular a partir de sua representação gráfica. Esta foi por exemplo a abordagem adotada por Kooiker (2016) para avaliação da motilidade ocular, na qual os avaliadores (cegos quanto ao paciente avaliado e quanto aos resultados de avaliações para diferentes procedimentos de um mesmo paciente) precisavam tomar uma decisão binária entre "impactado" ou "adequado" para verificação da motilidade ocular de cada criança, em uma abordagem que pode ser considerada semiautomatizada. Os mesmos autores utilizaram critérios objetivos para detecção de cada estímulo (abordagem totalmente automatizada), utilizando direções de sacadas concordantes com a posição dos estímulos, duração e estabilidade das fixações (fixações precisavam ter pelo menos 2 segundos), mas não reportaram os limiares (ou outro critério) para cada uma dessas medidas utilizadas para determinar a detecção ou não do estímulo.

Critérios baseados em tempo relativo de fixação, tal como o adotado no presente estudo, são potencialmente sensíveis à idade da criança e ao seu grau de controle oculomotor. O comportamento visual exploratório de linha de base (observado na ausência

de grade) pode variar com base em considerações como o fato de crianças mais novas terem dificuldade em desengajar de um estímulo após fixado (Hunnus, 2007; Butcher et al., 2000), ou o fato de crianças mais velhas terem padrão de busca visual mais ativo, podendo desengajar de um estímulo fixado mais rapidamente (Regal et al., 1983). Uma saída possível é adotar critérios que sejam ótimos para determinadas faixas etárias, a partir de ampla análise empírica do impacto diferencial destes critérios sobre os resultados.

Outros tipos de critérios que não a duração relativa da fixação são encontrados na literatura, como os de Chang et al. (2021), que reportam que sacadas foram rotuladas quando a velocidade do movimento ocular superou $30^\circ/\text{seg}$, aceleração maior que $8000^\circ/\text{seg}^2$ e amplitude maior que 0.1° . Fixações foram definidas como eventos quando os olhos moveram $< 0.1^\circ$ por duração maior ou igual a 100 milissegundos. Para determinar que um padrão tivesse sido visto, ele precisaria ser fixado preferencialmente pela criança 2/3 das vezes em que cada estímulo foi apresentado. Olhar para o padrão preferencialmente acontecia quando pelo menos 50% do tempo de todas as fixações de cada tentativa estivesse na região de interesse do padrão. Hathibelagal et al. (2015) utilizaram como critério a seguinte sequência de decisões: (1) se a primeira fixação fosse na área-alvo por pelo menos 2 segundos; caso contrário, (2) se houvesse uma sequência de fixações por pelo menos 2 segundos dentro da área alvo dentro de uma janela de apresentação de 6 segundos, e (3) caso a proporção de tempo olhando para o alvo fosse de pelo menos 75% durante toda a janela de apresentação. Uma decisão com base em visto ou não visto era feita por um observador baseado na visualização de gráficos gerados pelo sistema de rastreamento ocular em tempo real.

Jones et al. (2015) estudaram a questão do critério de detecção com profundidade através de estudo de simulação. Os autores começam notando que uma tarefa de rastreamento ocular não pode ser formalizada como de escolha forçada. Na tarefa de escolha forçada, a probabilidade de acerto ao acaso pode ser quantificada como $1/n$, onde n é o número de escolhas, o que dá margem para escolha da proporção de tentativas corretas que serão definidas para determinar o limiar. Na tarefa de rastreamento ocular, um erro (lapso) não pode ser quantificado em uma quantidade pré-determinada de opções (o paciente pode olhar para outra região do monitor, pode realizar sacadas não relacionadas ao estímulo), todas com probabilidade desconhecida. Uma primeira análise para delineamento

da tarefa psicofísica se baseia nas taxas de lapso dos pacientes (probabilidade de não fixar dado que o estímulo está acima do limiar). Baseados em seu estudo, os autores calcularam a taxa de lapso em 25,9%. Em compensação, as crianças também apresentaram taxas de acerto ao acaso (probabilidade de fixar o estímulo quando ele está acima do limiar) mais baixas ainda (6,9%), provavelmente devido ao arranjo experimental dos autores, que posicionavam o estímulo em qualquer posição de uma circunferência no monitor, e exigiam que a fixação ocorresse em um curto intervalo de tempo. Dadas estas características, os autores determinaram que a proporção de acertos adequada para a tarefa de rastreamento ocular seria de 40%, e que uma taxa de acerto de 33% gerou valores de AV muito mais próximos dos valores normativos do que uma taxa mais tipicamente utilizada na psicofísica com adultos, de 71%. Os autores também recomendam utilizar como procedimento final não a média das últimas reversões, mas o valor máximo atingido durante a sessão, uma métrica que é menos afetada pela falta de atenção, pois ela depende apenas de a criança ter mantido a atenção durante algum ponto da tarefa, enquanto outras regras são menos robustas a falta de atenção ao final da tarefa, o que tipicamente ocorre com crianças. A ressalva é a de que esta regra só pode ser aplicada quando existem baixíssimas taxas de acerto ao acaso, caso contrário ela pode superestimar o limiar.

A análise de olhar preferencial tradicional impõe poucas restrições sobre os critérios que o observador utiliza para realizar um julgamento de direção do olhar. Enquanto o método tradicional da escolha forçada (*forced preferential looking* - FPL) exige que o observador realize um julgamento sobre esquerda ou direita, o método dos cartões de acuidade não impõe tal exigência, sendo que o observador realiza o julgamento sobre "visto" ou "não visto" a partir de múltiplas dicas que podem não ser facilmente descritas pelo observador. A utilização de um equipamento de rastreamento ocular, por outro lado, exige que uma decisão seja tomada a partir de uma representação ainda mais abstrata, baseada no sinal de posição do olhar durante uma janela temporal de apresentação do estímulo. Enquanto a técnica aqui se aproxima do procedimento FPL, no sentido de que uma decisão binária precisa ser tomada a cada apresentação (o que não existe no procedimento TAC-II, no qual múltiplas apresentações podem ser realizadas antes que o aplicador tome uma decisão), precisamos, aqui, apontar a seguinte assimetria no critério de julgamento, que difere da técnica FPL: A partir de um sinal calibrado, a completa ausência de amostras do sinal de

ponto do olhar sobrepostos à região de interesse do estímulo permite uma decisão de "não fixado" de forma não-ambígua (excluindo limitações de acurácia do equipamento), mas o contrário não se aplica (isto é, marcar um estímulo como "fixado" depende de uma decisão metodológica). É inteiramente possível que durante o comportamento visual exploratório da criança, algumas amostras estejam sobrepostas à região de interesse do estímulo, sem que este seja efetivamente fixado, situação que seria marcada como fixação caso o critério adotado de duração mínima da fixação seja mais liberal. Da mesma forma, a criança pode realizar fixações breves, que não sejam registradas como tal caso o critério temporal seja mais conservador. Nota-se, portanto, a vinculação dos valores de AV medidos ao critério de detecção adotado. A comparabilidade dos resultados de diferentes sessões, entretanto, exige a uniformidade no critério adotado.

No presente estudo, analisamos o desempenho do algoritmo de detecção utilizando métricas da teoria da detecção do sinal (Stanislaw & Todorov, 1999), comparando o resultado do algoritmo com as decisões de um observador treinado. Nosso resultado, de $A' = 0,84$ e maximização da acurácia do classificador em 315 milissegundos, pode não ser universalmente aplicável para outras tarefas experimentais. Notamos especialmente a dependência deste critério ao tamanho do estímulo e à duração da apresentação, já que estímulos maiores e apresentados por um intervalo maior de tempo implicam maior risco de falsos positivos. Ao adotar o julgamento de um observador treinado como a linha de base para elaboração da matriz de confusão a partir da qual o critério de detecção é levantado, pode-se argumentar que o resultado de uma aplicação do teste dependerá, em última instância, de um julgamento subjetivo. Quanto a este ponto, ressaltamos que ainda que haja tal componente dependente de observador (ou do consenso de múltiplos observadores) na elaboração do critério, o procedimento ainda mantém a uniformidade do critério adotado para múltiplos participantes testados, diferentemente do teste tradicional.

6.6 Variações procedimentais na técnica de rastreamento ocular

Os resultados da avaliação podem ser impactados não apenas pelo critério de detecção, mas também por parâmetros de processamento da imagem e função de mapeamento da posição do olhar. Diferentes equipamentos utilizam diferentes algoritmos para gerar tal mapeamento. Mesmo que múltiplos equipamentos e sistemas de software

resultem em valores comparáveis de acurácia e precisão em suas especificações, o grau de adequação de cada sistema à avaliação com a população infantil precisa ser criticamente examinado. Procedimentos de calibração (cálculo da função de mapeamento da posição do olhar) tipicamente utilizados com adultos possuem problemas ao serem aplicados com crianças muito novas (de 2-24 meses). Como crianças nesta faixa etária colaboram com a tarefa experimental por tempo limitado, corre-se o risco de exaurir a capacidade de cooperação da criança já durante o procedimento de calibração, caso este precise ser repetido por um tempo mais extenso, ou precise ser re-aplicado em fase mais avançada da tarefa, quando a criança já pode estar pouco colaborativa (Harris et al., 1981). Observamos este problema na aplicação durante o desenvolvimento do presente trabalho, no qual as crianças, a despeito de prestarem atenção ao estímulo animado e com som, apresentavam baixa atenção à tarefa experimental durante os estímulos de teste. Estes estímulos muitas vezes estarão próximos ou acima do limiar da criança, portanto apresentando intervalos relativamente longos sem conteúdo que atraia a atenção da criança para o monitor. A movimentação e reposicionamento excessivo por parte de alguns participantes (principalmente na faixa etária dos 12-18 meses) também pode causar rápida degradação da calibração, que precisa ser continuamente revista, tomando tempo que poderia ser dedicado à tarefa experimental.

Os procedimentos de calibração disponibilizados por rastreadores oculares comerciais possuem algum grau de falibilidade com a população infantil, e inclusive já chegaram a ser reportados em alguns dos trabalhos aplicando a técnica de olhar preferencial. Esteban-Ibanez (2021) excluiu de seu estudo participantes que não puderam completar o procedimento de calibração (o autor realizou excluiu 3 dos 60 participantes para o teste de acuidade, e 16 dos 60 participantes para o teste de sensibilidade ao contraste, mas não especificou quantos dos excluídos foram em função de falhas na calibração, que foi um dos critérios de exclusão na metodologia do trabalho). Kooiker et al. (2016), que usaram um sistema comercial Tobii, realizaram o procedimento de calibração, mas nos casos em que a calibração falhou, os autores aplicaram um procedimento de re-calibração offline, no qual as fixações eram remapeadas manualmente por um examinador. Esta pode ser uma solução que garanta maior testabilidade do procedimento, apesar da desvantagem de exigir mais tempo para análise dos resultados, o que pode ser impraticável se houver alto número de

pacientes e/ou alto número de exames realizados com uma mesma criança. Além disso, requer um avaliador treinado para realizar o procedimento de calibração, e tal procedimento pode não ser facilmente generalizável de um hardware para outro, ou aplicável em situações clínicas em que não haja equipe treinada para realizar o procedimento.

Jones et al. (2015) desenvolveram uma abordagem original para o procedimento de calibração em seu trabalho. A abordagem, denominada pelo autor de *"bootstrapping"*, pode ser realizada em poucos segundos e exige fixação em apenas um ponto (em contraste com procedimentos com calibração em adultos, que usam cinco ou nove pontos). A técnica se inicia com um modelo de mapeamento do olhar de propósito geral, não calibrado para o participante atual, oferecido pelo sistema comercial que os autores utilizaram. A suposição básica do método é de que a calibração para cada participante difere sistematicamente deste modelo de propósito geral apenas por um vetor de translação nas direções horizontal e vertical. Os autores constataram que este vetor é menor na dimensão horizontal, a partir da qual as coordenadas da primeira fixação são calculadas. Um modelo probabilístico de detecção do estímulo baseado apenas na dimensão horizontal é construído para esta finalidade. Caso o participante fixe o olhar no estímulo, um modelo probabilístico normal com média na posição central do estímulo apresentado prevê a posição do olhar razoavelmente bem; caso contrário, um modelo probabilístico baseado na distribuição uniforme com limites inferior e superior em *pixels* cobrindo toda a largura do monitor é utilizado em seu lugar. A tentativa é considerada um acerto (ou seja, um olhar na direção do estímulo) quando a razão de verossimilhança entre estes dois modelos é superior a um limiar estabelecido previamente. O estímulo é apresentado sucessivamente até que uma detecção seja realizada. Após a primeira fixação no estímulo, um vetor de erro de translação nas duas coordenadas é calculado baseado no deslocamento da direção do olhar registrada e o centro do estímulo. Todas as fixações seguintes nos estímulos de teste são utilizadas para refinar o vetor de correção do erro de calibração, de forma que a calibração é reajustada continuamente durante a sessão. Os autores não chegaram a publicar uma análise formal da acurácia desta técnica, mas relatam que a julgaram adequada por inspeção visual do olhar em tempo real sobreposto sobre a tarefa experimental. Esta parece ser uma abordagem promissora, mas ainda carece de melhor especificação para que possa ser reproduzida.

