



Augmented reality for the Sight Impaired

A thesis submitted to the Nuffield Department of Clinical Neurosciences for the degree of
Doctor of Philosophy

Iain Robert Wilson
Wadham College

Supervisors: Professor Robert MacLaren & Professor Susan Downes

Michaelmas Term 2021



WADHAM
COLLEGE
UNIVERSITY OF OXFORD

Acknowledgements

Foremost I would like to thank my supervisors, Robert MacLaren and Susan Downes, for their guidance and support.

I am very grateful for the Eye Research Group Oxford (ERGO) support in the Oxford Eye Hospital. Notably, the support of Alexina Fantato and Caroline Justice, without whom this work would not have been possible.

I am grateful to my friends and colleagues - Jasleen Jolly, Caroline Peddle, Laura Wood, Laurel Chandler, Harry Orlans, Tina Storm, Alun Barnard, Maria Inês Patrício, Lewis Fry, Imran Yusuf, Mital Shah, Aislin Sheldon and Serena Dillon.

Finally, my family; My wife Kathryn for her ever-present support, and my two sons Henry and Oscar – they are my inspiration.

Abstract

Vision loss is rarely absolute; most people referred to as blind have some form of residual vision. This residual vision can create a confusing view of the world. In this thesis, I explore how a camera system and a wearable display can augment an image of the world to improve its visibility and understanding to a visually impaired wearer. I focus on two crucial parts of life; recognising faces and reading.

In the first part of this thesis, I design and validate a new face perception test to assess different image enhancement forms. I used this new test to measure the impact of image enhancement on a visually impaired population (n=64). The results found that approximately 40% of participants showed a significant improvement in test scores.

Performance was related to reading acuity and lower spatial frequency loss. I then develop and integrate a complete real-time image processing solution for face perception enhancement with an augmented reality display, demonstrating a practical application of the image enhancements.

The second part of this thesis's focus is reading. I compare paper-based reading performance against reading with a virtual reality (VR) display with a visually impaired population (n=53). I further assess the impact of different colours on reading with the VR display. The results found no difference in reading performance, comparing VR to paper-based. The colour of text and background had a significant effect, with 75% of participants having a preferred colour combination. A participant with a higher overall magnitude of colour vision loss is more likely to benefit from white text on a black background than black text on a yellow background.

Augmenting an individual's residual vision can improve the quality of information they receive about the visual world. While not universal, this improvement has the potential to enhance a visually impaired person's quality of life significantly.

Table of Contents

Acknowledgements.....	2
Abstract	3
Table of Contents.....	5
List of figures	9
List of Tables	14
List of abbreviations.....	16
1 Introduction	19
1.1 Visual Impairment	19
1.1.1 Blindness	20
1.2 Central Vision Loss.....	22
1.2.1 The Eye	23
1.2.2 Macular degeneration.....	27
1.2.3 How is central vision loss measured?	31
1.3 Impact of central vision loss.....	35
1.4 Aids for the visually impaired.....	36
1.4.1 Traditional low vision aids.....	36
1.4.2 Smart Low Vision Aids.....	37
1.4.3 Smart Apps	38
1.4.4 Acceptance of LVAs.....	39
1.5 Head-mounted displays.....	40
1.5.1 Virtual and Augmented reality	41
1.5.2 Head-mounted displays and low vision	43
1.6 Thesis Overview.....	47
2 Face Perception.....	49
2.1 Introduction	49
2.1.1 Why faces?	49
2.1.2 The whole and the part.....	50
2.1.3 Are all parts equal?	50
2.1.4 Friend or Foe?.....	51
2.2 What causes face perception loss?	51
2.2.1 Age-related decline – retinal, optical or neurological?	52
2.2.2 Vision loss and Face Perception	52
2.2.3 Behavioural patterns with the loss of face perception	53
2.3 How is Face Perception Assessed?.....	54

2.4	Image Enhancement.....	58
2.4.1	Previous work	58
2.4.2	Cartoonisation	60
2.5	Face Perception Tests	62
2.6	Conclusion.....	63
3	Face perception test.....	65
3.1	A new approach to an existing test.....	65
3.2	Realistic Face Stimulus Generation	66
3.3	Experimental Paradigm.....	70
3.4	Test Validation.....	72
3.4.1	Participants.....	72
3.4.2	Results	73
3.4.3	Measuring difficulty	80
3.5	Discussion.....	85
4	Face Perception - Patients	89
4.1	Materials and Methods	89
4.1.1	Participants.....	89
4.1.2	Experimental Protocol	90
4.2	Results.....	94
4.2.1	Control Participants	94
4.2.2	Patient Participants.....	94
4.2.3	Visual assessments	95
4.2.4	Face Perception Test.....	99
4.2.5	Spectral analysis	111
4.2.6	Factors relating to performance.....	112
4.2.7	Questionnaires	119
4.3	Discussion.....	127
5	Realtime Face Enhancement	133
5.1	Introduction	133
5.1.1	Previous work	134
5.1.2	Automatic Face Segmentation	134
5.1.3	Face Detection.....	136
5.1.4	Face Tracking	137
5.1.5	Skin Detection.....	138
5.1.6	Real-time performance	143

5.2	Proposed System	145
5.2.1	Face detection	146
5.2.2	Face tracking.....	146
5.2.3	Face Tracking persistence	148
5.2.4	Skin Detection.....	152
5.2.5	Real-time image enhancement system	174
5.3	Discussion.....	178
5.3.1	Further work and improvements	179
5.4	Experimental results.....	180
5.4.1	Introduction.....	180
5.4.2	Materials and Methods.....	181
5.4.3	Results	182
5.4.4	Discussion.....	188
5.5	Conclusion	191
6	Visual Acuity with a Head-mounted Display	193
6.1	Introduction	193
6.2	Methods.....	193
6.2.1	Binocular Visual Acuity.....	193
6.2.2	Smart Glasses Visual Acuity.....	194
6.3	Results.....	198
6.4	Discussion.....	201
6.5	Future work.....	203
6.6	Conclusion.....	203
7	Reading.....	205
7.1	Introduction	205
7.1.1	Reading and central vision loss	205
7.1.2	How is reading measured.....	205
7.1.3	Why is reading hard for CVL?	206
7.1.4	How can reading be improved?.....	208
7.1.5	Previous work HMD and reading.....	212
7.2	Reading with a virtual reality display	214
7.2.1	Methods	214
7.2.2	Results	219
7.2.3	Discussion.....	232
7.2.4	Conclusion	236

7.3	How vital is the display size?.....	237
7.3.1	Display size and critical character count.....	238
7.4	Discussion.....	239
7.4.1	Coordination issues.....	241
7.4.2	Low-light situations.....	244
7.4.3	Text to speech	244
7.5	Conclusion.....	245
8	General Discussion	247
8.1	Conclusion.....	260
	Bibliography.....	262
	Appendix – 1	281
A1.1	Visually impaired participants and experimental protocol	281
A1.2	Recruitment	281
A1.3	Protocol.....	281
A1.4	Patient Participants Ethics approval.....	283
A1.5	Patient Participant information sheet.....	291
A1.6	Patient Participant Consent Form	298
A1.7	Patient Case Report Form.....	300
A1.8	Normally sighted participant ethics approval.....	306
A1.9	Normally sighted participant information sheet	307
A1.10	Normally sighted participant consent form.....	311
A1.11	Normally sighted participant recruitment poster	312
A1.12	DLTV Questionnaire.....	313
A1.13	VFQ-25 Questionnaire	314
	Appendix - 2.....	330
A2.1	Wearable Display.....	330
A2.2	Control Computer.....	332
A2.3	Camera	332
A2.4	Latency reduction	333
A2.5	Face detection	335
	Appendix – 3	338
A3.1	Optotype Conversions	338
A3.1.1	M Units	338
A3.1.2	Computer display font size.....	339
A3.2	Virtual Optotype Size.....	339

A3.2.1	HTC Vive	341
A3.2.2	Epson BT350	341
A3.3	Critical Character Count.....	342

List of figures

Figure 1-1 – Simulation of the effect of vision loss – A: Normal Vision, B: Macular Degeneration, C: Diabetic Retinopathy, D: Glaucoma, E: Cataract.....	21
Figure 1-2 The anatomy of the human eye (Created with BioRender.com).....	23
Figure 1-3 The structure of the retina (Created with BioRender.com)	25
Figure 1-4 The retina - source https://en.wikipedia.org/wiki/Macula_of_retina#/media/File:Macula.svg	26
Figure 1-5 Autofluorescence images of A normal eye, B Age-related macular degeneration, C Stargardt. Autofluorescence Imaging uses naturally occurring fluorescence from the retina to provide an indicator of RPE health. In the above image A of the normal eye, the blood vessels and optic nerve head can be seen not to fluoresce, with the fovea slightly darker in the centre of the image. B, the AMD eye, geographic atrophy is observed as the darkening central area over the location of the fovea, producing a significant central scotoma and sight loss. Image C is a Stargardt eye with similar atrophy to the AMD at the fovea and ‘flecks’ over the retina that is typical of the disease.....	27
<i>Figure 1-6 - Centre-surround receptive field. Figures 1) and 2) represent the receptive field, 1) the exterior is deactivated with no illumination (Off-surround), relative to the interior (On-Centre). 2) circles represent the opposite effect; the centre is deactivated and the surround is activated. Figures a) and b) represent spatial frequencies that evoke a positive response. c) and d) represent spatial frequencies that do not evoke a response.....</i>	32
Figure 1-7 Early Treatment Diabetic Retinopathy Study (ETDRS) chart – source: https://commons.wikimedia.org/wiki/File:ETDRS_Chart_R.svg	33
Figure 1-8 - Smart Specs. Control processing unit and headset. The control unit houses the image processor, battery and control electronics. The headset is comprised of an Epson BT-200 display, colour and depth cameras.	45
Figure 2-1 Image enhancements (left) Original face, (centre) Cartoonisation, (right) Adaptive filter.....	60
Figure 3-1 morphing a photograph using Psychomorph. (Top left) Original image with green points denoting the facial feature points. (Top right) Mean face texture and green points denote the mean feature location. (Bottom left) Mean face morphed using the original image features in the positive direction. (Bottom right) mean face morphed in the negative	

direction using the original feature points. The blue lines indicate the distance to a reference point on the mean face.	68
Figure 3-2 processed and cropped images, left (original), centre (cartoon), right (adaptive filter).....	69
Figure 3-3 (top left) Cartoonisation, (top-right) Peli, (bottom-left) Normal – low difficulty, (bottom-right) Normal – high difficulty. Note: a white dot located below the target (the odd-one-out), and the distance value is log10 Euclidean distance between the target and distractor; distance is inversely related to difficulty. The dot and scores are for illustration purposes only and were not present during the experiment.	71
Figure 3-4 (Left) test score distribution for all participants, (right) test scores separated into the first and second tests.	73
Figure 3-5 (Left) test duration distribution for all participants, (right) test duration separated into the first and second tests.....	74
Figure 3-6 (Left) test reaction time distribution for all participants, (right) test reaction time separated into the first and second tests.	74
Figure 3-7 (Top) comparison of the two test scores, line of best fit presented and Pearson’s product-moment correlation. (Bottom) Bland-Altman plot, dashed light-blue lines indicate Coefficient of Repeatability 95th percentile.	76
Figure 3-8 Plot shows how CoR95 changes with task length. The final value after 50 trials was 0.23. The value starts to plateau after 40 trials [CoR95=0.25].	77
Figure 3-9 (left) Table presents the CoR95 and correlations between the tests. There is a notable reduction in variability from the first->second to the second->third. (right) Score box plots for each test.	77
Figure 3-10 (left) Embeddings vs Face feature points distances for all faces – colours represent a stimulus identity. (right) Embeddings vs Face feature points distances for a single stimulus step, and colours represent stimulus identity.....	83
Figure 3-11 – Comparison of threshold score estimates. Pearson’s correlation coefficient presented.	83
Figure 4-1 – Contrast sensitivity test example stimulus – (left) High contrast, (right) low contrast.	92
Figure 4-2 Cambridge Colour Test example stimulus. The top disc is the target, containing colour; the other three are the distractors.....	93
Figure 4-3 Patients Visual Acuity distribution	96
Figure 4-4 Patient Contrast sensitivity results. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.....	97

Figure 4-5 Patient colour sensitivity results. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.	98
Figure 4-6 Patient colour sensitivity results – comparing AMD to Stargardt. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.	99
Figure 4-7 The left plot does not converge, with a consistently high score estimate. The right plot does, stabilising after approximately 300 seconds.	100
Figure 4-8 Shows a lip region from the cartoon enhanced images. Left is the target, and right is the distractor. An example of a feature that some participants used to differentiate the stimulus is highlighted with a red circle.....	102
Figure 4-9 - All face perception scores by visual acuity. The coloured points represent the enhancement type; Yellow triangle = Normal, Red circle = Cartoon, and Blue square = Peli.	103
Figure 4-10 Test score distribution by stimulus type.	104
Figure 4-11 – Box plot of patient face perception scores relative to baseline, by enhancement type. A negative score indicates an improvement.	104
Figure 4-12 - face perception improvement by Visual Acuity. Colours indicate the enhancement type (orange=Cartoon, Yellow = Peli).....	105
Figure 4-13 All performances grouped by their baseline face perception scores (unenhanced normal).	106
Figure 4-14 Control participant Face perception score relative to baseline. A negative score is an improvement in performance.	109
Figure 4-15 (left) Test duration by enhancement type. (right) Test duration distribution plot.	109
Figure 4-16 (left) Test reaction time by enhancement type. (right) Test reaction time distribution plot.	110
Figure 4-17 Participant reaction time by response type. Where the response is a correct or incorrect answer to an individual trial.	111
Figure 4-18 Spatial frequency analysis of the face stimulus. Three lines on the plot represent all of the stimuli with the three modalities; Cartoon (top plot), Peli (middle plot) and normal (bottom plot).	112
Figure 4-19 The overall distribution of scores for the questionnaires totals and subtotals. Each plot represents a total. DLTV and VFQ-25 are for the whole questionnaires; the rest are sub-totals from the VFQ-25.....	124

Figure 5-1 Epson Moverio BT-25E headset	146
Figure 5-2 Face detect and track flowchart.	151
Figure 5-3 Skin classifier results – The plot shows four classifiers (grouped on the x-axis) and the five colour spaces. The colour of the box plots represents the colour space used. E.g. the left hand five boxes are the results for the colour spaces used with the Decision Tree classifier. The upper and lower boundaries of the coloured boxes are the interquartile range, and the error bars are the limits at 1.5 times the Interquartile range.	157
Figure 5-4 A sample of images from the ECU dataset (Son Lam Phung et al., 2005) and the output from the classifiers. Columns left to right are; Original colour image, Ground truth labels (white=skin, black=non-skin), Decision Tree labels, Random Forest labels, Logistic Regression labels and Gaussian Naïve Bayes labels.	159
Figure 5-5 Camera Comparison of three cameras, original image left column, and the skin classification, right column (Random Forest – Lab colour space). Rows are labelled; A = Epson Bt-35E camera, B = Apple MacBook pro webcam, C = Econ Systems See3Cam_130	160
Figure 5-6 The online learning process. Columns: (First from Left) – original image with a green box produced by the face detector. (Second) – the masks used for labelling, the small squares are non-face labels, and the centre mask is for face labels. (Third) - results from GrabCut showing segmented skin. (Fourth) – skin segmented using Random Forest trained with GrabCut data. (Fifth) – skin segmented using Radom Forest trained with offline data. Rows represent different cameras: (top) Econ Systems See3Cam_130, (bottom) – Epson BT300.....	162
Figure 5-7 shows the f-scores (y-axis) of the online (blue) vs offline (orange) for each classifier (x-axis) applied to a sample of 1700 images from the ECU dataset.	163
Figure 5-8 Examples of segmentation failures. Columns; Original image (overlaid face detector results), Ground Truth, Offline Segmentation, Online Segmentation. Rows A & B present failure of the online classifier A is a failure in the face detector, and B the hair has been included. Rows C to E present failures with the offline classifier. C shows a skin colour not represented in the offline dataset. D shows poor camera colour balance, and E shows the background colour too similar to skin.	164
Figure 5-9 Example of good quality face keyframes segmented with GrabCut.....	166
Figure 5-10 - Example of mixed quality face keyframes segmented with GrabCut. Notice the non-face keyframes (a portion of a door in the room).....	166
Figure 5-11 Example face mask created with good quality keyframes (figure 5.9) – with Cartoon enhancement applied.....	167
Figure 5-12 Example face mask with mixed quality keyframes (figure 5.10) – with Cartoon enhancement applied – notice the holes in the mask on the left portion of the image.	167

Figure 5-13 Mask comparison of 'good' (left) and 'bad' (right) grab cut segmentation of a face. All labelled masks' pixels were averaged and normalised, with a colour lookup applied; blue = low occurrence, red = high occurrence.	170
Figure 5-14 Processing flow chart for the skin segmentation method.	172
Figure 5-15 Original unprocessed image – screen capture from processing laptop.....	175
Figure 5-16 Cartoon enhancement overlay, blended with the underlying image.....	176
Figure 5-17 Background removed using a segmented face mask. Cartoon enhancement is blended with the underlying image/ Note the artefacts in the background as the raw mask is used (figure 5.18 uses a cleaned mask).	176
Figure 5-18 Background removed and cleaned, with cartoon enhancement applied (no blending of underlying image is used).....	177
Figure 5-19 Face tracking enabled along with the background removal - the zoom and crop is fully automated and will follow the face around the image.....	177
Figure 5-20 boxplot of visual acuity vs a participant liking the smart glasses	183
Figure 5-21 2 boxplot of age vs a participant liking the smart glasses.....	184
Figure 5-22 Scatter plot showing the relationship between age and visual acuity to preference. Red circles represent patients who did not like the smart glasses; green triangles are those who did.	184
Figure 6-1 Screen capture from the Epson BRT350 display. Snellen M = 19.	196
Figure 6-2 Screen capture from the remote-control application. The "UP SIZE" and "DOWN SIZE" buttons controlled which ETDRS line was displayed, automatically changing the M size. Manual control of the M is set with the "+M" and "-M" buttons.	197
Figure 6-3 Bland-Altman plot of ETDRS VA against Smart Glasses VA. The black line represents the average difference in between the two measurements, the two grey dashed lines represent the 95% confidence interval limits for the average difference. The two blue dashed lines represent ± 0.2 logMAR a clinically relevant change.	199
Figure 6-4 Visual acuity by improvement group. The X-axis is the group of negative (orange), neutral (yellow) and positive (turquoise) improvers. The upper and lower boundaries of the coloured boxes are the inter-quartile range. The median is the horizontal line within the box. The error bars represent 1.5 times the Interquartile range. Figure 2.6.4 Visual acuity by improvement group. The X-axis is the group of negative (orange), neutral (yellow) and positive (turquoise) improvers. The upper and lower boundaries of the coloured boxes are the inter-quartile range. The median is the horizontal line within the box. The error bars represent 1.5 times the Interquartile range.	200
Figure 7-1 HTC Vive virtual reality headset (centre) with two handheld controllers (left & right).....	217

Figure 7-2 – Screen capture from Unity – The virtual paper, black text on the white background, showing an example sentence. Note - The image's low resolution, and blocky text, is an unavoidable artefact of the capture method used, it is much clearer when viewed in the HMD.	218
Figure 7-3 Relationship between Critical print size (CPS) and visual acuity (VA). The blue line is the regression line from the estimated coefficients. The grey band indicates the 95% confidence intervals for the regression fitting.....	222
Figure 7-4 . Example data output from the four Virtual reality reading tests. The circular points represent Black on White presentation, Triangle = Black on Yellow, Square= White on Black, Cross = Yellow on Black. It can be seen that this participant smallest text read, identified by the red circle, was presented as Yellow on Black (Reading acuity = 0.66 logRAD), and the fastest read text, identified by the green circle, was also presented Yellow on Black (Critical Print Size = 1.15 logRAD).	223
Figure 7-5 : : Presents a comparison of the binocular visual acuity with VR binocular visual acuity. The X-axis is the difference between the BVCA and virtual reality BVCA, where a negative value indicates a worse score in the VR BCVA. The red horizontal line indicates the theoretical limit of the VR display (0.7 logMAR). The grey vertical area is a boundary of ± 0.1 logMAR, which is considered a clinically relevant change.....	225
Figure 7-6 Comparison of best and worst scores. A) Presents Reading Acuity best / worst scores, B) presents critical print size (CPS) best / worst scores. The starred bars indicate levels of significance; **** = $p < 0.0001$. RA = Reading Acuity, CPS = Critical Print Size.	228
Figure 7-7 Presentation preference selection method distribution. Bar colours represent the selection method used to choose the preference. The numbers represent the totals in each group.	229
Figure 8-1 None of these people are real; they are generated by StyleGAN (Karras et al., 2020). Source: https://github.com/NVlabs/stylegan	249

List of Tables

Table 1-1 Commercial Head-mounted displays	46
Table 3-1 Stimulus target locations	75
Table 4-1 (left) summary of primary diagnosis, (right) age and visual acuity statistics.	95
Table 4-2 Contrast sensitivity results.....	97
Table 4-3 Colour Sensitivity results.....	98
Table 4-4 Patient participants who could not perform the baseline face perception test.	99

Table 4-5 Participant test failures; summarises the participants whose scores were marked as invalid; the trend is toward lower visual acuity.	101
Table 4-6 - Patient face perception scores by stimulus type.	102
Table 4-7 Control participants face perception score summary.	108
Table 4-8 Linear regression model coefficients for normal enhancements	115
Table 4-9 Linear regression model coefficients for cartoon enhancements	116
Table 4-10 Linear regression model coefficients for Peli enhancements.....	116
Table 4-11 Estimates of the linear regression model Coefficients. The Characteristic column represents the frequencies tested (cycles per degree).	117
Table 4-12 Regression Results.....	125
Table 4-13 Regression results - Individual Questions.....	125
Table 4-14 Selected questions from the DLTV and VFQ-25	126
Table 5-1 Labelled face data sets summary	155
Table 5-2 Skin classification results – best scores for each colour space. AUC = area under the curve. The training time is the total time taken to train the classifier. The results are from classifiers trained and tested on the combined datasets.....	158
Table 5-3 Speed of execution	173
Table 6-1 Character and acuity limits of the BT-350 display.....	195
Table 6-2 Patients with significant score differences	200
Table 7-1 Frequency of conditions	220
Table 7-2 Patient participants	221
Table 7-3 Reading Acuity correlation.....	227
Table 7-4 Effective reading acuity range of commercial HMDs.....	239

List of abbreviations

AF	Fundus autofluorescence
AMD	age-related macular degeneration
API	Application programming interface
4-AFC	4-alternate forced-choice test
AR	Augmented Reality
BCVA	Best-corrected binocular visual acuity
CCC	Critical character count
CCT	Cambridge colour test
CCTV	Closed-circuit television
CNN	Convolutional neural network
CPD	Cycles per degree
CPS	Critical Print Size
CVL	Central vision loss
DLTV	Daily Living Tasks Dependent on Vision
DR	Diabetic retinopathy
ETDRS	Early Treatment Diabetic Retinopathy Study
EU	European Union
GAN	generative adversarial networks
GBT	Gradient boosted trees
GPU	Graphics Processing Unit
HMD	head-mount display
HSV	Hue Saturation Value

LCD Liquid crystal displays

LGN Lateral geniculate nucleus of the thalamus

logMAR Logarithm of the Minimum Angle of Resolution

logRAD Log Reading Acuity Determination

MRS Maximum reading speed

OLED Organic light-emitting diodes

OVAT Oxford Enhanced Vision Assessment Tool

PCA Principle Component Analysis

PRL Preferred retinal locus

RA Reading Acuity

RF Random Forest

RMSE Root mean squared error

RPE Retinal pigment epithelium

RSVP Rapid serial visual presentation

SGVA – Smart Glasses Visual Acuity

UK United Kingdom

VA Visual Acuity

VEGF Vascular Endothelial Growth Factor

VFQ-25 Visual Functioning Questionnaire – 25

VR Virtual Reality

WHO world health organisation

Chapter 1:

Introduction

My contributions to this chapter: The writing is my original work and was conducted over a 6-month period.

Other contributions: Professor Robert MacLaren and Professor Susan Downes assisted with proofreading.

1 Introduction

The aims of this thesis are to develop and refine wearable display technology to assist people living with central vision loss and develop a novel assessment tool to assess its utility.

1.1 Visual Impairment

Visual impairment impacts a significant portion of the global population, estimated to be about 285 million (Pascolini & Mariotti, 2012).

The causes of visual impairment vary according to geographic location, primarily driven by affluence. In low to middle-income countries, the primary causes of visual impairment are uncorrected refractive errors and cataracts. These problems are well treated in higher-income countries, and the common causes are age-related macular degeneration, glaucoma, and diabetic retinopathy (Bourne et al., 2017). Globally, the leading causes of vision impairment are:

- uncorrected refractive errors
- cataract
- age-related macular degeneration
- glaucoma
- diabetic retinopathy

The impact these different forms of vision loss have upon a person is significant, with increased depression, hospital visits, reliance on others, loss of income and independence, all observed for even mild visual impairment (Jacobs, Hammerman-Rozenberg, Maaravi, Cohen, & Stessman, 2005). The economic impact is significant; for working-age individuals, the loss of income is commonplace, even with positive action by governments.

Thus, anything that reduces the impact of vision loss is highly desirable.

1.1.1 Blindness

In common parlance, the term blindness is often interpreted as someone with no vision.

This is rarely the case. A person classed as blind can have differing degrees of residual vision.

This is an important distinction as the main aim of this thesis is to attempt to make better use of this remaining vision.

The level of visual impairments as categorised by the WHO (world health organisation)

(Stevens et al., 2013) follows:

- Mild – presenting visual acuity worse than 6/12
- Moderate – presenting visual acuity worse than 6/18
- Severe – presenting visual acuity worse than 6/60
- Blindness – presenting visual acuity worse than 3/60

Vision loss can be grouped into categories;

Central vision loss:

Vision loss starts at the centre of vision and increases to the periphery.

Symptoms: difficulty reading, blurring of vision, inability to see detail.

Examples; Age-related macular degeneration (figure 1.1 B)

Peripheral loss:

Vision loss starts at the edges of vision, increasing towards the centre as diseases progress.

Symptoms: Night-blindness, mobility problems.

Examples; retinitis pigmentosa, glaucoma (figure 1.1 D)

General loss:

Vision loss affects the whole of vision, increasing in severity as diseases progress.

Symptoms; blurriness, cloudy vision

Examples; Uncorrected refractive errors, Cataracts (figure 1.1 E)



Figure 1-1 – Simulation of the effect of vision loss – A: Normal Vision, B: Macular Degeneration, C: Diabetic Retinopathy, D: Glaucoma, E: Cataract.