Um trabalho que propõe uma vantagem metodológica significativa neste sentido

é o de Hathibelagal et al. (2015), que utiliza solução de mapeamento do ponto do olhar que dispensa calibração. Este trabalho se baseia em tecnologias de rastreamento ocular baseadas em modelos geométricos tridimensionais, que dispensam a calibração quando duas câmeras são utilizadas (Guestrin & Eizenman, 2006). Esta pode ser uma interessante via de pesquisa para futuros trabalhos; salvaguardadas as necessidades de calibração da câmera (que não exige colaboração do participante, mas exige maior conhecimento técnico para montagem) e o requisito de que duas câmeras precisam ser utilizadas. Nenhum dos trabalhos de aplicação da técnica de olhar preferencial até o momento utilizou o método do invariante de razões cruzadas (Yoo & Chung, 2005), que não exige nem colaboração do participante, e nem calibração prévia da câmera, sendo outra via de pesquisa interessante para futuros trabalhos. Por fim, existem outras técnicas de mapeamento de ponto do olhar baseadas em modelos que não usam regressão de pontos fixados: Brillhault et al. (2019) propõem o uso de algoritmos de agrupamento para calibração de rastreadores oculares aplicados com crianças e primatas não-humanos. A técnica permite ignorar pontos que não sejam os de interesse para o processo de calibração. Esta abordagem funcionará desde que a criança fixe os estímulos de calibração com alta probabilidade, relaxando a restrição de determinar de fato quais estímulos foram fixados ou não.

No presente trabalho, procuramos realizar uma calibração utilizando os próprios estímulos da tarefa para a calibração, nos quadros em que o avaliador julgasse que o estímulo tivesse sido visto, um procedimento similar ao realizado por (Kooiker et al., 2016). Para que isto fosse possível, a região ocular foi gravada juntamente com as marcações temporais e rótulos para cada estímulo apresentado. Após escolhidos os quadros, a calibração pode se aplicar retroativamente aos estímulos apresentados até então. A principal desvantagem deste procedimento (na forma como ele foi aplicado até aqui) é que ele preclui a realização de procedimentos adaptativos, nos quais a calibração precisa ser aplicada em tempo real durante a tarefa. É possível, entretanto, realizar esta marcação em tempo real, durante a tarefa (tal como feito por (Hathibelagal et al., 2015)), se o avaliador puder examinar o registro da posição do olho na medida em que a sessão é realizada, mas a situação de aplicação (disposição da sala, participante e examinador) não permitiu que isso fosse realizado no presente trabalho sem que o avaliador distraísse a criança durante a tarefa.

Por fim, paradigmas que usam a vídeo-oculografia, mas usam respostas reflexas ao

invés de respostas voluntárias, podem ser uma opção que dispensa totalmente a calibração (D. S. Lima & Ventura, 2023). A detecção de padrões típicos de reflexos oculomotores como o nistagmo optocinético não exige calibração, mas apenas a análise da forma do sinal de posição do olhar em sua forma bruta. A técnica de vídeo-oculografia já vem sendo utilizada para avaliação da função visual com base em respostas reflexas em adultos ao invés das respostas voluntárias exigidas pela técnica de olhar preferencial. Hyon et al. (2010) avaliou 83 pacientes adultos em um paradigma de nistagmo optocinético, no qual os pacientes precisavam fixar em grades movendo-se continuamente da direita para a esquerda, com o objetivo de detectar a menor frequência espacial que eliciava o reflexo (paradigma de indução) ou o menor tamanho de um objeto ocluser das grades que inibia o reflexo (paradigma de supressão), encontrando bom grau de concordância com testes de acuidade visual tradicionais. Resultado similar foi encontrado por Shin et al. (2006), que também testaram pacientes adultos afetados por uma série de condições como erros refrativos, glaucoma, maculopatias e neuropatias ópticas com a vídeo-oculografia, encontrando altas correlações com o relato subjetivo do paciente de detecção de optótipos. Estes trabalhos sugerem que talvez a testabilidade da vídeo-oculografia para fins de detecção do limiar possa ser maior em crianças não-colaborativas caso respostas reflexas sejam utilizadas ao invés das respostas voluntárias exigidas pelo paradigma de olhar preferencial. Dentre as primeiras técnicas para determinação da acuidade visual em população infantil, estava o tambor para eliciação da resposta optocinética (Lee R. Wolin, 1964), no qual a criança era posicionada dentro de um aparato circular, que tinha sua superfície interna revestida com padrões de frequência espacial variável, que capturavam todo o campo visual da criança. A presença ou ausência de nistagmo optocinético era então determinada por um observador para cada frequência espacial. A técnica de vídeo-oculografia pode ser facilmente aplicada em situação similar, posicionando a criança de forma mais próxima do monitor, e realizando o registro das respostas reflexas, mas não é de conhecimento do autor nenhum trabalho que tenha realizado este procedimento.

Uma questão que precisa ser considerada durante o processo de validação é a replicabilidade do procedimento por diferentes equipamentos de rastreamento ocular. No presente trabalho, buscamos criar uma solução de software que seja relativamente agnóstica com relação a hardware, utilizando câmeras USB de propósito geral, tal como realizado

em outros trabalhos (Matsuda et al., 2017; Hosp et al., 2020; Ivanchenko et al., 2021). A principal motivação para tal escolha é o desenvolvimento de protocolos experimentais com base em especificações técnicas reprodutíveis da *pipeline* de processamento de imagem, desvinculadas de soluções comerciais específicas (que estão sujeitas à descontinuidade e/ou revisões de hardware e software fora do controle do pesquisador). Na medida em que estudos futuros de olhar preferencial por vídeo-oculografia venham a utilizar outros equipamentos, a existência de especificações dentro das quais uma tarefa precisa operar será necessária para garantir a reprodutibilidade do procedimento. Idealmente, qualquer sistema de rastreamento ocular poderia ser capaz de executar um determinado conjunto de protocolos de avaliação, desde que atingidas um conjunto de especificações estabelecidas.

Para fins de validação e aplicação clínica, é importante que o software adotado seja em certa medida independente do hardware utilizado para aquisição, ou pelo menos que existam especificações técnicas replicáveis por múltiplos sistemas, caso contrário qualquer dado de validação pode ser condicionado aos procedimentos de processamento de imagem e calibração do sistema utilizado por um sistema particular. Isso impõe certos riscos: O software utilizado pelo sistema pode não ser mantido, ou pode passar por alterações durante seu ciclo de desenvolvimento que invalidem o procedimento futuramente. Não existem ainda especificações técnicas de vídeo-oculografia com a população infantil para a técnica de olhar preferencial, sendo as vantagens relativas das diferentes técnicas existentes com esta população ainda objeto de estudo.

Existem vários modelos de câmeras de custo relativamente baixo que atendem ao protocolo GenICAM, o que a princípio facilita a execução de estudos de validação em maior escala. Para que o programa de avaliação desenvolvido no presente projeto pode ser executado com sucesso por outros sistemas, um conjunto de especificações técnicas ainda precisam ser levantadas (luminância do ambiente, tipo de display, resolução e taxa de quadros da câmera), para determinar com melhor precisão as condições sobre as quais um eventual estudo de validação em larga escala se aplicará, separando os aspectos essenciais que são objeto da validação (como características do estímulo e dinâmica da tarefa) de aspectos acidentais (como equipamento utilizado). É necessário também estimar a precisão do algoritmo de rastreamento do olhar (expressa em graus de ângulo visual) para diferentes metodologias de calibração que podem ser adotadas para a tarefa experimental,

para determinar os limites aceitáveis de tamanho do estímulo da tarefa. Estas questões são importante objeto de pesquisa futura, pois estudos de validação precisam, em certa medida, balancear a necessidade de padronização da tarefa experimental com a variedade de soluções técnicas possíveis para o problema de rastreamento ocular.

6.7 Engajamento do participante com a tarefa

Diferentemente da avaliação por rastreamento ocular com adultos, com a população pré-verbal não se pode pressupor a colaboração da criança por toda a duração do experimento, sendo os registros reflexo do comportamento visual exploratório espontâneo da criança. Na ausência da possibilidade de estabelecer instruções explícitas para o engajamento da criança com o experimento (ao menos para as crianças mais novas), reportar alguma métrica de engajamento (tempo que a criança permanece orientada para a tela do experimento) se torna uma importante dimensão na avaliação dos resultados. A presença de variações no grau de engajamento de crianças com mídias audiovisuais durante seu curso de desenvolvimento são bem estabelecidos na literatura: Em um experimento que examinou o engajamento espontâneo com conteúdo televisivo, foi constatado um claro e significativo aumento tanto da frequência de orientação para a televisão quanto do tempo assistindo ao conteúdo entre os 12 e 38 meses de idade (10% do tempo assistindo televisão para crianças com 12 meses de idade contra 56% do tempo para crianças com 48 meses de idade) (Anderson & Levin, 1976). Existem também fortes evidências de correlação do tempo de engajamento com o tempo que as crianças usualmente passam assistindo televisão em casa (Anderson & Levin, 1976; Barr, Zack, Garcia, & Muentener, 2008). O conteúdo apresentado possui papel importante na manutenção do engajamento, e a presença de conteúdo com significado linguístico têm impacto positivo para manter o engajamento de crianças com grau mais avançado de desenvolvimento da linguagem (Pempek et al., 2010). Na presente tarefa, introduzimos um tempo de aclimatação com a situação experimental ao apresentar um conteúdo animado e com música pela duração de um minuto, previamente à apresentação dos estímulos de teste. Tanto o adulto quanto o experimentador se comunicavam livremente com a criança durante este período. Introduzir um período de aclimatação mais elaborado, com dinâmica mais próxima da tarefa (porém com conteúdo mais engajador), e que se seja apresentado não apenas no início, mas também de forma

intercalada com a tarefa (permitindo maior grau de comunicação e engajamento ativo da criança) também pode ser uma direção relevante a ser explorada.

O fato de existir um viés atencional para estímulos socialmente relevantes (Johnson, Dziurawiec, Ellis, & Morton, 1991) sugere que o engajamento da criança pode ser melhor na sua interação com um avaliador presente. Entretanto, a avaliação psicofísica exige a apresentação de estímulos relativamente padronizados, sem qualquer conteúdo com maior saliência competindo pela atenção da criança. No procedimento clássico de olhar preferencial por escolha forçada, há menor interação da criança com o avaliador, pois este permanece por trás de um aparato. O procedimento clínico TAC-II, por outro lado, propõe que haja maior grau de interação com o avaliador, ocorrendo de forma intercalada com os estímulos de teste. Presumivelmente, este arranjo garante o bom engajamento da criança mesmo durante a apresentação dos estímulos de teste, gerando a expectativa da próxima interação. A presente tarefa reproduz na medida do possível tal filosofia, apresentando um estímulo animado de forma intercalada com os estímulos de teste. Se uma paridade maior com o procedimento dos cartões de acuidade é o objetivo, uma direção relevante para pesquisas futuras utilizando estímulos apresentados por computador pode ser a de introduzir variações experimentais em tais intervalos: aumentar o grau de relevância social dos estímulos (mostrar filmagens de pessoas ao invés de animações, com conteúdo na língua natural da criança), aumentar a saliência e/ou variedade dos estímulos, ou instruir o adulto participante a adotar um papel mais ativo para estimular o engajamento da criança (atenção conjunta). Há também a possibilidade de criação de tarefas nas quais o conteúdo apresentado seja contingente ao olhar, estimulando um papel mais ativo da criança.

O examinador precisa, durante a aplicação, considerar não apenas o aspecto comportamental, mas também ergonômico sobre o engajamento. Um marco particularmente importante do desenvolvimento motor é o do suporte pélvico e abdominal para se sentar, atingido tipicamente entre 5-6 meses (Adolph & Robinson, 2015), que em muito muda a interação espontânea do paciente com a tarefa. Na situação de aplicação do presente estudo, para crianças mais novas, o adulto responsável precisa sustentar a criança orientada para o monitor com mais cuidado, garantindo tanto o conforto da criança quanto o bom enquadramento no campo de visão da câmera. Após desenvolvida a capacidade de se sentar, a criança passa a movimentar com maior liberdade não apenas a cabeça e os olhos,

mas todo o corpo durante sua resposta de orientação. Idealmente, montagens experimentais em que a criança permanece em outras posições que minimizem o esforço da resposta de orientação (deitada de costas, mantida com apoio de cabeça nos ombros do adulto) podem melhorar o engajamento da criança com a tarefa.

6.8 Desafios para a aplicabilidade clínica

Sistemas automatizados de diagnóstico têm sido tratados na literatura como tendo um papel de suporte à tomada de decisão clínica, e não de substituição a esta. Um dos requisitos básicos para implementações práticas efetivas tem sido o de garantir alto grau de envolvimento dos profissionais clínicos responsáveis pela avaliação em todos os aspectos do sistema, como entrada de dados, curadoria do conjunto de dados utilizado para construções de bases de dados e ajuste de parâmetros de modelos, e conhecimento dos mecanismos de inferência utilizados pelos sistemas (Rajkomar, Dean, & Kohane, 2019). Na medida em que ferramentas automatizadas de avaliação da função visual atinjam um grau de desenvolvimento tal que sejam aplicáveis em ampla escala (isto é, sua testabilidade e usabilidade estejam demonstradamente adequadas para a população de interesse, as metodologias de análise estabelecidas, e validação em larga escala esteja disponível), ainda assim não se pode desconsiderar a hipótese de falso diagnóstico, caso, por exemplo, as fixações do paciente sejam demasiado breves para serem registradas como fixação, ou o paciente esteja desatento, sonolento ou não cooperativo para concluir a avaliação. A avaliação destes aspectos continua sendo uma questão de julgamento clínico do profissional responsável, mesmo que alguns aspectos da avaliação sejam automatizados.