Economic factors have a significant part to play in the global prevalence of vision loss. Lack of access to eye care is the leading cause of most global cases of vision loss. Uncorrected refractive errors are preventable by correcting refractive errors at a young enough age. This condition is rare in the developed world, with affordable access to eye care. Cataracts are treatable conditions with a routine procedure but require modern facilities, which are

commonplace in high-income countries. However, there still are a wide range of conditions for which there are limited treatments.

In a 2015 review of the prevalence of vision loss in Western Europe by Bourne et al. (Bourne et al., 2018), Cataract was the most common cause, followed by age-related macular degeneration (AMD), glaucoma, uncorrected refractive error, diabetic retinopathy and cornea-related disorders.

For this thesis, I am focusing on wearable digital displays, and the technology used is unlikely to be available in resource-limited regions. Thus, I consider the target population in the developed world. Discussions and prevalence statements will be in that context unless otherwise stated.

For most vision loss, there are few cures. Treatments are focused on limiting degeneration. Developments in gene therapy and stem cells are showing promise. But they are still in early development, and availability is limited.

1.2 Central Vision Loss

In industrialised countries, the most prevalent causes of vision loss principally affect the central part of vision.

1.2.1 The Eye

The eye is the organ of sight, containing structures that focus and capture light photons. The focusing elements are the lens and cornea, located at the front of the eye, focusing objective light onto the retina, the light-capturing part (figure 1.2).

The retina contains cells responsible for capturing and converting light into chemical and electrical energy, known as visual phototransduction. This energy is then communicated via the optic nerve to the brain.

Anatomy of the Human Eye

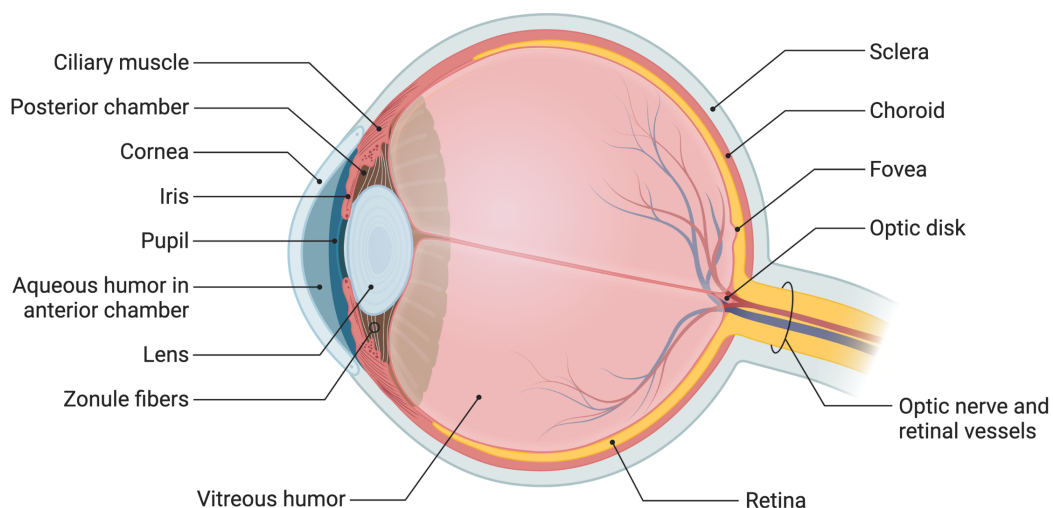


Figure 1-2 The anatomy of the human eye (Created with BioRender.com)

1.2.1.1 The retina

The retina has two classes of light-capturing cells (photoreceptors) called rods and cones.

Rods are highly sensitive and are optimised for low light conditions. Cones are optimised for

brighter conditions; and have three types, distinguished by the wavelength of light they are sensitive to.

The distribution of rods and cones is uneven; rods are significantly more numerous than cones, and their density increase with eccentricity from the fovea (Christine A. Curcio, Sloan, Kalina, & Hendrickson, 1990). Cones are more densely packed into the fovea, which is the focal point of the lens, making our high detail vision.

The retina is a complex structure that can be subdivided into several layers (figure 1.3). The visual sensitivity of the retina is not just driven by the photoreceptors but also by the connective circuitry. The photoreceptor cells are connected to the bipolar and horizontal cells, which in turn are connected to the ganglion cells whose axons collect in the nerve fibre layer and exit the eye through the optic nerve head. The connections terminate in the lateral geniculate nucleus of the thalamus (LGN) (figure 1.3).

The bipolar cells receive signals from the photoreceptors and can be subdivided into cells that are excited by light (ON) or by the lack of light (OFF) (Field & Chichilnisky, 2007). The bipolar cells can also receive input from horizontal cells, which create feedback signals that modulate the gain (S. M. Wu, 1991).

Ganglion cells send signals to the LGN, forming three main pathways, the parvocellular (P), magnocellular (M) and koniocellular (K) pathways. The P-pathway consists of small parvocellular (type-P) ganglion cells that project to the parvocellular layer in the LGN. The receptive field of type-P cells is small and driven by different classes of cones (typically L and M cones). The P-pathway is important for high spatial acuity and colour vision (Solomon, 2021).

The M-pathway consists of large magnocellular (type-M) ganglion cells that project to the magnocellular layer in the LGN. The type-M cell receptive field is larger than type-P and is

driven by all summed activity of L and M cones, so most sensitive to luminance. M-pathway is important for high visual sensitivity and motion vision.

The K-pathway is less well understood, consisting of non-M and non-P ganglion cells projecting to the koniocellular layer of the LGN. Many non-M and non-P receptive fields are from S cones and are likely to contribute to colour perception (Solomon, 2021).

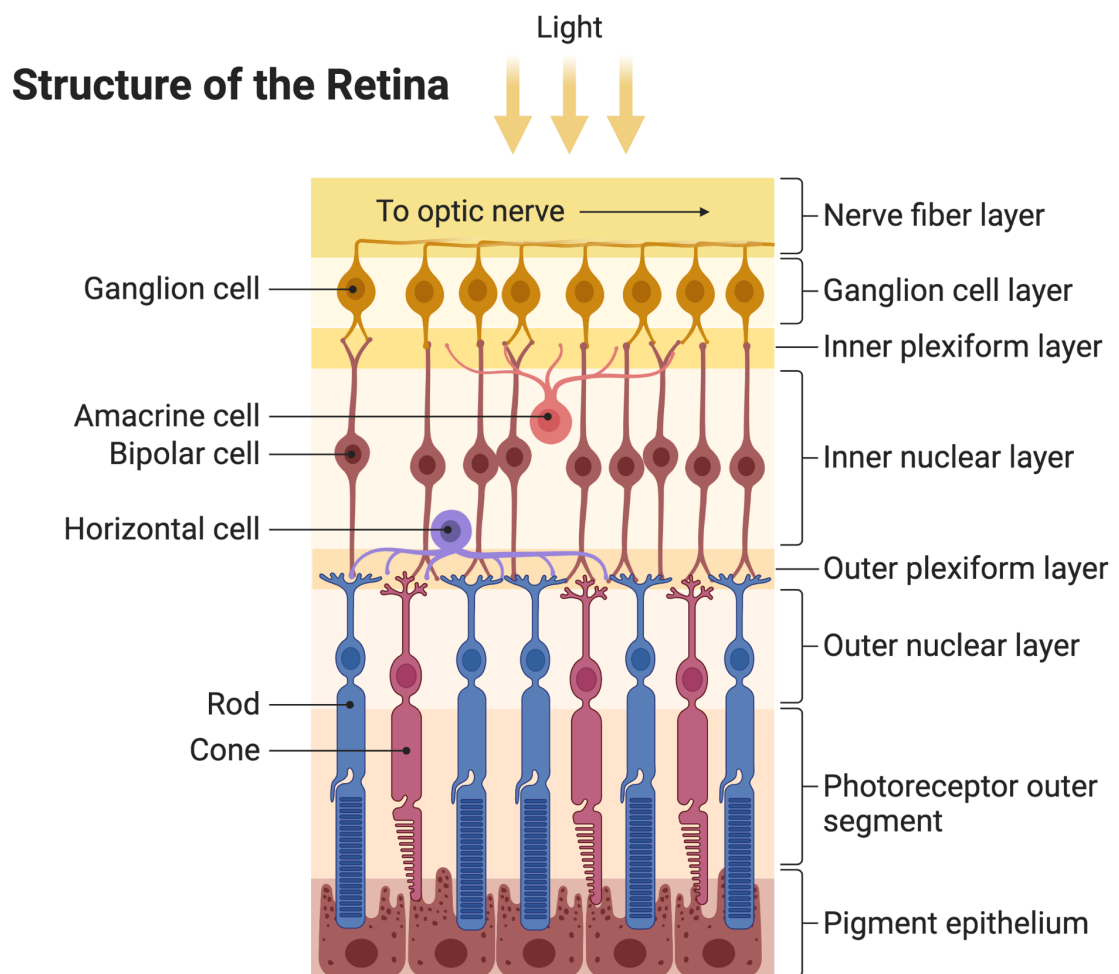


Figure 1-3 The structure of the retina (Created with BioRender.com)

1.2.1.2 The Macula

The macula is an oval area on the retina. It contains several subareas (figure 1.4), including the fovea, which specialises in high-acuity vision containing the largest concentration of cones. The field of view of the fovea is approximately 5 degrees; a standard visual field is an island of vision measuring 90 degrees temporally to central fixation, 50 degrees superiorly and nasally, and 60 degrees inferiorly (SPRY, 2001).

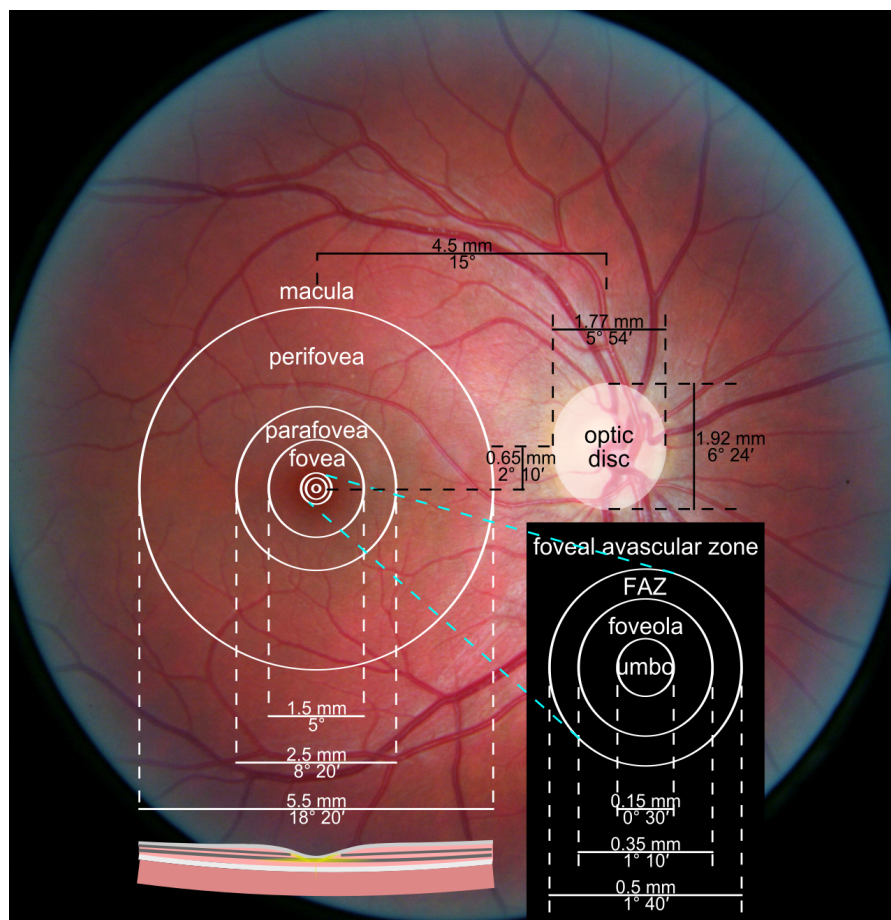


Figure 1-4 The retina - source https://en.wikipedia.org/wiki/Macula_of_retina#/media/File:Macula.svg

Any damage to the cells in this area will impact the central part of vision and will likely be noticed immediately. Conditions that progressively affect this area are known as macular degeneration.

1.2.2 Macular degeneration

Central vision loss is symptomatic of many different diseases. The most prevalent form of central vision loss in the developed world is age-related macular degeneration (AMD)(Bourne et al., 2018).