Partindo desta constatação, o desenvolvimento de aplicações de avaliação clínica baseadas em rastreamento ocular pode ter como requisito garantir tal papel de apoio à decisão clínica. O registro em vídeo pode auxiliar, por exemplo, que um avaliador possa examinar criticamente eventuais erros de software. A ferramenta pode apontar para momentos específicos da avaliação (isto é, indexar o registro em vídeo e/ou sinal bruto para agilizar a análise posterior pelo avaliador) onde há maior ambiguidade, caracterizada por momentos em que a fixação relativa esteja em posição limítrofe ao critério automatizado adotado, e cuja avaliação possa se beneficiar da supervisão e eventual intervenção por parte do clínico sobre a detecção de algum estímulo específico que possa impactar o diagnóstico.

Esta discussão se insere em um contexto maior, na medida em que na última década temos presenciado a incorporação de ferramentas de inteligência artificial para diagnóstico clínico. Ferramentas de apoio diagnóstico em visão podem atuar de forma análoga ao que acontece na radiologia, onde as soluções operam no sentido de aumentar a homogeneidade dos critérios diagnósticos utilizados por diferentes observadores, e reduzir a carga cognitiva humana na avaliação (eliminando erros devido a fadiga por exemplo) (Rajkomar et al., 2019). Não há como se falar em substituição de um avaliador, mas apenas na mudança na caracterização da atividade que realiza. O julgamento do observador, que antes era a única fonte de informação durante um procedimento, passa a ser uma das fontes de informação. Tanto o julgamento do(s) observador(es) quanto o resultado da análise automatizada passam a estar sujeitos a crítica mútua a fins de resolução de eventuais ambiguidades e contradições.

A literatura médica tem empregado o conceito de viés de automação para situações nas quais o aumento da confiança em sistemas de automação de decisões pode levar a erros (Lyell & Coiera, 2017). Tais erros podem ser de duas naturezas: de comissão, quando o desempenho sub-ótimo do sistema em questão leva o clínico a realizar um julgamento equivocado com base em informação fornecida pelo sistema; ou de omissão, quando o clínico ignora informações concretas do caso em função de ter prestado atenção de forma desproporcional às informações fornecidas pelo sistema. A segunda classe de erros tem origem em fatores humanos, e independe da acurácia do sistema: sistemas com maior acurácia tendem inclusive a serem depositários de maior grau de confiança por parte da equipe clínica, aumentando a chance deste tipo de erro (Lyell & Coiera, 2017).

A existência do risco de viés de automação coloca a questão de se, durante o processo de desenvolvimento e validação de ferramentas para avaliação da função visual baseadas em rastreamento ocular, os requisitos de transparência, facilidade de supervisão e intervenção por parte do avaliador não podem conflitar com os requisitos de melhor acurácia da análise automatizada. Por exemplo, reduzir o tamanho do estímulo (portanto, da região de interesse) e apresentá-la em múltiplas regiões de forma aleatorizada reduz a taxa de falsos-positivos (Jones et al., 2015), entretanto, diminui a margem para que um clínico faça a revisão posterior por meio do registro de vídeo, dado que o julgamento de posição do olhar qualitativo realizado por um observador humano exige apresentação em posições previsíveis e relativamente espaçadas. Procedimentos psicofísicos adaptativos são sem

dúvida mais eficientes, mas uma adesão estrita ao protocolo depende exclusivamente do resultado da análise automatizada, com pouca margem para intervenção externa por parte do clínico. Se minimizar a complexidade da tarefa de supervisão e verificação por parte de um observador humano é um requisito importante para o uso efetivo de ferramentas automatizadas de diagnóstico, conforme a revisão de Lyell & Coiera (2017) aponta, o processo de desenvolvimento precisa considerar em que grau tal demanda precisa ser contemplada, sem prejudicar a natureza e as vantagens oferecidas pela análise automatizada. Essa questão torna-se ainda mais relevante na avaliação de uma população pré-verbal, já que o grau de cooperação e engajamento com a tarefa varia de participante para participante.

O requisito de maior grau de supervisão por parte de um clínico pode significar também que medidas mais diretas sejam utilizadas. Ao passo em que características mais complexas extraídas do dado de rastreamento ocular sejam incorporadas ao modelo de detecção do estímulo, pode-se argumentar que há maior dependência aos aspectos técnicos do rastreador, tornando a ferramenta mais opaca ao processo de exame e revisão por parte do avaliador. Um critério por exemplo que segmente o sinal em fixação e sacadas, por exemplo, requer um limiar de dispersão ou velocidade para realizar esta segmentação. Ora, tal aspecto depende não apenas do desempenho do participante, mas também de características técnicas como distância de aplicação, comprimento focal da câmera, frequência temporal de aquisição, etc. Portanto, a adoção de um critério mais simples, como tempo relativo de fixação (*dwell time*), presumivelmente favorece a transparência. Tal medida, por ser mais diretamente calculável a partir de um sinal bruto, pode ser também, a princípio, mais robusta no tocante a possíveis variações entre múltiplos sistemas de rastreamento ocular.

Ao mesmo tempo em que uma maior possibilidade de intervenção por parte do avaliador é desejável na clínica, os impactos no uso de ferramentas automatizadas de rastreamento ocular na pesquisa precisam ser considerados, dado que o desenvolvimento de programas de pesquisa robustos é condição essencial para o aperfeiçoamento da técnica a ponto de criar soluções aplicáveis à clínica. A boa prática científica sugere a criação de critérios objetivos e uniformes aplicados à análise, eliminando ao máximo aspecto subjetivo da avaliação. Um grau maior de flexibilidade na análise pode, inadvertidamente, introduzir

maiores graus de liberdade que minem a reprodutibilidade dos resultados (Simmons, Nelson, & Simonsohn, 2011). Critérios de extração de características (intervalos de sacadas e fixações) são dependentes não apenas das características do paciente avaliado, mas também dos parâmetros de aquisição da imagem. Eventuais ajustes nesse sentido por parte do avaliador pode não ter o mesmo efeito para todos os pacientes, e podem exigir conhecimento de aspectos que fogem da esfera de competência profissional clínica.

Reportar ao avaliador múltiplos valores de acuidade (ao invés de um valor único) baseados em múltiplos pontos de corte da métrica adotada (por exemplo, tempo relativo de fixação) pode ser uma saída que permita ao mesmo tempo uma análise crítica por parte do avaliador e garanta certa homogeneidade das avaliações entre múltiplas situações de avaliação. É possível reportar diferentes valores de acuidade para diferentes valores de duração mínima da fixação, sendo o ponto máximo da curva interpretado como um limite superior para a acuidade. Comparações entre múltiplos estudos podem ser feitas utilizando-se sempre os mesmos pontos de corte, e avaliações individuais podem permitir certa flexibilidade no ponto adotado, desde que haja justificativa plausível para a escolha. A partir do exame crítico da sensibilidade do resultado ao critério adotado, o profissional pode emitir um julgamento a respeito do desempenho da avaliação automatizada em uma determinada avaliação. Na hipótese de um procedimento adaptativo, ao invés de se basear a decisão do próximo estímulo a ser apresentado em um único ponto de corte, pode-se tomar por base a probabilidade marginal de detecção, a partir de uma série de estimativas de acuidade ao se adotar diferentes pontos de corte.

Estudos futuros que se proponham a normatizar o processo de avaliação por olhar preferencial podem inclusive se dedicar à coleta de dados normativos considerando múltiplos critérios simultaneamente ao invés de convergir para um critério universal. Por exemplo, pode-se gerar um conjunto de dados normativos para o critério de sacadas na direção do estímulo, outro conjunto de critérios normativos de deslocamento de fixação, e outro conjunto de dados normativos ainda de latência de sacada, todos baseados em um mesmo conjunto de avaliações. Na eventual ocasião de uso dos dados normativos para fins de avaliação clínica, o valor de acuidade considerado "verdadeiro" pode ser aquele que tiver o melhor valor, dentre cada um dos múltiplos critérios adotados. Adotar o melhor valor (e não a média, ou mediana, ou o menor valor dentre todos os critérios) baseia-se na heurística de

que a probabilidade de detecção (positivo verdadeiro) é superior à probabilidade de erro por omissão (falso positivo), o que em certa medida é uma heurística baseada no desenho da tarefa experimental, mas que possui certo grau de constatação empírica (Jones et al., 2014).

7 Conclusão

No presente trabalho, desenvolvemos uma solução de software para realização de tarefas experimentais com vídeo-oculografia, e exploramos uma aplicação particular desta solução, que é a realização do exame psicofísico de AV pela técnica de olhar preferencial com crianças pré-verbais (que é o objetivo geral do trabalho e contempla os objetivos específicos 1 e 2). Determinamos que o teste subestimou os valores de AV medidos quando comparado ao teste tradicional em 0,38 logMAR em média (aproximadamente dois cartões do teste original), de acordo com os objetivos específicos 3 e 4. Apesar dessa diferença, observamos que o teste é capaz de capturar uma tendência esperada de melhora na AV com o desenvolvimento da criança de forma similar ao teste tradicional (-0,018 logMAR/mês). Ao final da execução do projeto, contamos com uma aplicação que pode servir a futuros estudos de validação em larga escala nesta linha de pesquisa, de acordo com o objetivo específico 5 (ver também Apêndice B, que descreve alguns aspectos técnicos da aplicação).

A diferença encontrada entre os valores de AV medidos pelo TAC e pelo exame de vídeo-oculografia (APL) precisa ser reduzida antes que a aplicação diagnóstica da tarefa seja considerada. A reduzida diferença entre o valor de AV calculado automaticamente pelo programa (APL) e pelo valor calculado a partir da análise manual do registro de vídeo da mesma sessão (VPL), de 0,17 logMAR (aproximadamente um cartão), sugere que a subestimação não se deve ao uso de uma análise automatizada, mas sim de diferenças procedimentais que impactam o engajamento da criança com a tarefa. Na medida em que o impacto destas diferenças seja minimizado através de melhorias metodológicas na tarefa (tempo e esquema de apresentação dos estímulos de teste, uso de procedimentos adaptativos, maior qualidade e variedade do conteúdo apresentado para engajamento, dentre outras), será possível examinar o poder diagnóstico do teste para diferentes condições clínicas. A possibilidade de revisão do desempenho do algoritmo por meio da inspeção visual do registro de vídeo é uma contribuição original do presente trabalho, que pode favorecer sua aplicabilidade clínica, na medida em que o examinador pode avaliar criticamente o desempenho do algoritmo.

O poder diagnóstico (sensibilidade e especificidade) do teste para identificar perdas de acuidade depende da diferença entre o valor de acuidade de um determinado quadro clínico (grau de severidade da perda) e o valor normativo de referência. Quadros de

degeneração retiniana, por exemplo, são geralmente caracterizados por perdas graves, da ordem de 2,8 oitavas (0,84 logMAR) (D. Mayer, Fulton, & Hansen, 1985), e podem a princípio ser detectados com maior acurácia, mesmo que haja diferença significativa entre os valores normativos levantados por diferentes procedimentos diagnósticos. Quadros de ambliopia, entretanto, são caracterizados por um amplo espectro da gravidade das perdas visuais, com distribuição variando entre perdas sutis da ordem de 0,3 logMAR (valor que compõe o critério diagnóstico) até perdas da ordem de 2,3 logMAR em adultos (McKee, Levi, & Movshon, 2003). Mesmo que a completa paridade entre os testes não seja demonstrada após futuras melhorias no procedimento, é preciso minimizá-la a ponto de que mesmo as perdas mais sutis (como as presentes nos casos mais leves de ambliopia, da ordem de 0,3 logMAR) possam ser detectadas, garantindo a utilidade da ferramenta para diagnóstico de perdas visuais em múltiplos níveis de gravidade.

8 Considerações finais

É concebível que os trabalhos futuros que se proponham a automatizar o procedimento de olhar preferencial procedam adotando critérios sobre as características quantitativas do sinal de posição do olhar que sejam universalmente aceitos. Mas essa alternativa não condiz com a realidade de que os comportamentos não-verbais ocorrentes nas tentativas julgadas como "não visto" irão variar de criança para criança. É pelo menos esperado que essas características sejam sistematicamente relacionadas à idade, dado o melhor controle do sistema oculomotor e por consequência o comportamento visual exploratório mais preciso na medida em que a criança se desenvolve. Mesmo considerando esta tendência, transpor a técnica para a clínica exige que diferenças individuais apresentadas por diferentes quadros clínicos possam ser acomodadas. A população que mais pode se beneficiar da técnica (quadros de nistagmo, baixa visão, ambliopia e estrabismo) pode se diferenciar, em sua resposta comportamental, da população padrão utilizada durante o procedimento de validação, e por consequência de quaisquer critérios universais de características do sinal de posição ocular que são aplicados, incorrendo no risco de subestimação da acuidade nestes casos.

Parte da motivação da criação dos Cartões de Acuidade de Teller foi permitir certa liberdade para que o avaliador decidisse quais comportamentos apresentados pela criança indicassem uma detecção do estímulo apresentado (Teller et al., 1986). A adoção de critérios universais de detecção de certa forma ignora a possibilidade de tornar a avaliação sensível a tais aspectos, o que em última instância leva ao risco de, na tentativa de uniformizar o processo de avaliação, uma fonte de viés seja introduzida na técnica. É de vital importância que um procedimento de automatização de medida de olhar preferencial com aplicabilidade clínica garanta a acomodação de diferenças individuais na forma como o paciente se engaja na tarefa. Acomodar tais diferenças ao mesmo tempo em que a reprodutibilidade da técnica seja garantida é um desafio reservado aos trabalhos futuros nesta linha de pesquisa.

Referências

- Abramov, I., Gordon, J., Hendrickson, A., Hainline, L., Dobson, V., & LaBossiere, E. (1982). The retina of the newborn human infant. *Science*, *217*(4556), 265–267.
- Adoh, T. O., Woodhouse, J. M., & Oduwaiye, K. A. (1992, Jun). The Cardiff Test: a new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. *Optom. Vis. Sci.*, *69*(6), 427–432. doi: 10.1097/00006324-199206000-00003
- Adolph, K. E., & Robinson, S. R. (2015). Motor development. *Handbook of child psychology and developmental science*, 1–45.
- Anderson, D. R., & Levin, S. R. (1976). Young children's attention to "sesame street". *Child Development*, 806–811.
- Antoniades, C., Ettinger, U., Gaymard, B., Gilchrist, I., Kristjánsson, A., Kennard, C., ... Carpenter, R. H. S. (2013, May). An internationally standardised antisaccade protocol. *Vision Res.*, *84*, 1–5. doi: 10.1016/j.visres.2013.02.007
- Aslin, R. N. (1977, Feb). Development of binocular fixation in human infants. *J. Exp. Child Psychol.*, *23*(1), 133–150. doi: 10.1016/0022-0965(77)90080-7
- Aslin, R. N., & Salapatek, P. (1975, May). Saccadic localization of visual targets by the very young human infant. *Percept. Psychophys.*, *17*(3), 293–302. doi: 10.3758/BF03203214
- Aslin, R. N., & Smith, L. B. (1988, Jan). Perceptual Development. *Annu. Rev. Psychol.*, *39*(1), 435–473. doi: 10.1146/annurev.ps.39.020188.002251
- Banks, M. S. (1980, September). The Development of Visual Accommodation during Early Infancy on JSTOR. *Child Dev.*, *51*(3), 646–666.
- Banks, M. S., & Bennett, P. J. (1988). Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. *JOSA A*, *5*(12), 2059–2079.
- Banks, M. S., & Salapatek, P. (1978, Apr). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Invest. Ophthalmol. Visual Sci.*, *17*(4), 361–365.
- Banks, M. S., & Salapatek, P. (1981, Feb). Infant pattern vision: A new approach based on the contrast sensitivity function. *J. Exp. Child Psychol.*, *31*(1), 1–45. doi: 10.1016/0022-0965(81)90002-3
- Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008, January). Infants' Attention and Responsiveness to Television Increases With Prior Exposure and Parental Interaction.

- Infancy*, 13(1), 30–56. doi: 10.1080/15250000701779378
- Bedell, H. E., & Stevenson, S. B. (2013, Sep). Eye movement testing in clinical examination. *Vision Res.*, 90, 32–37. doi: 10.1016/j.visres.2013.02.001
- Bennett, C. R., Bex, P. J., Bauer, C. M., & Merabet, L. B. (2019, Oct). The Assessment of Visual Function and Functional Vision. *Semin. Pediatr. Neurol.*, 31, 30–40. doi: 10.1016/j.spen.2019.05.006
- Bharadwaj, S. R., & Candy, T. R. (2008, Dec). Cues for the control of ocular accommodation and vergence during postnatal human development. *J. Vis.*, 8(16), 14. doi: 10.1167/8.16.14
- Birch, E. E., Shimojo, S., & Held, R. (1985, Mar). Preferential-looking assessment of fusion and stereopsis in infants aged 1-6 months. *Invest. Ophthalmol. Visual Sci.*, 26(3), 366–370.
- Blanke, O., Morand, S., Thut, G., Michel, C. M., Spinelli, L., Landis, T., & Seeck, M. (1999, Apr). Visual activity in the human frontal eye field. *Neuroreport*, 10(5), 925–930. doi: 10.1097/00001756-199904060-00006
- Borchert, M. S., & Angeles, C. H. L. (2000, Jul). *Methods and apparatus for measuring visual acuity in preverbal children*. ([Online; accessed 26. Mar. 2022])
- Braddick, O., & Atkinson, J. (2011, Jul). Development of human visual function. *Vision Res.*, 51(13), 1588–1609. doi: 10.1016/j.visres.2011.02.018
- Brilhault, A., Neuenschwander, S., & Rios, R. A. (2019). Using Clustering in Eye-Tracking Calibration Task. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 15–18). IEEE. doi: 10.1109/BRACIS.2019.00115
- Bronson, G. W. (1990, Feb). Changes in infants' visual scanning across the 2- to 14-week age period. *J. Exp. Child Psychol.*, 49(1), 101–125. doi: 10.1016/0022-0965(90)90051-9
- Burger, W., & Burge, M. J. (2016). *Digital image processing: an algorithmic introduction using java*. Springer.
- Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development*, 23(1), 3–21.
- Cerrolaza, J. J., Villanueva, A., & Cabeza, R. (2008, Mar). Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In *ETRA*

- '08: *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 259–266). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/1344471.1344530
- Chang, M. Y., & Borchert, M. S. (2020, Nov). Advances in the evaluation and management of cortical/cerebral visual impairment in children. *Surv. Ophthalmol.*, 65(6), 708–724. doi: 10.1016/j.survophthal.2020.03.001
- Chang, M. Y., & Borchert, M. S. (2021, Dec). Validity and reliability of eye tracking for visual acuity assessment in children with cortical visual impairment. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 25(6), 334.e1–334.e5. doi: 10.1016/j.jaapos.2021.07.008
- Chen, A. H., & Mohamed, D. (2003, Oct). New paediatric contrast test: Hiding Heidi low-contrast 'face' test. *Clin. Experiment. Ophthalmol.*, 31(5), 430–434. doi: 10.1046/j.1442-9071.2003.00691.x
- Clark, R., Blundell, J., Dunn, M. J., Erichsen, J. T., Giardini, M. E., Gottlob, I., ... Williams, C. (2019, Aug). The potential and value of objective eye tracking in the ophthalmology clinic. *Eye*, 33(8), 1200–1202. doi: 10.1038/s41433-019-0417-z
- Clifford, C. E., Haynes, B. M., & Dobson, V. (2005). Are norms based on the original teller acuity cards appropriate for use with the new teller acuity cards ii? *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 9(5), 475–479.
- Cornsweet, T. N., & Crane, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth purkinje images. *JOSA*, 63(8), 921–928.
- Coutinho, F. L., & Morimoto, C. H. (2013). Improving head movement tolerance of cross-ratio based eye trackers. *International journal of computer vision*, 101, 459–481.
- Curtis, C. E., & Connolly, J. D. (2008, Jan). Saccade Preparation Signals in the Human Frontal and Parietal Cortices. *J. Neurophysiol.*
- Daniel, B. M., & Lee, D. N. (1990, Oct). Development of looking with head and eyes. *J. Exp. Child Psychol.*, 50(2), 200–216. doi: 10.1016/0022-0965(90)90039-B
- De Courten, C., & L. J., G. (1983, Jan). [Morphological development of the primary visual pathway in the child]. *J. Fr. Ophtalmol.*, 6(2), 187–202.
- de Oliveira Dias, J. R., Ventura, C. V., de Paula Freitas, B., Prazeres, J., Ventura, L. O., Bravo-Filho, V., ... Maia, M. (2018, Sep). Zika and the Eye: Pieces of a Puzzle. *Prog.*

- Retin. Eye Res.*, 66, 85–106. doi: 10.1016/j.preteyeres.2018.04.004
- Dobson, V., & Teller, D. Y. (1978, Jan). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Res.*, 18(11), 1469–1483. doi: 10.1016/0042-6989(78)90001-9
- Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- Esteban-Ibañez, E., Perez-Roche, T., Prieto, E., Castillo, O., Fanlo-Zarazaga, A., Alejandre, A., ... Pueyo, V. (2021, Oct). Age norms for grating acuity and contrast sensitivity in children using eye tracking technology. *Int. Ophthalmol.*, 1–10. doi: 10.1007/s10792-021-02040-4
- Fantz, R. L. (1958). Pattern vision in young infants. *The psychological record*, 8, 43.
- Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999, May). Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5), 476–480. doi: 10.1109/34.765658
- Fulton, A. B., & Mayer, D. L. (1988). Esotropic children with amblyopia: effects of patching on acuity. *Graefes Arch. Clin. Exp. Ophthalmol.*, 226(4), 309–312. doi: 10.1007/BF02172956
- Garey, L., & De Courten, C. (1983). Structural development of the lateral geniculate nucleus and visual cortex in monkey and man. *Behavioural Brain Research*, 10(1), 3–13.
- Gaymard, B., Ploner, C. J., Rivaud-Péchéux, S., & Pierrot-Deseilligny, C. (1999, Nov). The frontal eye field is involved in spatial short-term memory but not in reflexive saccade inhibition. *Exp. Brain Res.*, 129(2), 288–301. doi: 10.1007/s002210050899
- GenICam - EMVA. (2022, Apr). Retrieved from <https://www.emva.org/standards-technology/genicam> ([Online; accessed 13. Apr. 2022])
- Gogate, P., Gilbert, C., & Zin, A. (2011, Apr). Severe Visual Impairment and Blindness in Infants: Causes and Opportunities for Control. *Middle East Afr. J. Ophthalmol.*, 18(2), 109. doi: 10.4103/0974-9233.80698
- Gredebäck, G., Johnson, S., & von Hofsten, C. (2009, Dec). Eye Tracking in Infancy Research. *Developmental Neuropsychology*, 35(1), 1–19. doi: 10.1080/87565640903325758
- Grönqvist, H., Gredebäck, G., & Hofsten, C. v. (2006, May). Developmental asymmetries between horizontal and vertical tracking. *Vision Res.*, 46(11), 1754–1761. doi:

- 10.1016/j.visres.2005.11.007
- Guestrin, E. D., & Eizenman, M. (2006, Jun). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng.*, 53(6), 1124–1133. doi: 10.1109/TBME.2005.863952
- Hainline, L., & Riddell, P. M. (1995, Dec). Binocular alignment and vergence in early infancy. *Vision Res.*, 35(23), 3229–3236. doi: 10.1016/0042-6989(95)00074-O
- Hansen, D. W., & Ji, Q. (2009, Feb). In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3), 478–500. doi: 10.1109/TPAMI.2009.30
- Harris, C. M., Hainline, L., & Abramov, I. (1981, Jan). A method for calibrating an eye-monitoring system for use with infants. *Behavior Research Methods & Instrumentation*, 13(1), 11–17. doi: 10.3758/BF03201864
- Harvey, E. M., Dobson, V., Tung, B., Quinn, G. E., & Hardy, R. J. (1999). Interobserver agreement for grating acuity and letter acuity assessment in 1-to 5.5-year-olds with severe retinopathy of prematurity. *Investigative ophthalmology & visual science*, 40(7), 1565–1576.
- Haslwanter, T., & Clarke, A. H. (2010, Jan). Eye movement measurement: electro-oculography and video-oculography. In *Handbook of Clinical Neurophysiology* (Vol. 9, pp. 61–79). Waltham, MA, USA: Elsevier. doi: 10.1016/S1567-4231(10)09005-2
- Hathibelagal, A. R., Leat, S. J., Irving, E. L., Nandakumar, K., & Eizenman, M. (2015, Jul). Measuring Infant Visual Acuity with Gaze Tracker Monitored Visual Fixation. *Optom. Vis. Sci.*, 92(7), 823–833. doi: 10.1097/OPX.0000000000000613
- Haynes, H., White, B. L., & Held, R. (1965). Visual accommodation in human infants. *Science*, 148(3669), 528–530.
- Hernández, E., Hernández, S., Molina, D., Acebrón, R., & García Cena, C. E. (2018, Feb). OSCANN: Technical Characterization of a Novel Gaze Tracking Analyzer. *Sensors*, 18(2), 522. doi: 10.3390/s18020522
- Holmes, J. M., & Clarke, M. P. (2006, Apr). Amblyopia. *Lancet*, 367(9519), 1343–1351. doi: 10.1016/S0140-6736(06)68581-4
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

- Hosp, B., Eivazi, S., Maurer, M., Fuhl, W., Geisler, D., & Kasneci, E. (2020, jun). RemoteEye: An open-source high-speed remote eye tracker. *Behav. Res. Methods*, *52*(3), 1387–1401. doi: 10.3758/s13428-019-01305-2
- Hunnius, S. (2007, Jan). The early development of visual attention and its implications for social and cognitive development. In *Progress in Brain Research* (Vol. 164, pp. 187–209). Waltham, MA, USA: Elsevier. doi: 10.1016/S0079-6123(07)64010-2
- Hunnius, S., Geuze, R. H., & van Geert, P. (2006, Jan). Associations between the developmental trajectories of visual scanning and disengagement of attention in infants. *Infant Behavior and Development*, *29*(1), 108–125. doi: 10.1016/j.infbeh.2005.08.007
- Huttenlocher, P. R., de Courten, C., Garey, L. J., & Van der Loos, H. (1982, Dec). Synaptogenesis in human visual cortex — evidence for synapse elimination during normal development. *Neurosci. Lett.*, *33*(3), 247–252. doi: 10.1016/0304-3940(82)90379-2
- Hyon, J. Y., Yeo, H. E., Seo, J.-M., Lee, I. B., Lee, J. H., & Hwang, J.-M. (2010, Feb). Objective Measurement of Distance Visual Acuity Determined by Computerized Optokinetic Nystagmus Test. *Invest. Ophthalmol. Visual Sci.*, *51*(2), 752–757. doi: 10.1167/iovs.09-4362
- Ierusalimschy, R., De Figueiredo, L. H., & Filho, W. C. (1996). Lua—an extensible extension language. *Software: Practice and Experience*, *26*(6), 635–652.
- Ivanchenko, D., Rifai, K., Hafed, Z. M., & Schaeffel, F. (2021). A low-cost, high-performance video-based binocular eye tracker for psychophysical research. *Journal of Eye Movement Research*, *14*(3). doi: 10.16910/jemr.14.3.3
- Jadon, S. (2020). A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (cibcb)* (pp. 1–7).
- Johnson, M. H. (1990, April). Cortical Maturation and the Development of Visual Attention in Early Infancy. *J. Cognit. Neurosci.*, *2*(2), 81–95. doi: 10.1162/jocn.1990.2.2.81
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991, August). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19. doi: 10.1016/0010-0277(91)90045-6
- Jones, P. R., Kalwarowsky, S., Atkinson, J., Braddick, O. J., & Nardini, M. (2014, Dec).