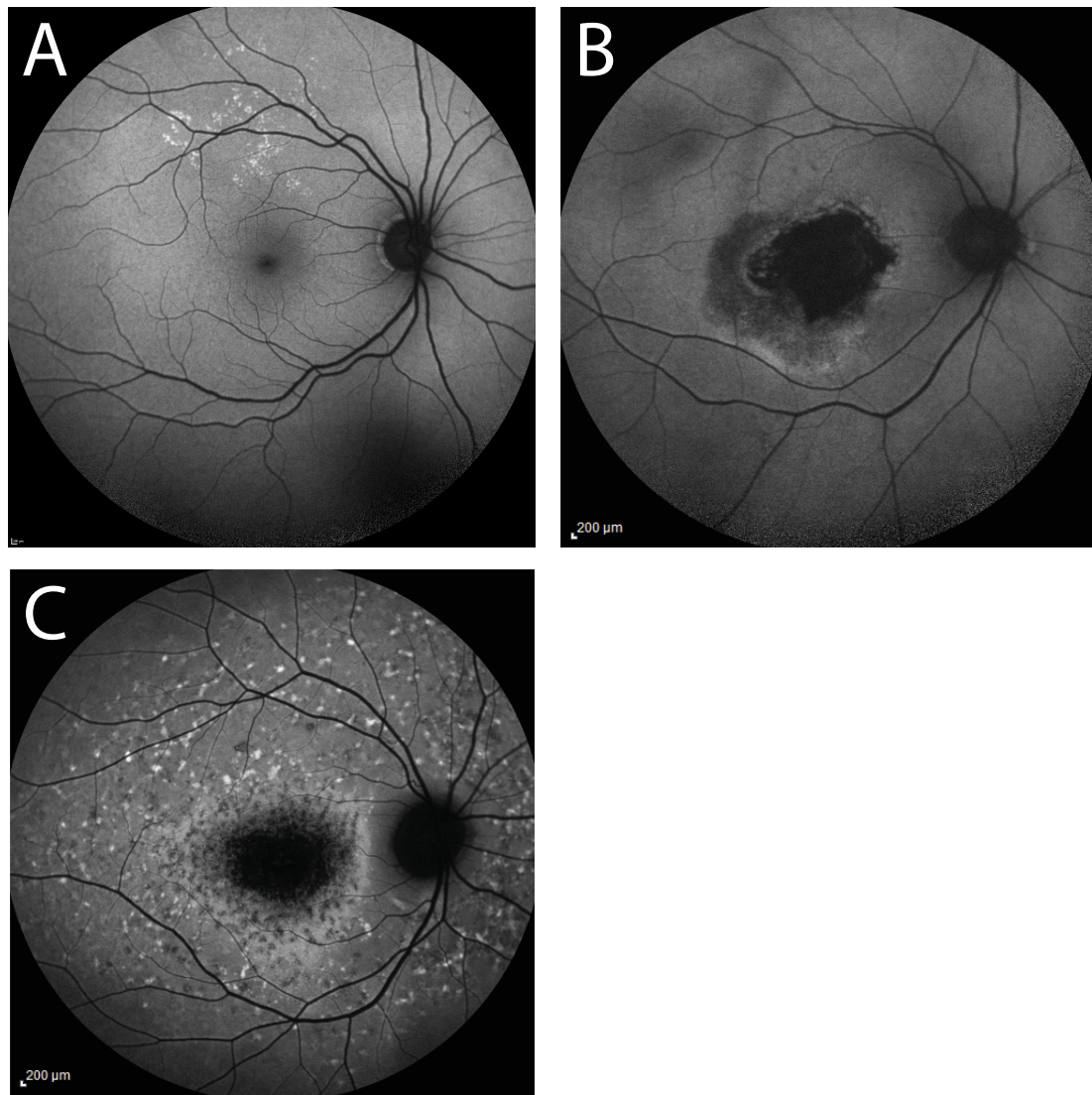


Figure 1-5 Autofluorescence images of A normal eye, B Age-related macular degeneration, C Stargardt. Autofluorescence Imaging uses naturally occurring fluorescence from the retina to provide an indicator of RPE health. In the above image A of the normal eye, the blood vessels and optic nerve head can be seen not to fluoresce, with the fovea slightly darker in the centre of the image. B, the AMD eye, geographic atrophy is observed as the darkening central area over the location of the fovea, producing a significant central scotoma and sight loss. Image C is a Stargardt eye with similar atrophy to the AMD at the fovea and 'flecks' over the retina that is typical of the disease.

1.2.2.1 Age-related macular degeneration

AMD is the most significant cause of sight loss in the United Kingdom (UK), affecting more than 600,000 people (Owen et al., 2012).

AMD is caused mainly by retinal pigment epithelium (RPE) damage and by damage to Bruch's membrane in the choroid. The RPE is a cell layer in the retina that nourishes the visual cells. Damage to the RPE can lead to neovascularisation (known as “Wet” AMD), where new blood vessels are formed and fluid can leak into the retina, causing scarring. Geographic atrophy, known as “Dry” AMD, may coexist or occur separately and is associated with photoreceptor death. The early stages of AMD may not cause symptoms, whereas the later stage is always associated with some loss of visual function.

AMD symptoms vary, but vision loss typically starts at the macula and progresses into the mid periphery; it is rare for total visual loss unless there is a catastrophic bleed. Sufferers begin to notice a loss of visual acuity; difficulties with reading are often the first sign. Optical distortion is common when the photoreceptor layer is disrupted, resulting in metamorphopsia. In the wet type, if left untreated, it can lead to scotomas; an area of the retina with no light sensitivity due to photoreceptor death (figure 1.5 B). There is currently no treatment for the dry type.

1.2.2.1.1 Prevalence

AMD affects those later in life; A study on the global prevalence (Wong et al., 2014) found the overall prevalence of any type of AMD is about 8% within the age range of 45->84. It ranges from ~3% in the 45-49 age group to ~20% in the 80->84 group. The majority of these cases are early-stage diseases; for the late-stage and more likely signs of vision loss, the total prevalence is ~0.3%, ranging from 0.05% (45->49) and 3.25% (80->84).

Wong et al. projected the global number of people with age-related macular degeneration in 2020 to be 196 million, increasing to 288 million in 2040 (Wong et al., 2014).

Further work by Li et al. (J. Q. Li et al., 2020) focused on the pre-Brexit European Union (EU) and found that the prevalence estimate of early or intermediate AMD and any late AMD in those 60 years and older was 25.3% and 2.4%, respectively.

Overall, the number of EU inhabitants (including the UK) with any type of AMD is expected to increase from 67 to 77 million by 2050. Incident late AMD is estimated to grow from 400,000 per year to 700,000 in 2050 (J. Q. Li et al., 2020).

1.2.2.1.2 Treatment

For the wet type, the current interventional therapies involve blocking Vascular Endothelial Growth Factor (VEGF), which reduces choroidal neovascularisation and vasopermeability. Delivered into the vitreous by injection, this has been shown to limit disease progression and improve visual acuity in some patients (Lazzeri, 2013). However, this is an expensive treatment and is limited to the more affluent countries. For the dry form of AMD, there are no available treatments.

1.2.2.2 Other conditions

Diabetic retinopathy

The second leading cause of irreversible central vision loss is diabetic retinopathy (DR), a complication of diabetes. Bleeding can occur when fragile new blood vessels which develop in uncontrolled diabetes. This can disrupt vision and, in later stages, scarring and sometimes

retinal detachment, leading to more significant vision loss. The global incidence rate of DR ranges from 2.2% to 12.7% (Sabanayagam et al., 2019). Within the United Kingdom, a 2014 study (Mathur et al., 2017) found the prevalence is 48% in the population with type 1 diabetes and 28% for type 2.

The visual symptoms are not necessarily focused on the macula, as mid and peripheral field loss is also experienced in DR due to ischaemia, laser treatment or complications of DR (figure 1.1 C).

Macular dystrophy

These are a group of inherited disorders that can cause central vision loss (Rahman, Georgiou, Khan, & Michaelides, 2020), the most common is Stargardt disease. Stargardt disease is an inherited macular dystrophy caused by a single gene disorder (ABCA4). Its prevalence in the population is much lower than AMD, affecting 1 in 8-10,000 (Genead, Fishman, Stone, & Allikmets, 2009). Vision loss typically occurs in the first 10 to 20 years of life. Vision loss starts in the centre and progresses centrifugally. Generally, the periphery is preserved, but complete vision loss may occur.

Other macular dystrophies include Best disease, X-linked retinoschisis, pattern dystrophy, Sorsby fundus dystrophy and autosomal dominant drusen. The group has various pathologies and genetics, but when symptomatic are generally so due to central vision loss.

1.2.3 How is central vision loss measured?

1.2.3.1 *Visual Acuity*

Central vision loss is typically measured in terms of visual acuity (VA). Visual acuity is a measure of the spatial resolution of the visual processing system. This resolution is dependent on several components within the eye; optical the light focusing elements (lens, cornea), retinal light-receiving parts (photoreceptors), and neural; the nerve connections from the retina and the visual processing areas of the brain. A loss of functionality of any of these components due to disease or trauma can cause a reduction in visual acuity.

Visual acuity is commonly measured by an individual identifying optotypes on a chart. The visual angle of the lowest size read is used to calculate their visual acuity. Optotypes are often a letter or similar symbol, and the visual angle is the angle that which the optotype appears at the eye. Normal vision is considered the ability to discern an optotype with a visual angle of 5 arc minutes. Of note, this is measured independent of any optical refractive errors (such as myopia); tests are performed with the optimal eyeglass prescription (up to date refraction).

Testing charts contain a series of decreasing-sized optotypes, which are typically used to measure visual acuity. The Snellen Chart has been used for many years and defines normal sight as “6/6,” where the standard viewing distance is 6 meters. However, it is gradually being replaced by the logMAR scoring system. In this thesis, I will be using the logMAR (Logarithm of the Minimum Angle of Resolution) scoring system, together with the Early Treatment Diabetic Retinopathy Study (ETDRS) chart for visual acuity measurements. The ETDRS chart was primarily designed for clinical studies and low vision evaluations, where accurate and repeatable visual acuity measurements are required (figure 1.7).

1.2.3.2 Contrast sensitivity

Contrast sensitivity refers to the ability to discern a light-dark transition. This ability is a feature of the organisation of photoreceptors and retinal ganglion cells, known as the centre-surround (Allman, J; Miezin, F; McGuinness, 1985). Centre-surround is where the response of the ganglion cell is determined by the differential of signals from the interior and exterior of a receptive field. (Figure 1.6). The size of the receptive field defines the spatial resolution to which they are sensitive. At the fovea, the field size is the smallest, where the ratio of photoreceptors to retinal ganglion cells is lowest and is most sensitive to high spatial frequencies. The receptive field size increases eccentrically, with this size increase the frequencies they are sensitive to are reduced (Christine A. Curcio et al., 1990).

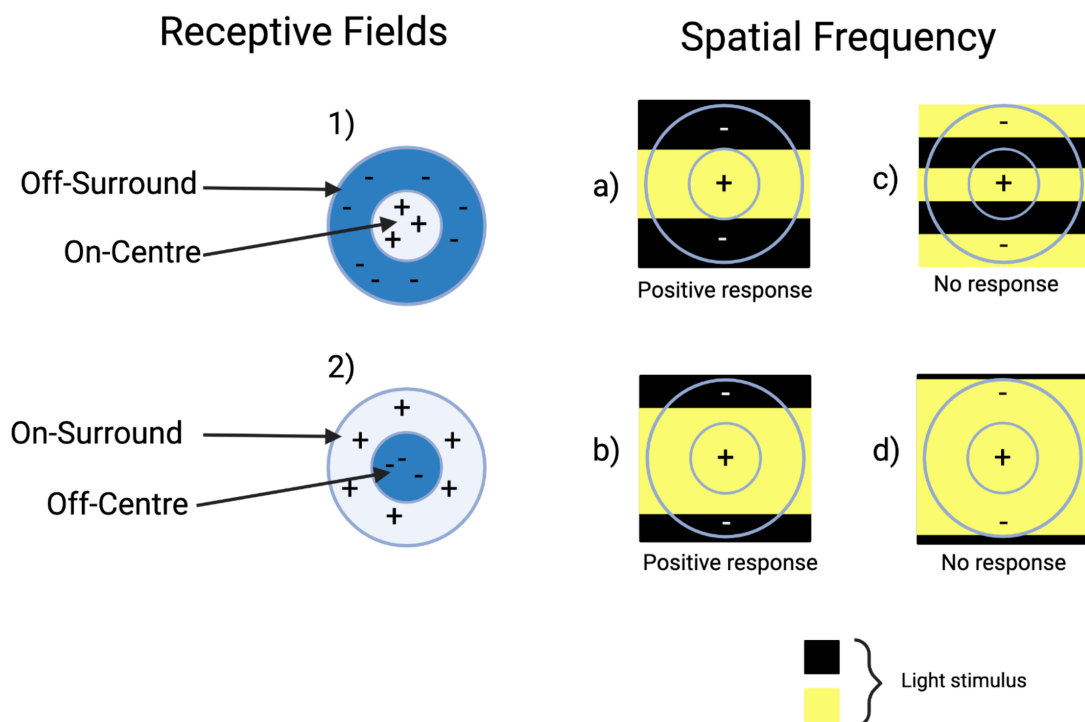


Figure 1-6 - Centre-surround receptive field. Figures 1) and 2) represent the receptive field, 1) the exterior is deactivated with no illumination (Off-surround), relative to the interior (On-Centre). 2) circles represent the opposite effect; the centre is deactivated and the surround is activated. Figures a) and b) represent spatial frequencies that evoke a positive response. c) and d) represent spatial frequencies that do not evoke a response.

An archetypal feature of central vision loss is a loss of contrast sensitivity, often occurring before any noticeable loss of visual acuity (Marmor, 1986). Contrast sensitivity loss is more prominent at the higher spatial frequencies at early disease stages, progressing to lower spatial frequencies (Enger, Alexander, & Fine, 1988). Sensitivity loss is linked with ageing, where visual acuity is not changed. Age has been seen to impact high and medium frequencies, but less the lower frequencies (Ross, Clarke, & Bron, 1985).

Contrast sensitivity is measured clinically with several methods; the most common is the Peli-Robson chart (Pelli, D. G., & Robson, 1988) a printed chart that produces a scalar value as a measure of sensitivity. Computerised tasks have the advantage of providing a more robust and detailed information (Pelli & Bex, 2013).