- Automated Measurement of Resolution Acuity in Infants Using Remote Eye-Tracking. *Invest. Ophthalmol. Visual Sci.*, 55(12), 8102–8110. doi: 10.1167/iops.14-15108
- Jones, P. R., Kalwarowsky, S., Braddick, O. J., Atkinson, J., & Nardini, M. (2015, Aug). Optimizing the rapid measurement of detection thresholds in infants. *J. Vis.*, 15(11), 2. doi: 10.1167/15.11.2
- Jung, R. (2020). Understanding and evolving the rust programming language.
- Kooiker, M. J. G., Pel, J. J. M., Verbunt, H. J. M., de Wit, G. C., van Genderen, M. M., & van der Steen, J. (2016, Sep). Quantification of visual function assessment using remote eye tracking in children: validity and applicability. *Acta Ophthalmol.*, 94(6), 599–608. doi: 10.1111/aos.13038
- Krause, A. (2007). *Foundations of gtk+ development*. Apress.
- Lappi, O. (2016, October). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neurosci. Biobehav. Rev.*, 69, 49–68. doi: 10.1016/j.neubiorev.2016.06.006
- LaRoche, G. R. (2011). Examining the visual system. In *Pediatric clinical skills* (pp. 101–121). Elsevier.
- Larrazabal, A. J., Garcia Cena, C. E., & Martinez, C. E. (2019, May). Video-oculography eye tracking towards clinical applications: A review. *Comput. Biol. Med.*, 108, 57–66. doi: 10.1016/j.combiomed.2019.03.025
- Leat, S. J., Yadav, N. K., & Irving, E. L. (2009). Development of Visual Acuity and Contrast Sensitivity in Children. *Journal of Optometry*, 2(1), 19. doi: 10.3921/joptom.2009.19
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436–444.
- Lee R. Wolin, P. (1964, June). Objective Measurement of Visual Acuity: Using Optokinetic Nystagmus and Electro-Oculography. *Arch. Ophthalmol.*, 71(6), 822–826. doi: 10.1001/archopht.1964.00970010838008
- Leigh, R. J., & Kennard, C. (2004, Mar). Using saccades as a research tool in the clinical neurosciences. *Brain*, 127(Pt), 3. doi: 10.1093/brain/awh035
- Lengyel, D., & Gottlob, I. (2003, Jan). Comparison between grating acuity measured by visual tracking and preferential looking in infants. *Strabismus*, 11(2), 85–93. doi: 10.1076/stra.11.2.85.15102
- Lewis, T. L., & Maurer, D. (2005). Multiple sensitive periods in human visual development:

- evidence from visually deprived children. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 46(3), 163–183.
- Lima, D. d. S., Baran, L. C. P., Hamer, R. D., Costa, M. F. d., Vidal, K. S., Damico, F. M., ... Ventura, D. F. (2020, Feb). Longitudinal visual acuity development in ZIKV-exposed children. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 24(1), 23.e1–23.e6. doi: 10.1016/j.jaapos.2019.11.005
- Lima, D. S., & Ventura, D. F. (2023). A review of experimental task design in psychophysical eye tracking research. *Frontiers in Human Neuroscience*, 17, 1112769.
- Luna, B., Velanova, K., & Geier, C. F. (2008, Dec). Development of eye-movement control. *Brain Cogn.*, 68(3), 293–308. doi: 10.1016/j.bandc.2008.08.019
- Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association*, 24(2), 423–431.
- Mash, C., & Dobson, V. (2005, Sep). Intraobserver reliability of the Teller Acuity Card procedure in infants with perinatal complications. *Optom. Vis. Sci.*, 82(9), 817–822. doi: 10.1097/01.opx.0000177809.75080.41
- Matsuda, K., Nagami, T., Sugase, Y., Takemura, A., & Kawano, K. (2017, May). A Widely Applicable Real-Time Mono/Binocular Eye Tracking System Using a High Frame-Rate Digital Camera. In *Human-Computer Interaction. User Interface Design, Development and Multimodality* (pp. 593–608). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-58071-5_45
- Maurer, D. (2017, Apr). Critical periods re-examined: Evidence from children treated for dense cataracts. *Cognitive Development*, 42, 27–36. doi: 10.1016/j.cogdev.2017.02.006
- Mayer, D., Fulton, A., & Hansen, R. (1985). Visual acuity of infants and children with retinal degenerations. *Ophthalmic Paediatrics and Genetics*, 5(1-2), 51–56.
- Mayer, D. L., & Dobson, V. (1982, Jan). Visual acuity development in infants and young children, as assessed by operant preferential looking. *Vision Res.*, 22(9), 1141–1151. doi: 10.1016/0042-6989(82)90079-7
- McConnell, B. A., & Bryson, S. E. (2005, Dec). Visual attention and temperament: Developmental data from the first 6 months of life. *Infant Behavior and Development*, 28(4), 537–544. doi: 10.1016/j.infbeh.2005.09.002

- McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of vision*, 3(5), 5–5.
- Miron, C., Grigoras, L. I., Ciucu, R., & Manta, V. (2021). Eye image segmentation method based on the modified u-net cnn architecture. *Bulletin of the Polytechnic Institute of Iași. Electrical Engineering, Power Engineering, Electronics Section*, 67(2), 41–52.
- Model, D., & Eizenman, M. (2010, October). An Automated Hirschberg Test for Infants. *IEEE Trans. Biomed. Eng.*, 58(1), 103–109. doi: 10.1109/TBME.2010.2085000
- Naegele, J. R., & Held, R. (1982, Jan). The postnatal development of monocular optokinetic nystagmus in infants. *Vision Res.*, 22(3), 341–346. doi: 10.1016/0042-6989(82)90149-3
- Narvaez Palazón, C., Sánchez Ventosa, A., Nieves Moreno, M., Redondo Ibáñez, A., & de Liaño Sánchez, R. G. (2021, Mar). Study of reliability and validity of VOG Perea and GazeLab and calculation of the variability of their measurements. *Archivos de la Sociedad Española de Oftalmología (English Edition)*, 96(3), 127–132. doi: 10.1016/j.oftale.2020.06.023
- Norcia, A. M., & Tyler, C. W. (1985, Jan). Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Res.*, 25(10), 1399–1408. doi: 10.1016/0042-6989(85)90217-2
- Norcia, A. M., Tyler, C. W., Hamer, R. D., & Wesemann, W. (1989, Jan). Measurement of spatial contrast sensitivity with the swept contrast VEP. *Vision Res.*, 29(5), 627–637. doi: 10.1016/0042-6989(89)90048-5
- Odom, J. V., & Green, M. (1984, Dec). VISUALLY EVOKED POTENTIAL (VEP) ACUITY: TESTABILITY IN A CLINICAL PEDIATRIC POPULATION. *Acta Ophthalmol.*, 62(6), 993–998. doi: 10.1111/j.1755-3768.1984.tb08451.x
- Owsley, C., & Sloane, M. E. (1987, Oct). Contrast sensitivity, acuity, and the perception of 'real-world' targets. *Br. J. Ophthalmol.*, 71(10), 791–796. doi: 10.1136/bjo.71.10.791
- Pang, B., Nijkamp, E., & Wu, Y. N. (2019, sep). Deep Learning With TensorFlow: A Review. *Journal of Educational and Behavioral Statistics*, 45(2), 227–248. doi: 10.3102/1076998619872761
- Peirce, J. W. (2007, May). PsychoPy—Psychophysics software in Python. *J. Neurosci. Methods*, 162(1), 8–13. doi: 10.1016/j.jneumeth.2006.11.017

- Pel, J. J. M., Manders, J. C. W., & van der Steen, J. (2010, Jun). Assessment of visual orienting behaviour in young children using remote eye tracking: Methodology and reliability. *J. Neurosci. Methods*, *189*(2), 252–256. doi: 10.1016/j.jneumeth.2010.04.005
- Pelli, D. G., & Zhang, L. (1991, Jan). Accurate control of contrast on microcomputer displays. *Vision Res.*, *31*(7), 1337–1350. doi: 10.1016/0042-6989(91)90055-A
- Pempek, T. A., Kirkorian, H. L., Richards, J. E., Anderson, D. R., Lund, A. F., & Stevens, M. (2010, September). Video Comprehensibility and Attention in Very Young Children. *Dev. Psychol.*, *46*(5), 1283. doi: 10.1037/a0020614
- Pieh, C., Proudlock, F., & Gottlob, I. (2012, Jan). Smooth pursuit in infants: maturation and the influence of stimulation. *Br. J. Ophthalmol.*, *96*(1), 73–77. doi: 10.1136/bjo.2010.191726
- Portnoi Baran, L. C., Fernades da Costa, M., Summer Vidal, K., Damico, F. M., Telles Salgueiro Barboni, M., da Silva Lima, D., ... Fix Ventura, D. (2019, Aug). Alterations in visual acuity and visual development in infants 1-24 months old either exposed to or infected by Zika virus during gestation, with and without microcephaly. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *23*(4), 215.e1–215.e7. doi: 10.1016/j.jaapos.2019.03.005
- Prager, T. C., Zou, Y. L., Jensen, C. L., Fraley, J. K., Anderson, R. E., & Heird, W. C. (1999, Oct). Evaluation of methods for assessing visual function of infants. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *3*(5), 275–282. doi: 10.1016/S1091-8531(99)70023-6
- Puasuaricua, A., Bozomitu, R. G., Tuarniceriu, D., Andruseac, G., Costin, H., & Rotariu, C. (2017). Analysis of eye image segmentation used in eye tracking applications. *Rev. Roum. Sci. Tech.*, *62*, 215–222.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347–1358.
- Regal, D. M., Ashmead, D. H., & Salapatek, P. (1983, Oct). The coordination of eye and head movements during early infancy: A selective review. *Behav. Brain Res.*, *10*(1), 125–132. doi: 10.1016/0166-4328(83)90158-4
- Richards, J. E., & Hunter, S. K. (1997, Nov). Peripheral stimulus localization by infants with

- eye and head movements during visual attention. *Vision Res.*, 37(21), 3021–3035. doi: 10.1016/s0042-6989(97)00082-5
- Richards, M., & Ford, N. (2020). *Fundamentals of software architecture: An engineering approach*. O'Reilly Media.
- Robinson, D. A. (1986, January). The systems approach to the oculomotor system. *Vision Res.*, 26(1), 91–99. doi: 10.1016/0042-6989(86)90073-8
- Robson, A. G., Nilsson, J., Li, S., Jalali, S., Fulton, A. B., Tormene, A. P., ... Brodie, S. E. (2018, February). ISCEV guide to visual electrodiagnostic procedures. *Doc. Ophthalmol.*, 136(1), 1–26. doi: 10.1007/s10633-017-9621-y
- Ronneberger, O., Fischer, P., & Brox, T. (2015, November). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-24574-4_28
- Salomão, S. R., & Ventura, D. F. (1995). Large sample population age norms for visual acuities obtained with vistech-teller acuity cards. *Investigative Ophthalmology & Visual Science*, 36(3), 657–670.
- Sasson, N. J., & Elison, J. T. (2012, Mar). Eye Tracking Young Children with Autism. *JoVE (Journal of Visualized Experiments)*(61), e3675. doi: 10.3791/3675
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th python in science conference* (Vol. 57, p. 61).
- Shin, Y. J., Park, K. H., Hwang, J.-M., Wee, W. R., Lee, J. H., & Lee, I. B. (2006, Feb). Objective Measurement of Visual Acuity by Optokinetic Response Determination in Patients With Ocular Diseases. *Am. J. Ophthalmol.*, 141(2), 327–332. doi: 10.1016/j.ajo.2005.09.025
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359–1366.
- Sjostrand, J., & Abrahamsson, M. (2022, Apr). *Prevention of amblyopia and the concept of cure* (Vol. 7) (No. 2). *Eur J Ophthalmol.* ([Online; accessed 6. Apr. 2022])
- Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998, January). Newborn infants prefer attractive faces. *Infant Behavior*

- and Development*, 21(2), 345–354. doi: 10.1016/S0163-6383(98)90011-X
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, 31(1), 137–149.
- Tafreshi, D., Thompson, J. J., & Racine, T. P. (2014). An analysis of the conceptual foundations of the infant preferential looking paradigm. *Human Development*, 57(4), 222–240.
- Teller, D. Y. (1979, Jan). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development*, 2, 135–153. doi: 10.1016/S0163-6383(79)80016-8
- Teller, D. Y. (1998, Nov). Spatial and temporal aspects of infant color vision. *Vision Res.*, 38(21), 3275–3282. doi: 10.1016/S0042-6989(97)00468-9
- Teller, D. Y., McDonald, M. A., Preston, K., Sebris, S. L., & Dobson, V. (1986, Dec). ASSESSMENT OF VISUAL ACUITY IN INFANTS AND CHILDREN; THE ACUITY CARD PROCEDURE. *Dev. Med. Child Neurol.*, 28(6), 779–789. doi: 10.1111/j.1469-8749.1986.tb03932.x
- The Vision in Preschoolers Study Group. (2004, Apr). Comparison of preschool vision screening tests as administered by licensed eye care professionals in the vision in preschoolers study. *Ophthalmology*, 111(4), 637–650. doi: 10.1016/j.ophtha.2004.01.022
- Villanueva, A., Daunys, G., Hansen, D. W., Bohme, M., Cabeza, R., Meyer, A., & Barth, E. (2009, Mar). A geometric approach to remote eye tracking. *Univ. Access Inf. Soc.*, 8(4), 241–257. doi: 10.1007/s10209-009-0149-0
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors (2020, March). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, 17(3), 261–272. doi: 10.1038/s41592-019-0686-2
- Vrabic, N., Juros, B., & Pompe, M. T. (2021). Automated Visual Acuity Evaluation Based on Preferential Looking Technique and Controlled with Remote Eye Tracking. *Ophthalmic Res.*, 64(3), 389–397. doi: 10.1159/000512395
- Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., & Cottrell, G. (2018). Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (wacv)* (pp. 1451–1460).