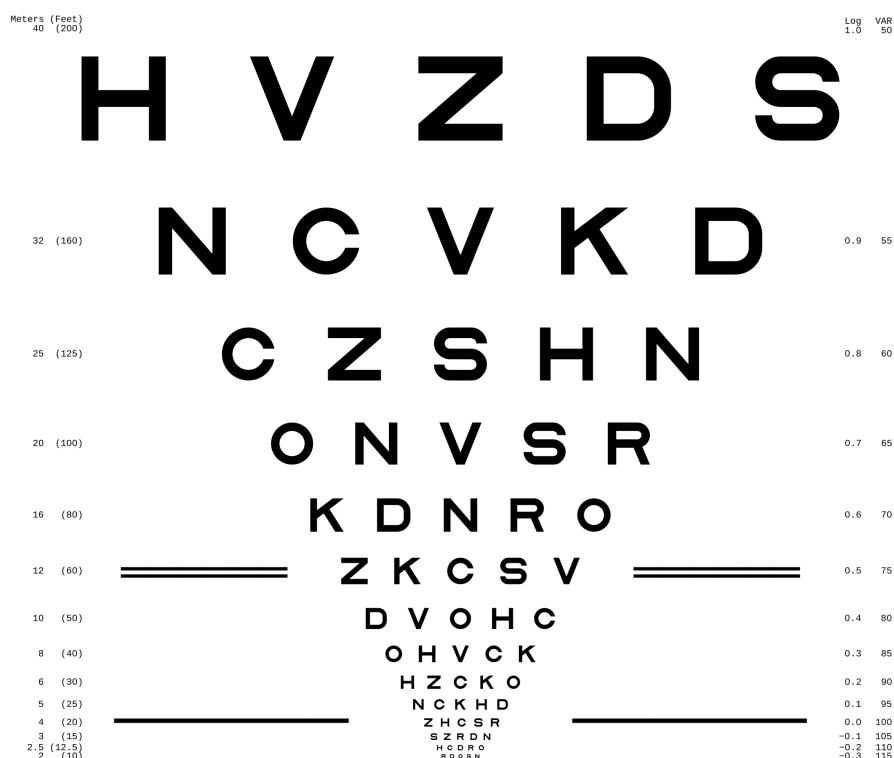


Figure 1-7 Early Treatment Diabetic Retinopathy Study (ETDRS) chart – source: https://commons.wikimedia.org/wiki/File:ETDRS_Chart_R.svg.

1.2.3.3 Colour Vision

Colour vision is trichromatic, meaning there are three channels for conveying colour information. Colour vision is an important function of cone photoreceptors, of which there are three types; differentiated by their sensitivity to different wavelengths of light. The three cones are named by their peak wavelength sensitivity L (long, 560 nm), M (medium, 530nm) and S (short,420nm). The density of these three types is asymmetric, with S-cones absent from the foveola and comprising approximately 5 to 10% of all cones (Solomon & Lennie, 2007). The ratio of L to M cones varies considerably but is thought to be in males 1.1:1 to 16:1 (Simunovic, 2016a).

The responses from the three cone types are combined in the bipolar layer of the retina. The L and M cones are contrasted to indicate Reds and Greens, and the S cones are contrasted to the L and M to indicate Blues and Yellows. This contrasting process is known as the cone opponent system, (L+M), (L-M) and (S-(L+M)) (Solomon & Lennie, 2007).

Defects in cone photoreceptors, can cause a bias in their peak wavelength or inhibit their response leading to a colour vision deficiency. These defects can be congenital or acquired and are classified by the type of photoreceptors they impact. Deficiencies in the L-cones are referred to protanopia, M-cones deuteranopia and S-cones tritanopia (Simunovic, 2016a), referred to as protan, deutan and tritan from herein.

The acquired form of colour vision deficiency is of interest in the context of this thesis and is common with central vision loss. Colour matching can detect early changes in AMD, with several studies identifying significant changes in colour sensitivity before visual acuity is impacted (Arden & Wolf, 2004; Smith, Pokorny, & Diddie, 1988). Tritan defects are generally found to be higher with AMD (Vemala, Sivaprasad, & Barbur, 2017). Conversely, evidence from Vandenbroucke et al. studying Stargardt disease found early colour loss is red/green, with 45% the deutan type, moving to protan type as the disease progresses, and tritan type in 25% of their cohort (Vandenbroucke et al., 2015).

In general, acquired colour deficiency is common with retinal diseases. Some diseases have more predictable manifestations, others less so. The exact mechanisms for colour deficiencies are complex and currently not fully understood (Simunovic, 2016b).

1.3 Impact of central vision loss

The impact of central vision loss is understandably significant. High-acuity vision is a vital component of life, and the loss of it will impact all activities of daily living.

The majority of conditions, i.e. AMD & Stargardt, are progressive and vision deteriorates over time. Initial symptoms may go unnoticed, e.g. slight difficulties with reading can be alleviated with better lighting. As diseases progress and patients become symptomatic, there is accompanying loss of visual acuity, contrast sensitivity, and visual distortions. The degree and severity of loss will depend on the individual and stage of disease (Taylor, Edwards, Binns, & Crabb, 2018).

The impact on quality of life is complex but significant. Work by Hassell et al. (Hassell, Lamoureux, & Keeffe, 2006) explored the effect of vision loss on life outside of reading

difficulties. They found that the most important items that impacted the quality of life were associated with reading, leisure activities, shopping, and emotional reaction to vision loss. They also found that the severity of impact is similar for those with mild to moderate vision loss. Sufferers are frequently depressed (O'Donnell, 2005) and have almost twice the risk of injury-causing falls (Szabo, Janssen, Khan, Lord, & Potter, 2010).

The economic impact of vision loss in developed countries is significant. Pezzullo et al. (Pezzullo, Streatfeild, Simkiss, & Shickle, 2018) studied the economic impact of sight loss on the UK population, estimating the 2013 cost to be £15.8 billion. This included the direct health care costs and indirect costs (e.g. lower workforce participation and the cost of informal care).

1.4 Aids for the visually impaired

Numerous aids for the visually impaired are available, ranging in simplicity and expense. The variability of aids reflects the heterogeneous nature of visual impairments and that there is currently no one perfect solution. As the focus of this thesis is central vision loss, I shall be discussing aids that are targeted at the typical challenges faced by those individuals.

1.4.1 Traditional low vision aids

Aids for central vision loss are numerous but fall into two categories; Near and Far. For near vision tasks, handheld optical magnifiers are commonly used and frequently contain an integrated light source to illuminate and increase contrast. Digital handheld magnifiers are widely used and desktop-based digital magnifiers, known as CCTV (closed-circuit television).

The latter provides people with a large field of view, not offered by handheld versions.

There is an extensive range of choice and availability of these devices, with an equal range in price. Optical magnifiers are a cost-effective solution but have limited functionality.

Digital magnifiers can improve on this functionality, but with an added expense. Computer screen readers are valuable tools, and modern computer operating systems have numerous accessibility features, such as zoom and contrast inversion.

For far-vision tasks, telescopes and binoculars are used. Some can be attached to spectacles or the more traditional handheld devices. Digital versions of these are less common, partly as they do not provide the advantages their traditional counterparts do. Golubova et al.

(Golubova, Starke, Crossland, & Wolffsohn, 2021) observed the behaviour of visually impaired participants using low vision aids and found that the satisfaction with the aids was lowest with distance viewing.

Support services such as low vision clinics and sight loss charities typically recommend these aids for the visually impaired. With support and training, they can provide significant benefits for the user.

1.4.2 Smart Low Vision Aids

The advent of smartphones and tablets has been beneficial for the visually impaired population. Most devices contain high-resolution cameras and integrated lights and can act as simple digital magnifiers. The two major operating systems (Apple IOS and Android) also include low-vision accessibility tools by default, such as magnification, text-to-speech and contrast inversion. Martiniello et al. (Martiniello, Eisenbarth, Lehane, Johnson, & Wittich,

2019) explored the increasing use of smartphones and tablets. They noted that they are replacing the more traditional visual aids but found that the traditional methods are still preferable for some tasks, particularly for extensive typing. Wittich et al. (Wittich, Jarry, Morrice, & Johnson, 2018) investigated the effectiveness of an Apple iPad as a magnifier for reading and found that it compared equally to a traditional CCTV device.

1.4.3 Smart Apps

The advantages of smart devices are their general-purpose nature and the extensive availability of application development tools. Creating new applications that provide simple functionality (e.g. magnification such as SuperVision+ used by (Wittich, Jarry, et al., 2018)) requires only a small amount of software engineering ability and are usually offered at little to no cost.

More complex applications for smartphones and tablet computers are constantly being developed, thus extending functionality beyond traditional low vision aids. Utilising the ever-increasing computing power and internet connectivity, they can provide functionality and services previously unobtainable. One of the most sophisticated is Seeing AI, developed by Microsoft Corporation (Redmond, Washington, USA). Leveraging advances in computer vision and machine learning and using the device's camera, the software can; semantically describe a scene, read text (even handwriting), recognise currency and people.

Video conferencing technology is now being used to connect a visually impaired user to a sighted individual (aid) who can provide visual assistance. Be My Eyes App (Be My Eyes, San Francisco, CA, US) is a free application that uses volunteers, and Aira (Aira Tech San Diego, CA, US) is a paid-for service using trained agents. Aira was the focus of a study by Park et al.

(Park, Kim, Nguyen, Chen, & Chao, 2020) and found it positively impacted individuals with very low vision.

Many applications support low-vision, such as the Amazon (Seattle, WA, USA) Kindle reading app and Instagram social media. Underlying accessibility support and guidance provided by the major operating systems (Apple IOS and Google Android) simplifies low-vision support for developers and will increase the usability of smartphones and tablet computers for the visually impaired.

The usefulness of smartphones and tablets as low-vision aids is apparent. However, the array of different applications and devices can be daunting. Robinson et al. (Robinson, Braimah Avery, Chun, Pusateri, & Jay, 2017) investigated the usage of accessibility options for the mobile operating system IOS (Apple, Cupertino, California, USA), concluding that while commonly used, little training is provided to make the best use of the options for someone's particular needs. This was also noted by da Silva et al. (da Silva, Leal, & Ferraz, 2020) who recommended that such devices be accompanied by more support from rehabilitation programmes.

1.4.4 Acceptance of LVAs

There is a low acceptance rate for the use of low-vision aids. A study by Sivakumar et al. (Sivakumar et al., 2020) identified that over half of the participants who noticed an improvement from using a low vision aid (n=413) were unwilling to use one. The most common cause was a social stigma, fear of loss of employment, and low necessity perception.

The acceptance of types of devices, Sivakumar found high acceptance with hand and stand magnifiers and lower acceptance with telescopes and electronic devices. Cost barriers are a factor, with digital readers starting from £125 for a small handheld device to £2000 for desktop CCTV (source <https://shop.rnib.org.uk/magnification/magnifiers/electronic-magnifiers>). These prices are a barrier for many, even in advanced economies.

Low acceptance rates highlight the challenges for low-vision aid designers. Devices need to be cost-effective, discreet, and ultimately provide a significant functional benefit.

1.5 Head-mounted displays

Broadly, a head-mount display (HMD) is a display that is worn on the head and images are presented to both eyes in a binocular or bi-ocular (same image shown to each eye) fashion.

The first head-mounted display, Headsight, was developed by the Philco Corporation (Philadelphia, US) in 1961. It was a helmet comprising of a cathode ray tube and a head tracking system. In 1965 Ivan Sutherland developed the “Ultimate Display” (Sutherland, 1965), where he outlined the idea of a virtual room where the user would not be able to distinguish simulation from reality. Then in 1968, Sutherland and Bob Sproull developed the first augmented reality head-mounted display, the “Sword of Damocles”, that required a ceiling mount to carry the weight. The name is derived from the Greek story of Damocles, in which a sword was suspended over the king’s throne by a hair. The first commercial HMD Virtuality was released by Virtuality Group in 1991. The application of HMDs has ranged from psychological research, entertainment and military training (Boas, 2013).

A more recent renaissance of interest in HMDs was triggered by the release of the Oculus Rift¹ (Facebook, Inc, Menlo Park, CA, US) in 2013, an affordable commercial virtual reality display. Since then, there has been a boom in the development of HMDs, with ever-increasing display resolution and reducing weight and cost.

1.5.1 Virtual and Augmented reality

There are two main categories of head-mounted displays, Virtual Reality (VR) and Augmented Reality (AR). VR is fully immersive; the visual information presented to the wearer is computer-generated, placing the user in a virtual environment. AR is an augmentation of reality; the images shown are overlaid onto the real world, combining the virtual with the real.

There are several display methodologies used, with different characteristics:

Fully-occluded

The displays are embedded within an enclosed head-mount; the wearer can only see the displays; the rest of the world is occluded from view. This is typical of modern Virtual Reality (VR) displays, such as the Oculus (Facebook) and HTC Vive. The field of view is typically large (90 degrees) and is usually binocular. The display technologies are usually Liquid crystal displays (LCD) and organic light-emitting diodes (OLED), which produce high luminance and good colour definition.

Semi-occluded

Displays are embedded in an open head-mount, where views of the world are not occluded. The wearer can see the display and perceive the world in their peripheral vision. Typically

¹ Oculus was purchased by the Facebook corporation in 2014

having a smaller field of view than the fully-occluded version, it uses the same display technology with similar luminance. However, the perceived brightness may not be as high in bright environments.

See-through

The display is transparent, and images are projected onto the transparent display. The field of view is limited and highly dependent on the exact technology used; modern commercial displays range from 20-40 degrees. These are also referred to as augmented reality (AR) devices, where the images presented in the displays are intended to augment and complement the view of the world.

AR devices can include fully, and semi occluded displays; however, in this thesis, I will use VR for all fully occluded devices and AR for see-through devices.

Modern HMDs can provide knowledge of the head pose; this is particularly important for fully occluded displays. To avoid any visual conflicts with the vestibular system and induce motion sickness, the rendering of the virtual environment must accurately reflect an individual's head pose. This is usually achieved by integrating an accelerometer and gyroscope into the HMD. This is acceptable for some applications, but it only provides pitch, roll and yaw, not translational motion.

Translational movement is monitored by either an outside-in or inside-out approach.

Outside-in uses external sensors to monitor the position of the HMD within the environment. Inside-out is where the HMD has sensors (usually cameras) facing outward and track its position relative to the world. The outside-in approach has the limitation of more components and limits the working area, but it is reliable and accurate. Inside-out is compact; all the parts are integrated into the HMD and is theoretically unlimited in its

working area. However, the tracking is less reliable as it is highly dependent on easy to track features in the environment.

Augmented reality HMDs have a requirement of at least one front-facing camera that enables the augmentation. Simple systems only have one camera (such as the Epson Moverio BT-200/300). Others utilise more complex imaging systems, utilising depth-sensing technologies and multiple cameras (e.g. Microsoft HoloLens) to enable complete and accurate mapping and augmentation of the world.

1.5.2 Head-mounted displays and low vision

In their simplest form, head-mounted displays work as a hands-free magnifier. But their digital nature allows for the opportunity to perform image enhancements. E Peli et al. (Fine & Peli, 1995; Eli Peli, Lee, Trempe, & Buzney, 1994) has shown contrast enhancement to improve the visual experience over simple magnifications alone. Edge detection techniques have also been used to enhance contrast using a monocular AR device (Hwang & Peli, 2014). Image remapping methods have been tried with some success (Aguilar & Castet, 2017). The concept is to remap visual information so that it does not fall within the patient scotoma; however, it has the added challenge of eye-tracking, and the benefits over simple magnification are not clear. Novel use of a scene depth-representation, with a custom-built semi-occluded HMD by Hicks et al. (Joram J. van Rheede et al., 2015) was shown to improve mobility. Using a similar depth-representation, Kinatader et al. (Kinatader et al., 2018) used a see-through HMD (Microsoft HoloLens), finding improved object recognition but not mobility.

Prior work by myself was the further development of the work by Hicks et al. (Joram J. van Rheede et al., 2015), using a depth and colour camera to improve the visibility of a scene for the visually impaired (I. R. Wilson, Van Rheede, Campbell, MacLaren, & Hicks, 2016). The Smart Specs (figure 1.4a) comprised a see-through display, a colour and a depth-sensing camera with custom control electronics. This device was assessed by 105 visually impaired participants in object recognition and navigation tasks (Joram Jacob Van Rheede et al., 2016). Approximately 30% of participants gained an objective benefit using the display. Of those, individuals with some spared central vision were more likely to perceive an advantage. A lack of benefit in patients with central scotomata was hypothesised to be a limitation of the small display size (23° field of view), thus excluding individuals with central vision problems. Another likely reason is the device was focused on assisting with mobility; mobility issues are rare for central vision loss conditions where peripheral vision is preserved. However, a few individuals with central vision loss did find an objective and subjective benefit whilst using the Smart Specs. This finding is a motivating factor for this thesis, investigating the advantages of head-mounted display technologies for people living with central vision loss.



Figure 1-8 - Smart Specs. Control processing unit and headset. The control unit houses the image processor, battery and control electronics. The headset is comprised of an Epson BT-200 display, colour and depth cameras.

1.5.2.1 Commercial HMD devices in low vision

Several commercial devices are currently available and have been shown to provide an improvement objectively: table 1 list presents currently available devices, their specifications and cost. A few research studies have focused on some of these devices. Crossland et al. (Crossland, Starke, Imielski, Wolffsohn, & Webster, 2019) investigated the SightPlus (GiveVision, UK, givevision.net) device, finding that approximately half of their participants perceived a benefit. Gopalakrishnan et al. (Gopalakrishnan, Chouhan Suwalal, Bhaskaran, & Raman, 2020) used a similar device (Samsung GearVR) and showed an improvement in visual acuity. Wittich et al. (Wittich, Lorenzini, et al., 2018) investigated the eSight display(eSight Corp, Toronto, Canada), which improved visual acuity and face recognition.

All devices listed in Table 1 are either semi or fully-occluded displays, except for the OxSight Crystal (OxSight, Oxford, UK). The see-through nature of the OxSight Crystal allows them to

be used whilst walking, but their main limitation is their comparatively small field of view.

All of the devices provide some variation of image contrast enhancement, magnification and colour enhancement techniques. The IrisVision (IrisVision Global, Inc, Pleasanton, CA, USA) and NuEyes Pro (NuEyes Technologies Inc. Newport, CA, USA) also offer optical character recognition and read text aloud to the user. The enhancement methods are passive, meaning that they do not rely on any semantic information within the scene. Users select the enhancement depending on the situation.

The target user for these devices is broad, as the manufacturers try not to narrow their market. Still, the features (magnification and contrast enhancement) are generally focused on central vision loss. However, the IrisVision display has a presentation mode described as “RP Mode” focused on visual field restricting conditions such as Retinitis pigmentosa and glaucoma. This mode ‘zooms out’, reducing the size of the displayed image to fit within the user’s field of vision. This feature is also available in the OxSight Crystal, which also targets field restricting conditions.

Table 1-1 Commercial Head-mounted displays

<i>Product Name</i>	eSight 3	GiveVision SightPlus	OxSight Crystal	IrisVision	NuEyes Pro
<i>Display resolution</i>	1024 x 768	1480 x 1440 *	1280 x 720	1210 x 920	1280x720
<i>Pixel visual angle arc min / px</i>	2.2 arcsin px	~3.3 arc min/px	1.0 arc min px	3.3 arc min/px	1.4 arc min/px
<i>Best visual acuity (logMAR)</i>	0.34	~0.5	0.0	0.5	0.14
<i>Field of view (deg)</i>	37.5 x 28	80 x 75	20 x 11	70 x 50	30 x 17 deg
<i>Display Type</i>	semi-occluded	fully-occluded	see-thorough	Fully occluded	Semi-occluded
<i>Cost</i>	£6,995.00	£2,995.00	£4000.00	£2,895.00	£5,995.00

Pricing obtained from <https://www.visionaid.co.uk/wearable-solutions> 01/03/2021

* Approximate as the manufacture does not report exact values.

1.6 Thesis Overview

While there is significant development with wearable displays for the visually impaired and several commercially available devices, these devices are targeted as broadly as possible, lacking focus on individual tasks. It is often unclear who would benefit from their usage. The image enhancements they provide are typically applied globally. In other words, they are applied to the whole image, rather than focusing on the part of the images of interest. There is little information about the methodologies used and their impact.

The aim of this thesis is to focus on reading and face perception, as the loss of these activities has the most significant impact on the lives of people with central vision loss. I assess the accuracy of performance with image enhancements and implement them with a head-mounted display. With the secondary goal of understanding for whom these approaches are most likely to confer a benefit.

The proceeding section will discuss face perception and how it impacts people with central vision loss.

Chapter 2:

Face Perception

My contributions to this chapter: The writing is my original work and was conducted over a 6-month period.

Other contributions: Professor Robert MacLaren and Professor Susan Downes assisted with proofreading.

2 Face Perception

This chapter discusses face perception in the context of central vision loss, and how it is measured and potentially improved.

2.1 Introduction

Face perception is a core feature of human social interactions. We can learn a considerable amount of information about an individual from a single image or glance at a face; Identity, age, sex, mood and attention. It is vital in communication; response cues and emotions are critical in understanding each other. Therefore, it should be no surprise that any impairment or loss of face perception can significantly impact the quality of life.

In this chapter, the aim is to improve face perception abilities for people with central vision loss. The intention is to apply relevant image processing to an image of a face presented on a wearable display. To achieve this, first, an image enhancement approach has to be identified, its performance investigated and quantified in individuals with central vision loss. Finally, the approach needs to be implemented to run in real-time on a wearable display.

2.1.1 Why faces?

Aside from the social importance, why should there be a need to focus an image enhancement just on faces? The reason for this is the unique way that faces are perceived.

2.1.2 The whole and the part

The process of face perception is thought to be holistic; a face is seen as a whole rather than a collection of its parts (Maurer, Le Grand, & Mondloch, 2002; Sergent, 1984; Tanaka & Simonyi, 2016). There is evidence that unique neurological mechanisms are dedicated to face perception, referred to as the fusiform face area (Kanwisher, McDermott, & Chun, 1997). This holistic process arguably separates face perception from other forms of object detection. Experimental evidence shows the degraded performance when comparing line drawings of faces to photographs (Davies, Ellis, & Shepherd, 1978). This is in contrast to object recognition, which does not suffer from such degradation, where the objects are identified by disassembly into constituent parts (Farah, 1991).

2.1.3 Are all parts equal?

While there is good evidence for the holistic processing of faces, the amount that each facial component (e.g. eye, nose, lips, etc.) is not equal, work by Sadr et al. (Sadr, Jarudi, & Sinha, 2003) identified that eyebrows are essential features. A more recent study by Abudarham and Yovel (Abudarham & Yovel, 2016) assigned a perceptual sensitivity value to 20 different facial features. They used this perceptual sensitivity value to identify a subset of critical features for face identification (lip thickness, hair colour, eye colour, eye shape, eyebrow thickness). This was experimentally validated by changing these features, which changed the perceived identity of a face. In contrast, changing lower perceptually sensitive features did not alter a face's perceived identity. It is important to note, as the authors did, that further work would need to be carried out in all ethnicities to confirm the same perceptual sensitivity as this has only been performed in white Caucasians.

2.1.4 Friend or Foe?

There is some evidence to suggest different viewing strategies are used depending on familiarity (Ellis, Shepherd, & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985), where the external features are thought to be more critical for unfamiliar faces, and the internal (i.e. nose, eyes, mouth) for familiar faces. However, Abudarham & Yovel (Abudarham & Yovel, 2019) did not find this the case and noted that a subset of facial features critical for face recognition (Abudarham & Yovel, 2016) did not change with familiarity.

While face perception is a well-studied research area, it is highly complex, and our understanding of all processes is far from complete. Yet, it is clear from the evidence that faces are unique and should be considered a specific case in image enhancements for the visually impaired.

2.2 What causes face perception loss?

There are several sources of face perception loss other than eye disease, for example, the natural ageing process and neurological conditions. Prosopagnosia is an example of a neurological disorder that can be either congenital, developed or acquired, which results in the loss of the ability to recognise faces. These forms of neurological conditions are usually distinguishable from eye disease. The most prevalent form of vision loss (in the industrialised world) is Age-related Macular Degeneration (AMD). However, we must take care to consider the natural ageing process as an influencing factor.

2.2.1 Age-related decline – retinal, optical or neurological?

Face recognition abilities are known to decline with age. Owsley et al. (Owsley & Sekuler, 1981), Lott et al. (Lott, Haegerstrom-Portnoy, Schneek, & Brabyn, 2005) and Norton (Norton, McBain, & Chen, 2009) showed that the age-related decline of contrast sensitivity (Sekuler, Hutman, & Owsley, 1980) is related to a loss in face perception. It is difficult to identify if this effect is retinal, neural or optical. Elliot (Elliott, 1987) indicated that the decline in contrast sensitivity with age is retinal and neural rather than optical. Later work by Elliot using adaptive optics noted that photoreceptor loss does occur with age (Panda-Jonas, Jonas, & Jakobczyk-Zmija, 1995). Curcio et al. (C. A. Curcio, Millican, Allen, & Kalina, 1993) found stability in cone photoreceptor density. Still, a loss of rods was noted, implying any drop in visual acuity with age is likely related to optical factors, not cone loss. Calder et al. (Calder et al., 2003) and Ruffman (Ruffman, Henry, Livingstone, & Phillips, 2008) found older adults showed a reduction in recognition of some facial emotions. They associated this change with the typical neurological ageing process.

2.2.2 Vision loss and Face Perception

The loss of face recognition ability is a common functional symptom of central vision loss (Taylor, Hobby, Binns, & Crabb, 2016) and significantly impacts the quality of life (Lane et al., 2018). Typically, as a central vision loss in a condition such as AMD progresses, photoreceptors and retinal ganglion cells are lost. This starts at the fovea and expands outwards into the posterior pole. This results in the loss of high spatial frequencies, where the finer details are attenuated and begin to become indiscernible. With progression, this reduction in sensitivity can become a complete loss, forming a central scotoma. Individuals

start to use a different fixation area, known as a preferred retinal locus (Whittaker, Budd, & Cummings, 1988). The larger the central scotoma, the more eccentric this locus is relative to the fovea. As we move eccentrically from the fovea, the retinal ganglion cells' receptive field increases (Christine A. Curcio et al., 1990). This can be considered the ratio of photoreceptors to retinal ganglion cells, which defines the highest spatial frequencies perceivable, or Nyquist frequency (A. B. Watson, 2016).

Face perception loss is not limited to conditions typically associated with central vision loss. Patients with glaucoma have been found to have difficulties with face perception (Glen, Crabb, Smith, Burton, & Garway-Heath, 2012). Conditions related to peripheral loss, such as retinitis pigmentosa, while initially may not have foveal involvement in the early stages, can show central loss later in disease (Grover, Fishman, & Brown, 1998). Patients can be left with isolated islands of sight outside the fovea or with minimal fields of view, both leading to severe difficulties with face perception.

2.2.3 Behavioural patterns with the loss of face perception

Viewing strategies change with the progression of a central scotoma. Individuals rely more on the external features of a face rather than internal in recognition (Berkeley, Joshua, Peace, & Iii, 2012; Seiple, Rosen, & Garcia, 2013). This is most likely due to the lower spatial frequency nature of the external facial features.