- Wass, S. V., Smith, T. J., & Johnson, M. H. (2013, Mar). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behav. Res. Methods*, 45(1), 229–250. doi: 10.3758/s13428-012-0245-6
- Wurtz, R. H. (2015, Sep). Using perturbations to identify the brain circuits underlying active vision. *Philos. Trans. R. Soc. London, Ser. B*, 370(1677), 20140205. doi: 10.1098/rstb.2014.0205
- Yoo, D. H., & Chung, M. J. (2005, Apr). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vision Image Understanding*, 98(1), 25–51. doi: 10.1016/j.cviu.2004.07.011
- Yuodelis, C., & Hendrickson, A. (1986, Jan). A qualitative and quantitative analysis of the human fovea during development. *Vision Res.*, 26(6), 847–855. doi: 10.1016/0042-6989(86)90143-4
- Zin, A. A., Tsui, I., Rossetto, J. D., Gaw, S. L., Neves, L. M., Zin, O. A., ... Lopes Moreira, M. E. (2018, Dec). Visual function in infants with antenatal Zika virus exposure. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 22(6), 452–456.e1. doi: 10.1016/j.jaapos.2018.07.352

ANEXO A - Termo de Consentimento Livre e Esclarecido

Instituto de Psicologia da Universidade de São Paulo

Departamento de Psicologia Experimental - Laboratório da Visão

Título do projeto Avaliação da função visual infantil a partir de solução automatizada de rastreamento ocular baseada em vídeo

Pesquisador Responsável Diego da Silva Lima

Você e seu filho(a) estão sendo convidados(as) a participar de um estudo que está sendo desenvolvido pela Universidade de São Paulo. A proposta da pesquisa é a criação de um novo exame de acuidade visual (capacidade de enxergar detalhes) para a população infantil. Como parte desta pesquisa, a acuidade visual será avaliada em seu filho(a) por dois exames:

1. Teste dos Cartões de Acuidade de Teller (CAT). Este teste consiste em mostrar para a criança uma série de cartões contendo padrões de listras pretas sobre um fundo cinza. O examinador determina se a criança consegue ou não fixar as listras, a partir da direção do olhar. Se achar que consegue, passa a apresentar cartões com listras cada vez mais finas até a criança deixar de fixá-las. Durante o exame, a criança ficará em seu colo ou sentada de frente para o examinador, a depender da idade. Este exame tem duração aproximada de 5-20 minutos, e você estará acompanhando a criança durante todo o procedimento.

2. Teste computadorizado de rastreamento ocular desenvolvido em nosso laboratório. Este teste é similar ao teste CAT citado acima, com a diferença de que a criança será posicionada sentada em frente a um computador (ou em uma cadeirinha do tipo bebê conforto, a depender da idade) utilizado para a apresentação de estímulos. O registro da direção do olhar é feito por uma câmera posicionada junto ao computador. Este teste também tem duração aproximada de 5-20 minutos, e você estará acompanhando a criança durante todo o exame. Poderão ser realizadas até duas sessões de avaliação com este procedimento em um mesmo dia. O teste (2) é o teste que está sendo desenvolvido pelo nosso laboratório, e a aplicação do teste (1) - que já é um teste amplamente utilizado na clínica - será uma forma de garantir que os dados que coletaremos pela segunda técnica são comparáveis e fornecem informações clínicas úteis para o diagnóstico de problemas de visão em crianças.

O pesquisador pode pedir para gravar em vídeo a sessão de avaliação para fins de análise posterior, e fica a seu critério permitir ou não esta gravação. Caso permita, os dados da sessão serão arquivados em base de dados no Laboratório da Visão (USP) e sob nenhuma hipótese serão publicados ou compartilhados para qualquer outro fim que não a realização da pesquisa. Os resultados do exame de CAT e de rastreio ocular serão publicados de forma a preservar a anonimidade do participante, sendo retirada qualquer informação que permita identificar você ou seu filho(a).

A sessão de avaliação será realizada exclusivamente nas dependências do Departamento de Pediatria do Hospital Universitário da Faculdade de Medicina de Jundiaí, onde seu filho(a) está realizando acompanhamento pediátrico. Como a pesquisa fará parte dos atendimentos já agendados para seu filho(a) durante uma mesma visita, a sua participação nessa pesquisa não incorrerá em despesas de transporte ou alimentação que já não estejam previstas para as consultas que você realizará no mesmo dia. A sua participação nesta pesquisa não prevê compensação financeira. Na eventualidade de qualquer dano decorrente da pesquisa, você tem o direito de buscar indenização, nos termos da lei.

Esta pesquisa não apresenta riscos à saúde física da criança, por se tratar apenas de procedimentos não-invasivos de avaliação da função visual, que exigem apenas que a criança interaja com o avaliador e observe os padrões visuais apresentados. Como a situação de avaliação será nova à criança, é possível que a criança sinta algum nível de desconforto emocional devido à interação com pessoas que lhe sejam estranhas, ou devido ao local de avaliação. Se for o caso, o avaliador se compromete a continuar o exame apenas se a criança estiver se sentindo confortável e segura, e todas as providências serão tomadas para essa finalidade. Você estará acompanhando a criança durante a realização dos procedimentos citados acima. Caso ela não esteja cooperativa para a realização dos exames (sonolenta, sob stress por alguma razão de saúde ou desconforto previamente ou durante a sessão), você estará junto e poderá dar qualquer assistência necessária, inclusive interromper a avaliação se for o caso. O avaliador se compromete a manter a criança sempre em situação em que se sinta confortável e segura, e a aderir a todos os protocolos sanitários de prevenção de transmissão doenças infecciosas (em especial da COVID-19) em vigência no Hospital de Medicina de Jundiaí no momento do exame.

Após realização do exame, você terá acesso ao relatório de acuidade visual da

criança, que permite determinar se ela está ou não tendo desenvolvimento normal deste aspecto da função visual. A qualquer momento, você pode entrar em contato com o pesquisador responsável para esclarecer qualquer aspecto do resultado da avaliação ou do procedimento, ou tirar qualquer dúvida sobre a pesquisa. Caso seja detectado no participante algum problema de saúde antes do início da pesquisa, ou como resultado dos exames realizados, o participante será encaminhado para a equipe médica do Hospital Universitário da Faculdade de Medicina de Jundiaí (Que oferece atendimento pelo Sistema Único de Saúde - SUS) para acompanhamento especializado.

O projeto tem por pesquisador responsável Diego da Silva Lima, que pode ser encontrado no endereço Avenida Professor Melo Moraes, 1721, Bloco D, Sala 206, Cidade Universitária, São Paulo- SP, e pode ser contatado pelo Telefone: (...) ou pelo e-mail (...). Se você tiver alguma consideração ou dúvida sobre a ética da Pesquisa, entre em contato com o Comitê de Ética em Pesquisa com Seres Humanos do Instituto de Psicologia da Universidade de São Paulo, no endereço Av. Prof.Melo Moraes, 1721, Bloco G, 2 andar, sala 27- Cidade Universitária – CEP 05508-030-São Paulo – SP; pelo Telefone (...); ou ainda pelo e-mail (...), ou pelo Comitê de Ética em Pesquisa do Hospital Universitário de Jundiaí, no telefone (...) ou e-mail (...).

Sua participação nesta pesquisa é voluntária, e mesmo após consentir em participar, você está livre para interromper sua participação a qualquer momento, sem qualquer tipo de prejuízo. Sua assinatura nesse termo de consentimento indica que você é maior de idade e concorda na utilização dos resultados dos testes para pesquisa, após ter lido as informações aqui contidas (ou alguém as ter lido para você). O presente documento será assinado em duas vias, uma das quais ficará com o pesquisador e a outra com o responsável pela criança.

Nome do Participante: _____

Nome do Responsável (grau): _____

Número do Documento de Identidade: _____

Data de nascimento: __/__/__

Sexo: ()M ()F

Endereço: _____ Complemento: _____

Bairro: _____ Cidade: _____

CEP: _____ Telefone: _____

Declaro que, após convenientemente esclarecido pelo pesquisador e ter entendido o que me foi explicado, consinto participar do presente Projeto de Pesquisa.

São Paulo, data: __/__/__

Assinatura do responsável: _____

Assinatura do pesquisador: _____

APÊNDICE A - Técnica de segmentação de pupila baseada em aprendizagem profunda

Segmentação semântica é um processo para obtenção de imagens segmentadas a partir de representações alta ordem, geralmente obtidas a partir de aprendizagem automática de representação através de algoritmos de otimização estocástica operando sobre redes neurais (Wang et al., 2018). Desde os anos 1990, redes neurais convolucionais surgiram como uma classe de funções otimizáveis com excelente aplicabilidade a problemas de análise de imagem (LeCun, Bengio, & Hinton, 2015). As primeiras redes neurais convolucionais foram planejadas para resolver problemas de classificação, utilizando uma arquitetura composta pela aplicação intercalada das operações de convolução e subamostragem, seguida pela geração de um rótulo de classe único para cada imagem a partir de uma camada de conectividade densa (LeCun et al., 2015). A arquitetura U-Net (Ronneberger, Fischer, & Brox, 2015) substitui a camada de conectividade densa por uma série de de-convoluções, que resulta em uma imagem rotulada de mesma dimensão da imagem original, na qual cada pixel representa o pertencimento a uma classe específica.

A arquitetura U-Net foi inicialmente aplicada ao problema de segmentação de células em lâminas de microscopia óptica (Ronneberger et al., 2015). Descrevemos aqui uma modificação da arquitetura para o problema de segmentação de pupila em imagens obtidas durante uma sessão de rastreamento ocular, visando a obtenção de uma representação segmentada robusta a mudanças de perspectiva, iluminação ambiente e características anatômicas do participante (cor da íris, presença de distratores como cílios, sobrancelhas ou limbo corneal escurecido). A arquitetura U-Net foi previamente aplicada ao problema de segmentação de pupila de forma independente por Miron et al. (Miron, Grigoras, Ciucu, & Manta, 2021), sendo que nossa solução se difere por operar em imagens monocromáticas e em menor dimensão, além de apresentar diferenças nas operações finais das camadas de reamostragem. Por serem capazes de capturar regularidades geométricas e espaciais entre o objeto a ser segmentado e seu fundo, a segmentação semântica se torna mais robusta que a segmentação por intensidade presente em alguns sistemas de rastreamento ocular (Puasuaricua et al., 2017). Outra vantagem é que o aplicador da sessão de rastreamento ocular não precisa reajustar o limiar manualmente ou empregar métodos de seleção automática de limiar que exigem fortes suposições sobre o tamanho da pupila

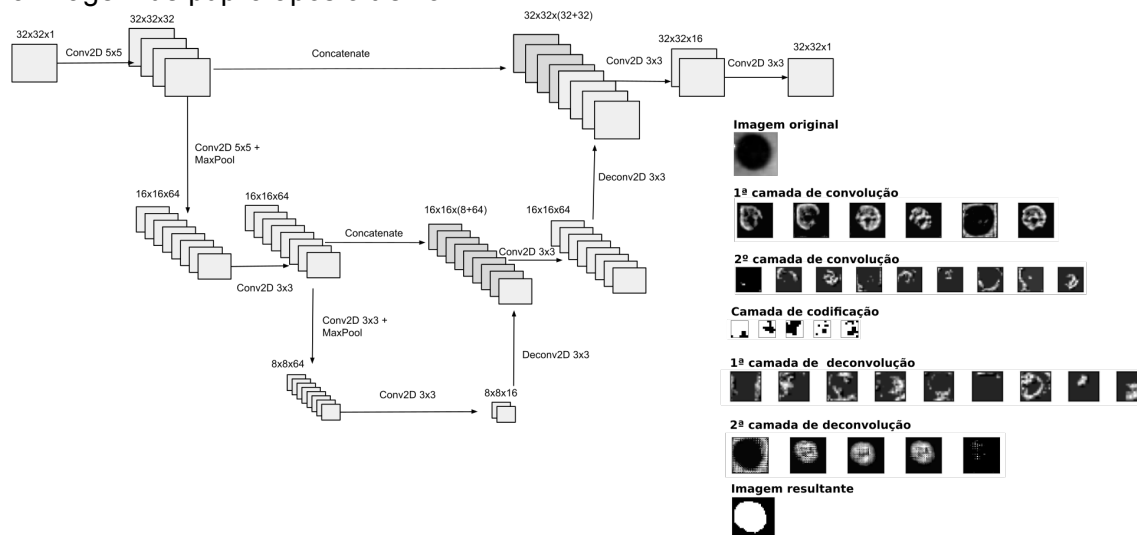
esperado (Puasuaricua et al., 2017).