The loss of face perception has a significant impact on an individual's quality of life. Most of the evidence for this is derived from studies considering vision loss as a whole (Taylor et al.,

2016). Recent work by Lane et al. (Lane et al., 2018) has focused on the specific impact of face perception. After focused face-to-face interviews, Lane et al. concluded the significant social impact of face perception loss. Their participants reported numerous difficulties, including; failing to recognise people, missing social nuances and concern about offending friends by unintentionally ignoring them.

2.3 How is Face Perception Assessed?

Low vision clinics do not measure face perception. It is not reported in the patient's notes in routine ophthalmic care. It might be recorded as part of behavioural analysis in a psychological assessment, but no quantification occurs. The only other clinically relevant speciality where a quantitative measurement of face perception might be performed is when a diagnosis of prosopagnosia has been made. Any measurement would usually be in collaboration with academic research teams. Thus, methods for measuring face perception are found predominantly in the research domain.

The field research into face perception is extensive, and covering all of the various methodologies is not within this thesis's scope. This section focuses on the approaches relevant to this chapter, i.e. face perception and measurement with vision loss.

The area of face recognition and vision loss is well studied, with much of the work focusing on AMD. This section reviews the most relevant literature on assessing face perception in the context of vision loss.

Owsley et al. (Owsley & Sekuler, 1981) investigated how the loss of contrast sensitivity due to ageing impacts face perception. Fourteen younger and 13 older normally sighted participants performed a face detection and discrimination task. For the detection task, the contrast of an image of a face was modulated, starting from zero contrast until the image of a face was perceived. The outcome measure was the contrast at which a face was detected. Their discrimination task used a similar contrast modulating approach, where the subject identified if two adjacent face images were identical or not. Their results indicated a higher contrast was required for both detection and discrimination for the older group compared to the younger.

Lott et al. (Lott et al., 2005) assessed face recognition in 572 older participants (mean age 78.1). Face recognition was measured by presenting standardised faces of varying sizes to simulate normal-sized faces at different viewing distances. Subjects were asked to identify the person's name and their facial expression, based on work by Bullimore et al. (Bullimore, Bailey, & Wacker, 1991). The marked decline in face recognition ability in elders is related to reductions in spatial vision and cognitive status.

Alexander et al. (Alexander, 1988) assessed facial expression identification as part of a more extensive visual function study in 100 patients with AMD. Subjects were asked to identify a facial expression from four hand-held black-and-white photographs held at a distance to replicate a person sitting next to them. The outcome measure was the number of correct responses. Only contrast sensitivity was found to correlate with performance.

Bullimore et al. (Bullimore et al., 1991) investigated face recognition in AMD patients. Subjects were presented with photographs of 8 individuals, displayed continuously, for comparison with the test photographs. The difficulty was modulated by increasing the

viewing distance until a threshold of 50% accuracy was achieved, with viewing time-limited to 10 seconds. The outcome measure was the equivalent viewing distance (EVD), defined as the viewing distance at which a real face would subtend the same angle as the test faces at threshold. 16 AMD patients were assessed, and performance was found to correlate with word-reading acuity, not contrast sensitivity.

Barnes et al. (Barnes, De l'Aune, & Schuchard, 2011) assessed face discrimination in 34 AMD patients. Nine faces were presented on a large screen (66 inches) containing eight different identities. The subject was directed to locate the target (central) face of 8 choices and report the facial expression. Errors and reaction times were reported. They found that performance accuracy and reaction times depended on both visual acuity and contrast sensitivity.

Boucart (Boucart et al., 2008) studied the recognition of facial emotions in low vision. They assessed 17 AMD patients on their ability to detect and identify emotion in grayscale photographs. They used an outcome measure derived from the combination of correct responses and false alarms. They reported that the AMD patients performed less well at detection and even less well at recognition when compared to normally sighted controls. They also observed that AMD participants mostly identified emotions using the lower part of the face (mouth).

Johnson et al. (Johnson, Woods-Fry, & Wittich, 2017) investigated the impact of magnification on face emotion perception with 20 AMD patients. Their task was first to detect and then recognise emotion in photographs of faces. Magnification of the stimulus was modulated during the test, and the outcome measure was percentage correct. They concluded that while magnification did impact emotion detection, it is not the only solution to improve emotion perception.

McCulloch et al. (McCulloch et al., 2011) used a face discrimination task with synthetic face stimuli. The synthetic faces were derived from photographs of faces, then morphed by varying percentages, allowing for more precise quantification of each task's difficulty. They simulated reduced contrast sensitivity and visual acuity with Bangerter foils. They found that face discrimination distance was correlated with both acuity and contrast sensitivity when manipulating face size.

Taylor et al. (Taylor, Smith, Binns, & Crabb, 2018) used a modified version of the Cambridge face memory test with 30 AMD patients categorised into early, intermediate and late-stage disease. The outcome measure was percentage correct, finding that face recognition was significantly impaired in the late-stage AMD group and correlated with contrast sensitivity.

In the aforementioned studies, there are many testing methodologies, with most studies developing their own test, with little or no discussion about the inherent variability of their methods. Only Taylor et al. (Taylor, Smith, et al., 2018) used an established face perception test. The sample sizes are generally small; Lott et al. (Lott et al., 2005) $n=527$, and Alexander (Alexander, 1988) $n=100$, the two exceptions. However, Lott only studied the elderly with no patient group. Alexander had a weak experimental design, using a limited number of stimuli (four) and little description of their testing protocol. Access and testing large numbers of patients is appreciably challenging, and most studies noted this limitation. As a result of the heterogeneity in methodologies and sample sizes, there is inconsistency in the reports on which factors relate to face perception performance.

The conclusion is that these studies all show a lack of commonality with testing methodologies, likely due to the lack of available standardised methods.

2.4 Image Enhancement

As previously discussed, there is evidence that any form of visual enhancement should treat faces as a unique case. Furthermore, any enhancement should be careful not to over or under-enhance specific facial features.

2.4.1 Previous work

There have been a few approaches to improving face recognition for central vision loss (other than treatment for the underlying condition). The simplest method to enhance face perception is magnification; in other words, increasing the face's size. Terjeria et al. (Tejeria, Harper, Artes, & Dickinson, 2002), using a biopic device (a small telescope attached to a pair of spectacles), showed an improvement in face recognition in 86% of participants and 79% regarding face emotion discrimination. Bullimore (Bullimore et al., 1991) modulated the viewing distance in their face perception task and found that the lower the performance, the further away was the face. Jonson et al. (Johnson et al., 2017) investigated magnification on emotion detection and found that magnification had a positive effect on performance for small image sizes, but not at larger sizes, concluding that image magnification is not a complete solution on its own.

Image processing can amplify spatial frequencies and potentially improve a patient's ability to recognise faces. There is a bulk of evidence that specific bands of spatial frequencies are essential for facial recognition (Costen, Parker, & Craw, 1996; Näsänen, 1999; Ojanpää & Näsänen, 2003; Eli Peli et al., 1994). These studies suggest that the most influential bands are from 8 to 16 cycles per face. This concept was the basis for a method proposed by E. Peli

et al. (E Peli & Peli, 1984) which selectively amplifies these spatial frequencies using an algorithm developed by T. Peli and J. Lim (T. Peli & Lim, 1982) (see figure 2.1). Using this methodology, E. Peli et al. found an improvement in face recognition for 16 of 38 central vision loss patients.

Leat et al. (Leat, Omoruyi, Kennedy, & Jernigan, 2005) investigated several different filters for generic image enhancement, including faces. They found that subjects preferred filters with lower gains, and Unsharp masking gave the most perceived visibility. (Unsharp masking is a process that generates a sharp image by subtracting an image from a blurred version of itself). In another study, Leat and Mei (Leat & Mei, 2009) investigated the effect of customising these filters based on a subject's contrast sensitivity, again using subjective responses. They found that the Peli adaptive filter (T. Peli & Lim, 1982) was the most frequently preferred and concluded that customising filters was did not give enough advantage over generic implementations. In a preliminary study, leat and Mei (Mei & Leat, 2009) performed objective assessments of subjectively chosen filters on face images. They asked nine maculopathy subjects to identify facial expressions and found a measurable improvement.

Several novel edge enhancement methods were introduced by Al-Atabany et al. (Al-Atabany, Memon, Downes, & Degenaar, 2010) that simplified the scene using anisotropic filters, colour quantisation and edge detection. They introduced a retinal degeneration model for simulating vision loss, which could be used to assess image enhancements. They also used a subjective comparison method with visually impaired patients and found a higher preference for the image enhancements when VA was greater than 0.9.

These studies show potential for image enhancements; however, they are limited in their methodologies. The experimental paradigms were poorly validated, Peli et al. (E Peli & Peli, 1984) used a famous faces recognition task that is prone to memory influences; Leat primarily used a subjective measure (Leat & Mei, 2009; Leat et al., 2005), and only used a quantitative measurement in a small preliminary study (Mei & Leat, 2009).

A different approach is caricaturing, used by Lane et al. (Irons, Barnes, Provis, & Kwa, 2014)(Lane et al., 2019). This has shown promise in simulated vision loss and with AMD patients. While novel, this approach does come with computational complexity, and a real-world implementation has not been presented. A system would have to reliably track all facial features and perform the caricaturing on the fly in real-time. As the tools exist, this is feasible, but producing a reliable, robust system is far from trivial.

2.4.2 Cartoonisation

A new method has been proposed called Cartoonisation based on work by Robson and Pearson (D. E. Pearson & Robinson, 1985); this method attempts to reduce the information in a face whilst maintaining the important features (figure 2.1).



Figure 2-1 Image enhancements (left) Original face, (centre) Cartoonisation, (right) Adaptive filter

Robson & Pearson developed an efficient method to communicate via video, where data communication rates are limited. Today's internet communications do not have this limitation, and full video communication is commonplace. They postulated that perceptually significant facial features occur when surfaces are tangential to the camera's line of sight. They developed a line drawing algorithm that, combined with a binary thresholded 'Mass' image, produced cartoon-like face images. This was further validated by Pearson et al. (D. Pearson, Hanna, & Martinez, 1990) by comparing the method against a human artist, finding a degree of correspondence.

Bruce et al. (Bruce, Hanna, Dench, Healey, & Burton, 1992) investigated the role of the "Mass" component in a famous person face recognition test. They found that the individual parts of the cartoon, lines and mass, did not perform as well as the full cartoon, which matched the source photographs' performance. A similar study by Gooch et al. (Gooch, Reinhard, & Gooch, 2004), who developed a similar concept to generate cartoon images, concluded that the cartoons were as effective as photographs in recognition tasks. Interestingly they reported that for learning tasks, the cartoons were learnt at a greater rate than that of photographs.

This Cartoonisation approach was implemented in the wearable display by myself (I. R. Wilson et al., 2016), but as it was targeted as a mobility aid, the impact was not fully realised. The high contrast provided by the cartoon images and the simplicity of the method can perform well as an enhancement method. This method is considerably more straightforward than the Al-Atabany (Al-Atabany et al., 2010) methods, which rely on anisotropic filters that are computational complex.

2.5 Face Perception Tests

As shown in the above short review, there is a significant variation in methods used to measure face perception. This lists those only within the context of visual impairments; the entire field of face perception would contain far too many to list.

The type of test used can be categorised into a few distinct relevant categories: Detection: can a participant tell if there is a face present or not. Recognition: can a participant identify the face. Discrimination: can a person identify the difference (or lack of) between a number of faces.

The most robust work in quantifying face perception has come from the diagnosis of prosopagnosia, also known as face blindness. The diagnosis and quantification of the condition have driven the creation of several tests; the most well-known are the Cambridge Face Memory Test (Brad Duchaine & Nakayama, 2006) and the Cambridge Face Perception Test (Bradley Duchaine, Germine, & Nakayama, 2007). The Cambridge face memory test, as its name suggests, has a memory component to it, and this can potentially compromise the effect being studied. The Cambridge Face Perception test is an alternative test in which participants are asked to order six face photographs in terms of their similarity to a target face.

Logan et al. (Logan, Wilkinson, Wilson, Gordon, & Loffler, 2016) developed the Caledonian face test extending the work by McCulloch et al. (McCulloch et al., 2011). This test is a face discrimination task, where participants are presented with four faces, three are identical (distractors), and one is different (target). The faces are synthetically generated from photographs of real faces (H. R. Wilson, Loffler, & Wilkinson, 2002a) and morphed toward or away from the statistical average. The use of synthetic faces was intended to control the

potential variation of photographic stimulus and allow for a more straightforward measure of difficulty.

2.6 Conclusion

One of the aims of this thesis is to assess a method of image enhancement and how it impacts face perception with the visually impaired. Many different studies exist with varying methodologies that assess face perception. Most of these have limitations and are unsuitable mainly because of poorly controlled stimulus (Barnes, De L'Aune, & Schuchard, 2011; Leat et al., 2005), memory components (Irons et al., 2014; Lott et al., 2005) or limited sensitivity range. The most well-known is the Cambridge face perception test (Bowles et al., 2009). However, this test has a floor effect, meaning that the minimum score is above what most visually impaired participants would realistically achieve and is unable to measure poorly performing participants. Logan et al. developed a new face discrimination task called the Caledonian Face Test (Logan et al., 2016). This attempts to address all the concerns of previous face perception tests. The limitation is the stimulus; Logan used synthetic faces to which the proposed image enhancements cannot be applied.

The following chapter presents the development and validation of a new face perception test designed to avoid the pitfalls of these previous testing approaches.

Chapter 3:

Face Perception Test

My contributions to this chapter: I designed and wrote the python software for the Face Perception test. I processed and curated all of the image stimuli. I designed, obtained ethics and conducted all experimentation. I performed all data processing and statistical analysis using Python and R.