A fim de segmentar a pupila, foi construída uma arquitetura simplificada inspirada pela U-Net. A simplificação foi motivada, principalmente, para poder ser executada em CPU de propósito geral (Intel x86) sem introduzir atrasos na *pipeline* de processamento do sinal de rastreamento ocular. A rede possui um total de 225,017 parâmetros treináveis. O caminho de subamostragem é composto por uma sequência de convoluções seguidas de subamostragem por máximos locais, e o caminho de reamostragem é uma sequência de deconvoluções (Figura 1) resultante em um mapa de probabilidade com a mesma dimensão da imagem original. A arquitetura U-Net assemelha-se às arquiteturas de codificadores automáticos (*autoencoders*) utilizadas em redes gerativas, que impõem um gargalo central na representação gerada. A diferença fundamental é a presença de conexões saltadas (*skip*), representadas por uma operação de concatenação entre a camada contração e a camada de expansão de dimensão correspondente, cabendo à camada convolucional posterior utilizar um filtro que integre a informação da camada de contração (mais próxima da imagem original) e a produzida pelo processo de codificação central. Pixels com probabilidade de pertencimento a região da pupila maior de 0.5 são então classificados como pertencentes à região da pupila. A rede foi treinada e executada com a biblioteca Tensorflow 2.13 (Pang, Nijkamp, & Wu, 2019). O algoritmo *RMSProp* foi utilizado para o treinamento (uma das variações adaptativas do algoritmo de descida de gradiente estocástica).

Para o processo de treinamento, foram utilizadas regiões de interesse na proximidade da pupila (32x32). Um conjunto de 8000 imagens obtidas a partir de sessões de rastreamento ocular foram utilizadas, com 30 participantes de diferentes idades e sexos, com uma divisão de 75% para o subconjunto de treino de 25% para o subconjunto de teste. As imagens foram aumentadas por translação da região de interesse em passos de 4 pixels na dimensão vertical e horizontal, e também por espelhamento vertical, horizontal e misto, de modo que a pupila estivesse completamente contida em todas as imagens geradas após as transformações. Após o aumento, obteve-se um total de 96000 imagens para o conjunto de treino. Cada imagem foi rotulada com uma máscara circular binária sobreposta à região da pupila.

Se Y representa o conjunto de *pixels* rotulados como pupila e X o conjunto de pixels rotulados como fundo (não-pupila), a função de erro minimizada durante o treino da rede é

Figura 10: Arranjo da rede codificadora inteiramente convolucional (U-NET) utilizada para segmentação de pupila (*Conv2D*: Camada convolucional; *Deconv2D*: Camada deconvolucional; *Concatenate*: Operação concatenadora). O painel inferior direito mostra exemplos de mapas de ativação de cada uma das camadas da arquitetura quando processa uma imagem de pupila após o treino.



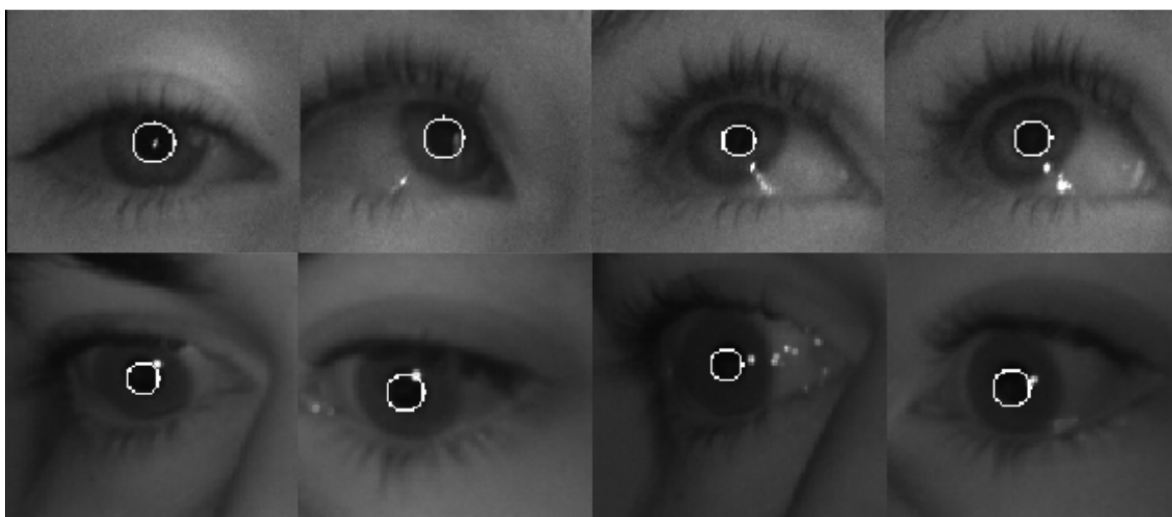
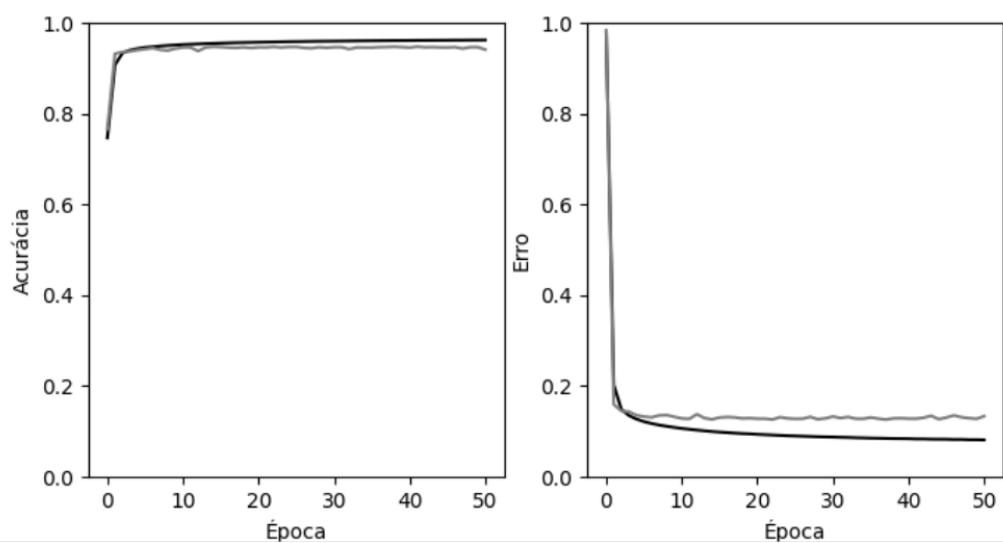
o complemento do coeficiente Dice (Jadon, 2020):

$$L(X, Y) = 1 - \frac{2 |X \cap Y|}{|X| + |Y|} \quad (1)$$

A acurácia final verificada ao fim do processo de treino foi de 98% para o conjunto de treino e 97% para o conjunto de teste. No software, eventuais falsos positivos podem ainda ser rejeitados com base na análise de conectividade através da erosão de todos os componentes que não aquele componente com massa dentro da região de interesse. O processo de treino pode ser reproduzido a partir de conjuntos de imagens de dimensão 32x32 centralizadas na pupila (em escala de cinza) e de imagens binárias com bits 0 ou 1 como resposta representando as máscaras de segmentação.

A rede pode ser utilizada em um sistema de rastreamento ocular tanto para processar cada quadro, como apenas para processar alguns quadros críticos (primeiro quadro de cada sequência de imagens). Neste caso, a rede é utilizada apenas para estimar o valor de intensidade dentro e fora da região da pupila para um algoritmo posterior de limiar de intensidade em escala de cinza, de forma a remover a necessidade de ajustes manuais do limiar e garantir robustez a variações de iluminação. A estratégia utilizada (realizar previsões

Figura 11: Painel superior: Representação do processo de treino da rede neural após 50 épocas (Erro final=0,0883; Acurácia final de treino=95,91%; Acurácia final de teste=94,29%).
Painel superior: Exemplos de elipses delimitantes da região da pupila calculadas a partir das segmentações produzidas pela rede.



a partir de segmentação semântica quadro a quadro, ou apenas em quadros críticos) pode inclusive ser ajustada dinamicamente, na medida em que a segmentação semântica mostre ou não presença de distratores importantes que possam comprometer o desempenho do cálculo de limiar baseado exclusivamente na intensidade. O procedimento de segmentação pode assim ser empregado de diferentes formas a depender de características do hardware e de resolução da imagem de um determinado sistema. A latência para cálculo de predição de cada imagem para um processador Intel x86 com 4 núcleos operando a 2.4 GHz é de aproximadamente 1 milissegundo, sendo portanto aplicável a sistemas de rastreamento ocular operando na faixa dos 60Hz-250Hz e realizando o processamento em hardware com características similares.

A posição final da pupila pode ser calculada através dos momentos de primeira ordem da imagem binária (Burger & Burge, 2016) resultante do processo de segmentação:

$$c = \begin{bmatrix} m_x \\ -m_y \end{bmatrix} \quad (2)$$

e os eixos de sua elipse delimitante são:

$$e_1 = 2\sqrt{\lambda_1}u_1 \quad (3)$$

$$e_2 = 2\sqrt{\lambda_2}u_2 \quad (4)$$

Onde u_1 , e u_2 são os autovetores e λ_1 e λ_2 os autovalores correspondentes da seguinte matriz:

$$M = \begin{bmatrix} m_{xx} - m_x^2 & m_{xy} - m_x m_y \\ m_{xy} - m_x m_y & m_{yy} - m_y^2 \end{bmatrix} \quad (5)$$

Composta por funções de m_x , m_y , m_{xx} , m_{yy} e m_{xy} , que são os momentos de primeira e segunda ordem normalizados da binária resultante da segmentação. Alternativamente, a elipse pode ser calculada através das coordenadas das transições da imagem binária do componente conectado de maior massa presente na imagem, através de algoritmos de ajuste elíptico (Fitzgibbon et al., 1999).

A rede neural para segmentação semântica opera sobre uma pequena região de interesse em torno de cada olho, e uma etapa de pré-processamento baseada em

características simples da imagem é aplicada previamente. Para determinar a posição dos olhos no campo de visão da câmera, foi aplicado um filtro de detecção de borda de dimensão 3 x 3 (filtro Laplace) sobre uma versão da imagem subamostrada em 8 vezes, seguido do cálculo dos máximos locais da imagem filtrada. Tal procedimento foi aplicado a fim de extrair regiões de interesse que poderiam possuir reflexos corneais. Regiões falhando este passo foram descartadas da análise posterior. Por se tratar de registro binocular, apenas regiões alinhadas horizontalmente em um arco máximo de 30 graus foram incluídas.

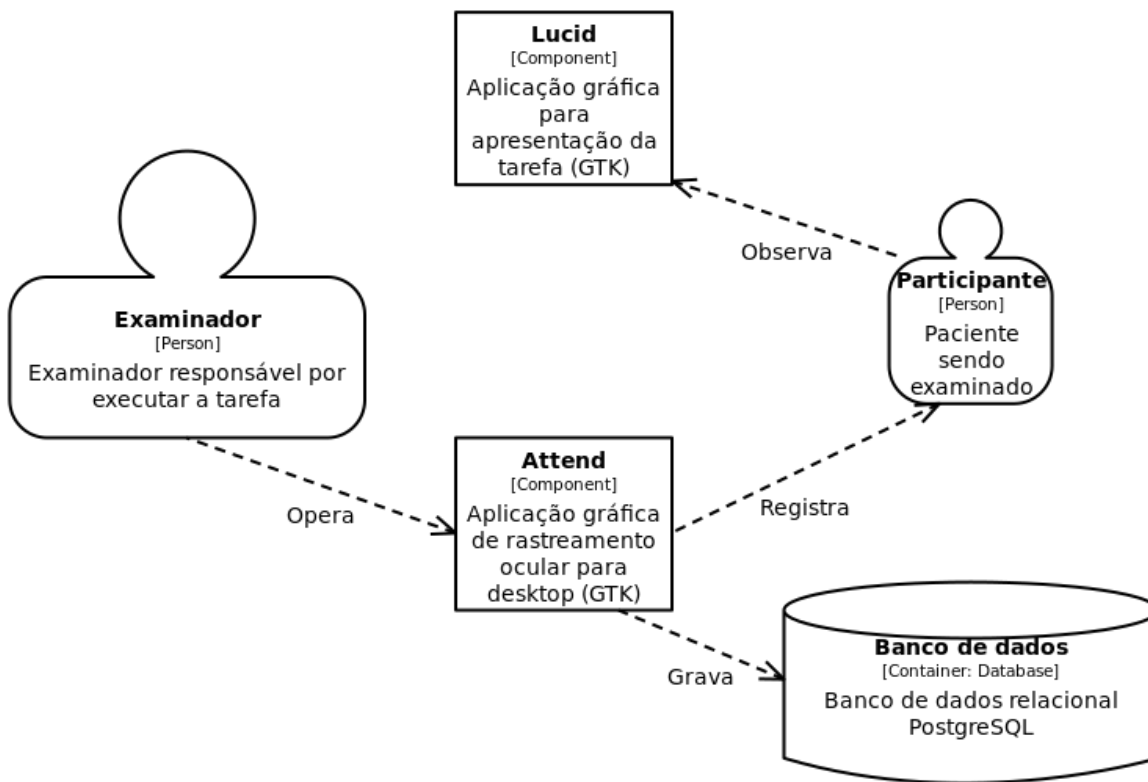
Após extração de tais regiões, foram delimitadas sub-regiões disjuntas que possuíssem valor médio de intensidade correspondente a no máximo o 10º percentil do histograma de intensidade. Tais regiões foram erodidas eliminando em ordem decrescente sub-regiões de intensidade mais clara, até que apenas sub-regiões com uma dimensão máxima (32x32) fossem mantidas. Quaisquer regiões de não estivessem suficientemente próximas de um reflexo corneal em potencial foram descartadas. As regiões restantes foram então submetidas à análise de segmentação semântica por uma rede neural convolucional para produzir uma imagem binária da segmentação da pupila. Um algoritmo de segmentação binária com ligação de vizinhança de quatro pixels (Burger & Burge, 2016) foi então aplicado, sendo preservado apenas a região conectada de maior massa, que foi então definida como a pupila. O reflexo corneal foi então validado como o máximo local da imagem de Laplace mais próximo da pupila.

APÊNDICE B - Especificações do software de vídeo-oculografia

A arquitetura da solução desenvolvida segue reproduzida na Figura 12 para referência, utilizando a convenção de diagrama de componentes do modelo C4 (M. Richards & Ford, 2020). Basicamente, o sistema utilizado para rastreamento ocular é constituído por: (1) uma aplicação gráfica para desktop que permite ao examinador controlar e verificar a qualidade da aquisição de imagens, bem como monitorar em tempo real o registro do sinal de posição do olhar e estado de calibração (aplicação "Attend"); (2) uma aplicação para apresentação da tarefa experimental ao participante, com interface programável para apresentação de imagens estáticas, animações simples e áudio, com comunicação bidirecional com a aplicação controlando a sessão de rastreamento ocular (aplicação "Lucid"). Os módulos da aplicação se comunicam utilizando protocolos de rede locais (ZeroMQ), que podem facilmente ser adaptados para arranjos com múltiplos computadores (para aquisição, apresentação da tarefa e arquivamento) ou para comunicação com outras aplicações.

Espera-se deste arranjo descentralizado, com módulos independentes se comunicando via protocolos de rede de propósito geral, duas vantagens: (1) Maior flexibilidade para uma eventual distribuição do exame: É possível distribuir apenas o módulo de apresentação dos estímulos e extração da direção do olhar para o aplicador, mantendo a análise em ambiente remoto (por exemplo, serviço de nuvem comercial de propósito geral com suporte para banco de dados relacional), com fins de garantir uma base de dados centralizada para fins de pesquisa e validação do instrumento; e (2) Diminuir a carga de processamento do computador utilizado para o teste, que já está apresentando uma aplicação gráfica e executando diversas rotinas numéricas de processamento de imagem. Caso a solução remota não seja uma opção de interesse, um arranjo de rede local (por exemplo, em um laboratório de pesquisa) pode ser igualmente planejado para executar a tarefa. A comunicação entre o programa de apresentação dos estímulos e o programa para monitoramento da sessão e controle do rastreador ocular foi implementada utilizando sockets locais (protocolo TCP). Sinais podem ser enviados tanto a partir do programa de controle do rastreador ocular (por exemplo, para envio de comandos especificando qual estímulo apresentar em seguida) quanto no sentido inverso (por exemplo, carregando os rótulos temporais de início de cada apresentação a fim de sincronizar o registro com o estímulo

Figura 12: Diagrama de componentes do sistema implementado para aquisição de análise dos registros de rastreamento ocular

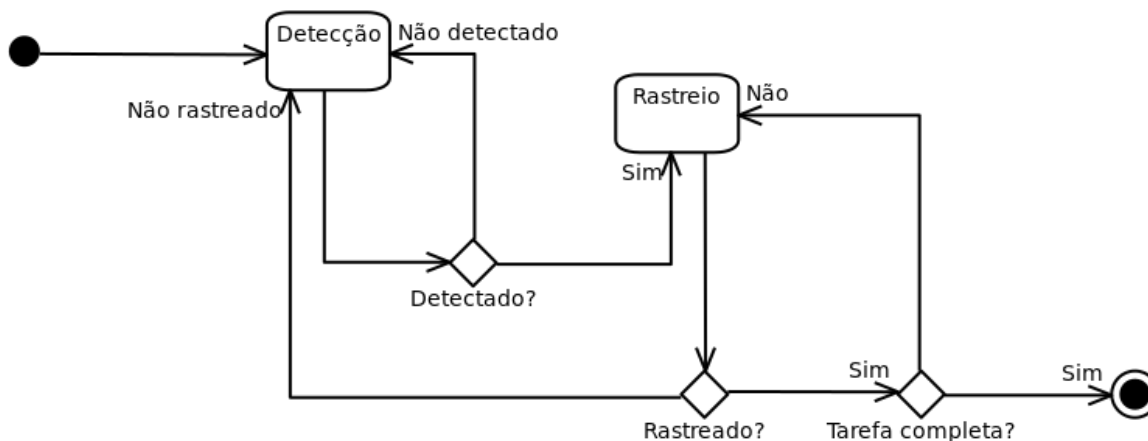


sendo apresentado).

A aplicação foi integrada com uma biblioteca de aquisição de imagens, a Spinnaker SDK (FLIR Systems), utilizada pela câmera com a qual será feita a validação do instrumento. Esta câmera funciona via protocolo amplamente utilizado na indústria, o protocolo GenICam (*GenICam - EMVA, 2022*). Este protocolo oferece uma interface de programação para controle de hardware para aquisição de imagens e é amplamente suportado pela indústria. A principal motivação para o uso de uma solução customizada de rastreamento ocular foi a possibilidade de manutenção do software de forma não atrelada a um hardware específico, sujeito a modificações ao longo do tempo ou descontinuação de suporte por parte do fabricante. Ao suportar esta interface, espera-se que o software possa ser reutilizado para aquisição com modelos variados de câmeras, e com maior liberdade para ajustes de parâmetros experimentais, facilitando o eventual processo de pesquisa translacional e desenvolvimento de produto para esta e potencialmente outras tarefas experimentais com potencial para aplicação clínica.

A aplicação Attend adota alguns aspectos de uma arquitetura de *plugins*, na medida

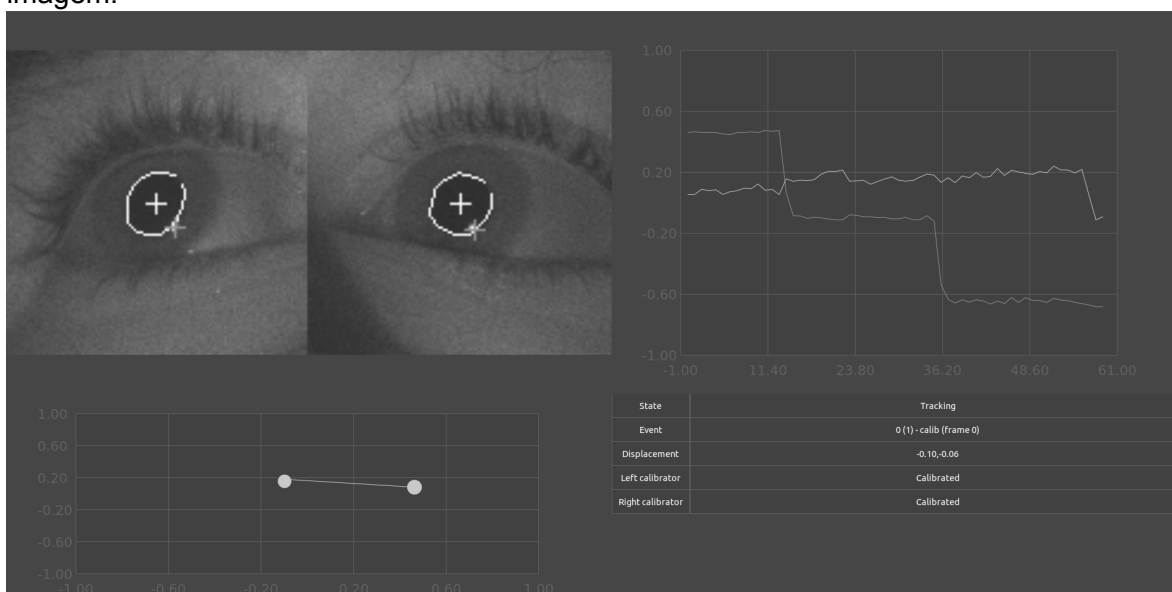
Figura 13: Diagrama de máquina de estados representando a *pipeline* de aquisição de imagens para rastreamento ocular



em que cada uma das etapas da *pipeline* de aquisição de imagem é composta por um algoritmo intercambiável por outro em tempo de execução. Cada quadro resulta em uma união discriminante que rotula o estado do algoritmo, contendo ou não uma coordenada de posição do olhar, que pode ou não estar calibrada a depender do estado do algoritmo. Usuários podem escrever *plugins* para ler tal estado e adotar alguma ação (controle de aplicação externa contingente ao olhar, gravar dados em disco ou banco de dados, etc.).

O software de controle de aquisição de imagem e apresentação de estímulos (Attend) foi desenvolvido na linguagem de programação *Rust* (Jung, 2020), uma escolha motivada pelo bom suporte da linguagem à programação concorrente, performance, interface binária compatível com a linguagem C, e disponibilidade de bibliotecas que suportam o protocolo GenICam para controle do hardware de aquisição de imagem. A interface gráfica do programa para monitoramento do registro de rastreamento ocular e controle de apresentação dos estímulos software foi desenvolvido com a biblioteca GTK4 (Krause, 2007), uma solução multiplataforma para desenvolvimento de interfaces gráficas, que permite curtos ciclos de desenvolvimento em função de sua modularidade e extensa coleção de componentes de interface já disponibilizados. Para acompanhamento em vídeo da sessão experimental em tempo real, componentes da biblioteca *GStreamer* foram utilizados, biblioteca que também foi utilizada para fins de codificação e arquivamento de vídeos da região de interesse (olho esquerdo e direito) durante a sessão. O núcleo da aplicação de aquisição é o controle da *pipeline* de processamento das imagens para

Figura 14: Exemplo de tela da aplicação "Attend" desenvolvida para rastreamento ocular. O painel superior esquerdo permite monitoramento das regiões de interesse selecionadas automaticamente pelo software, e ajustadas dinamicamente na medida em que o participante se move. O painel superior direito mostra as séries temporais dos sinais horizontal e vertical de posição do olhar. O painel inferior esquerdo mostra o *scanpath* mais recente. O painel inferior direito permite ao examinador monitorar o estado atual da *pipeline* de aquisição, informando ao observador possíveis ajustes que precisam ser feitos para obtenção da imagem.



rastreamento ocular, representada na Figura 13. Durante a etapa de detecção, a imagem é escaneada e separada em múltiplas sub-regiões que podem conter possíveis reflexos corneais. Soluções de aprendizado profundo (Apêndice A) são aplicadas para segmentar a pupila de duas destas regiões. Ao encontrar duas pupilas, inicia-se a etapa de rastreio, na qual as duas regiões de interesse examinadas para encontrar as novas coordenadas de posição do olhar são dinamicamente ajustadas com base nas coordenadas de cada quadro prévio.

O desenvolvimento da aplicação prevê uma estratégia descentralizada para o armazenamento e processamento dos resultados da sessão experimental, na qual o computador que realiza a apresentação de estímulos e extração das trajetórias de direção do olhar não é necessariamente aquele utilizado para análise dos resultados. Após a aquisição das trajetórias de rastreamento ocular e sincronização com os dados de

apresentação dos estímulos, os dados são enviados a um servidor de banco de dados relacional, permitindo tanto o acompanhamento da tarefa em tempo real na medida em que os dados sejam processados, quanto para fins de arquivamento. O banco de dados relacional PostgreSQL 11 (PostgreSQL Global Development Group, 2018) foi escolhido como solução para armazenar dados do módulo de controle da tarefa experimental e do módulo de detecção da direção do olhar. O uso de banco de dados relacional para salvar os dados de trajetória em arquivos de texto ou binários oferece uma série de benefícios, como compressão automática dos dados realizada pelo servidor de banco de dados, suporte a acesso concorrente por qualquer tecnologia que o usuário queira utilizar para monitorar a sessão em tempo real, e interface utilizando SQL. O PostgreSQL oferece ainda bom suporte para salvar dados não-estruturados, permitindo por exemplo salvar e recuperar as trajetórias de direção do olhar como vetores unidimensionais, uma representação natural para o problema de rastreamento ocular.

O software permite o registro opcional em vídeo da região de interesse centralizada no olho, para fins de calibração offline, verificação da qualidade do registro e análise qualitativa (classificação da direção do olhar por um avaliador, tal qual a avaliação realizada por observadores no teste de olhar preferencial tradicional. Para esse fim, os quadros de vídeo da região de interesse de tamanho 128 x 128 centralizados em cada olho são gravados utilizando codificador h264, logo após o término de cada sessão. Após decodificação, o software separa os segmentos de vídeo sincronizados à apresentação de cada estímulo, permitindo que a sessão seja reproduzida posteriormente, resgatando a mesma ordem da apresentação dos estímulos, ou agrupando-os por cada categoria de estímulo apresentada. Como apenas as duas regiões de interesse (olho esquerdo e direito) são gravadas, nenhuma característica global da região da face que permita identificar o paciente é salva, o que resguarda a anonimidade do paciente.