Other contributions: Professor Robert MacLaren and Professor Susan Downes assisted with proofreading.

Timeline: All data collection was performed between March 2017 and November 2018. The initial test design and software development took 4 months, with the data analysis taking a further two months.

3 Face perception test

As described in the previous section, the goal is to assess image enhancements techniques for improving face perception. There is no currently applicable face perception test. This chapter describes the development and validation of a new face perception test. For the purposes of this thesis, I have called this Oxford Enhanced Vision Assessment Tool (OVAT).

The aim is to quantify the performance of an image enhancement approach for face perception, and a fit for purpose assessment tool is required.

The assessment tool must meet specific criteria:

- 1) The assessment tool range must include those with varying degrees of vision loss to avoid floor and ceiling effects.
- 2) The stimulus must be able to have different image processing applied.
- 3) There must be no memory component, thus removing any chance of memory loss compounding task performance.
- 4) It must be quick to perform as it will be used in a clinical setting.

3.1 A new approach to an existing test

The test that best fits the required criteria is the Caledonian face test (Logan et al., 2016).

The limiting factor is the stimulus used. The test uses synthetic face stimuli (H. R. Wilson, Loffler, & Wilkinson, 2002b) which have been shown to perform well compared to real face stimuli and have been validated already with an fMRI study (Loffler, Yourganov, Wilkinson, & Wilson, 2005). However, they are not suitable for the application of different image enhancements.

The primary reason given by Logan et al. against using real photographs was the difficulty in the generation and for controlling for all the subtle features. Although the finer details are not very important with face perceptions (Abudarham, Shkiller, & Yovel, 2019).

The Caledonian test is a 4-alternate forced-choice test (4-AFC). Participants are presented simultaneously with four faces, three identical distractors and one target, which has been morphed away from the distractors by a variable amount. The participant has to identify the odd face out of the 4, and difficulty levels are varied with an adaptive staircase algorithm. For the work reported in this project, I have developed a new test based on the Caledonian Face test's design principles, but where the stimulus has been replaced with more realistic face images.

3.2 Realistic Face Stimulus Generation

Logan et al. (Logan et al., 2016) generated their synthetic stimulus from photographs of real faces. Thirty-seven feature point locations were chosen on the photograph to define the synthetic face's head shape, hairline, and internal feature locations. A mean of all the faces (mean face) was created by averaging the 37 feature points across all the faces, of the same sex, in their dataset.

The target stimulus was then generated by morphing to differ from the distractor by a certain percentage. The mean face appeared in every trial and was randomly assigned to be either distractor or target; the morphed identity was chosen randomly from a collection of 40 male, 40 female faces.

Following this approach, a photographic face stimulus was created whilst maintaining the same difficulty measure, percentage morph, as Logan et al. (Logan et al., 2016). This was performed by morphing photographs of real faces. The face photographs' source was a dataset from Stirling university [Stirling/ESRC 3D Face Database: pics.stir.ac.uk] containing 68 frontal images faces [32 male] with labelled feature points.

Feature point labels accompany each face within the Stirling dataset, 179 in total, labelled with Psychomorph (Tiddeman, Stirrat, & Perrett, 2005). Male and female mean faces were generated from the dataset, where the mean face is the statistical average of all the face feature points and texture. The stimulus was generated by morphing mean faces towards or away from an identity, with varying percentages see Figure 3.1.



Figure 3-1 morphing a photograph using Psychomorph. (Top left) Original image with green points denoting the facial feature points. (Top right) Mean face texture and green points denote the mean feature location. (Bottom left) Mean face morphed using the original image features in the positive direction. (Bottom right) mean face morphed in the negative direction using the original feature points.

The blue lines indicate the distance to a reference point on the mean face.

The feature point calculations are defined as:

(Eq 2.2.1)

$$p_{new} = p_{org} + m(p_{dst} - p_{src})$$

Where p is a face feature point, the subscript indicates the point image source, and m is the morph percentage.

The mean face was chosen over the original to avoid unquantifiable cues. There is a significant variation in skin tone, skin texture and eye colour within the dataset, which could guide the decision-making process and is not easily quantifiable.

Differences are quantified by the base 10 logarithm of the sum of the Euclidean distance between two morphed faces, which contains only structural information.

(Eq 2.2.2)

$$d = \log_{10} \sum_{i=1}^n \sqrt{(p_i - \hat{p}_i)^2}$$

The use of the mean face avoids the extra cues; skin texture, tone etc. The mean face was also more robust to artefacts introduced by the morphing process. Some image artefacts did occur at the higher morph percentages and were not included in the dataset.

The faces were then masked to display only the face, removing the background. The ears were removed in the masking process to produce stimulus closer to that of the Caledonian Face test, and large amounts of image artefacts were introduced. Finally, each processed face had the image processing applied (Figure 3.2).



Figure 3-2 processed and cropped images, left (original), centre (cartoon), right (adaptive filter).

3.3 Experimental Paradigm

The following definitions have been used:

Trial – each event or presentation of 4 faces

Test – A collection of trials.

Experiment – A collection of tests.

The proposed test was implemented in OpenSesame (Matht, Schreij, and Theeuwes 2012), an open-source experiment builder, with stimulus generation and calculations written in python.

Fifty identities (25 female) were selected from the Stirling database. For each identity, morphed faces were generated, ranging from -100% and +100% at 5% increments, resulting in 40 individual faces per identity.

Each test consisted of 50 individual trials. Within each trial, four faces, three identical distractors and one target are presented simultaneously on a grey background (Figure 3.3). For each trial, a single identity is chosen from the set of 50 different identities; the target and distractors differed by varying the morph towards or away from the mean. The location of the target was randomly assigned. Participants were instructed to identify the odd one out, responding via a button box, and were free to take as long as they desired to make their decision.

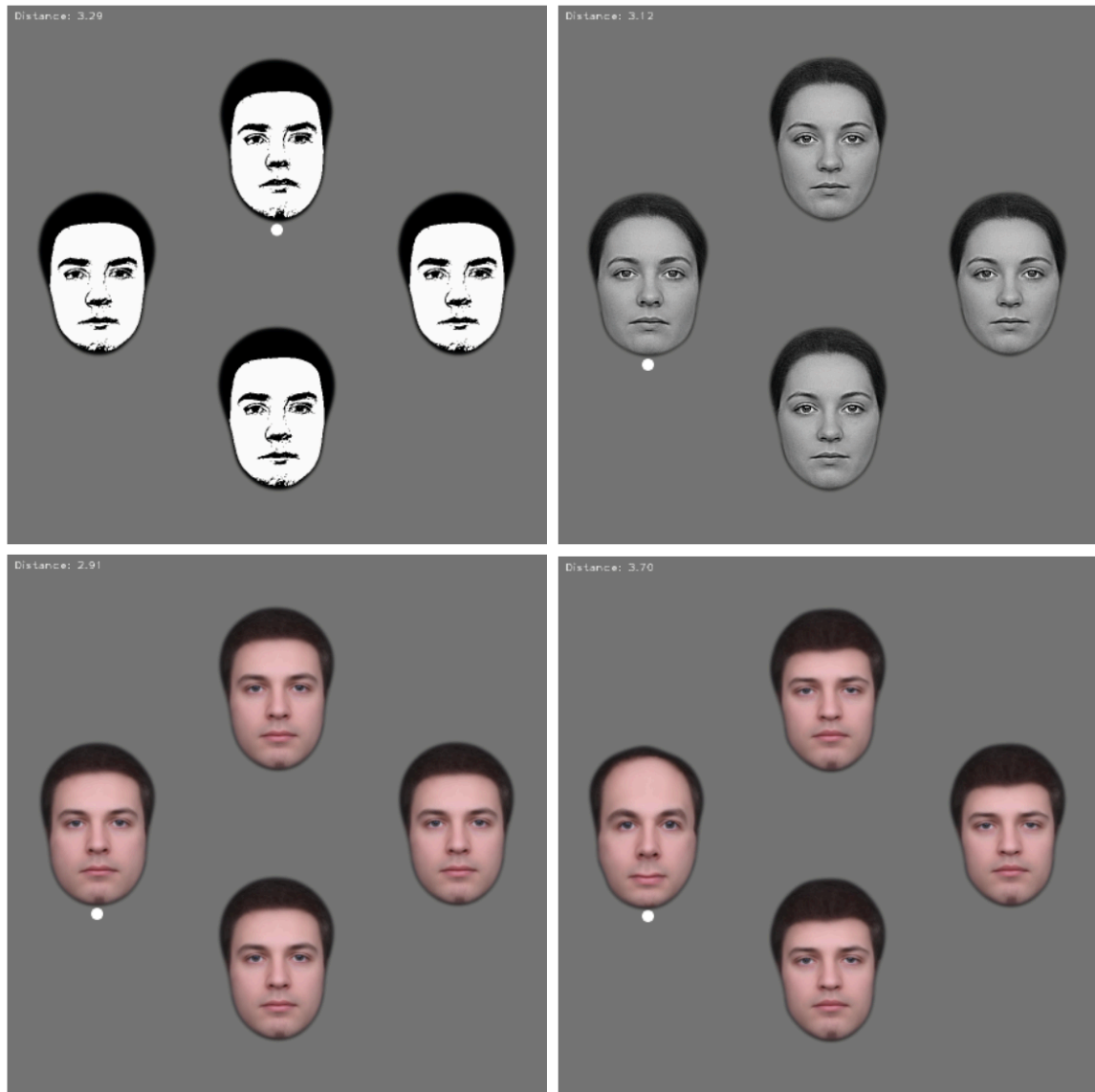


Figure 3-3 (top left) Cartoonisation, (top-right) Peli, (bottom-left) Normal – low difficulty, (bottom-right) Normal – high difficulty. Note: a white dot located below the target (the odd-one-out), and the distance value is $\log_{10}$ Euclidean distance between the target and distractor; distance is inversely related to difficulty. The dot and scores are for illustration purposes only and were not present during the experiment.

The difficulty was varied using the QUEST (a B. Watson & Pelli, 1983) adaptive staircase strategy. QUEST is a Bayesian prediction technique, which attempts to fit a psychometric function to collected data.

The strategy adapts the difficulty of the test such that the proportion of responses meets some criterion, in this case, 62.5% correct (Kingdom & Prins, 2010) chosen by the function:

$$t = \gamma + (1 - \gamma)/2$$

Where γ is the guessing rate, for a 4-AFC, it is 0.25. Upon test completion, the resulting difficulty (or distance value) is the measured threshold and the score assigned to that participant.

3.4 Test Validation

To confidently use this new test to measure face perception, I must first validate the test.

This section aims to understand the repeatability of intra-subject scores and establish performance with a normally sighted population.

3.4.1 Participants

The proposed face perception test was assessed on 49 normally sighted participants {F=35 (71%), age mean=25.1, SD=7.0} recruited as part of an ethically approved trial [R50271/RE001]. Participants were asked to perform the test twice, separated by a 5-minute break.

All participants viewed the test on a large display, seated 5 metres away. The size of the faces was equivalent to viewing the average person at 2 meters. All participants were given the same verbal instructions; “Be fast but accurate”.

3.4.2 Results

Scores are reported as the final threshold from the QUEST algorithm, as the distance value defined in (Eq 2.2.2).

The test scores were found to be not normally distributed [Shapiro-Wilk test for normality, $p=0.03$]. With a median = 2.96 IQR(0.21). [range 2.73 -> 3.4] (Figure 3-4)

There is no evidence for a difference between the first test (Median=2.96, IQR=0.22) and the second test (M=2.94, SD=0.21); [Wilcoxon rank-sum test, $W=1295$, $p=0.51$].

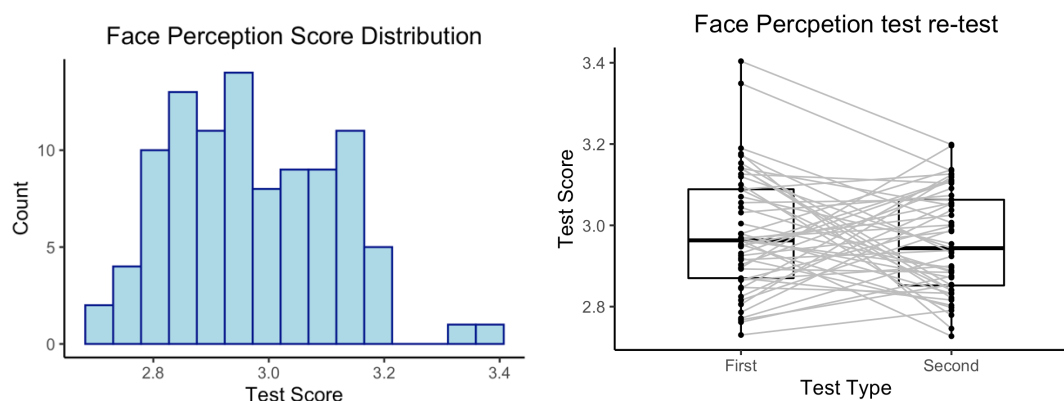


Figure 3-4 (Left) test score distribution for all participants, (right) test scores separated into the first and second tests.

3.4.2.1 Test Duration

The test duration is the time a participant took to complete an individual test; in this experiment, participants performed two tests.

Median duration for all tests 7.5, IQR 5.8] [$5^{\text{th}}=3.57$, $95^{\text{th}}=17.43$] (minutes).

Comparing the two tests performed by each participant, first and second. The medians were found to be not significantly different [Wilcoxon rank-sum test, $W=1243$, $p=0.77$] (figure 3-5). Test duration negatively correlated with test score [Spearman's rank correlation $\rho=-0.50$, $p<0.0001$], participants scored worse the faster they performed the test.

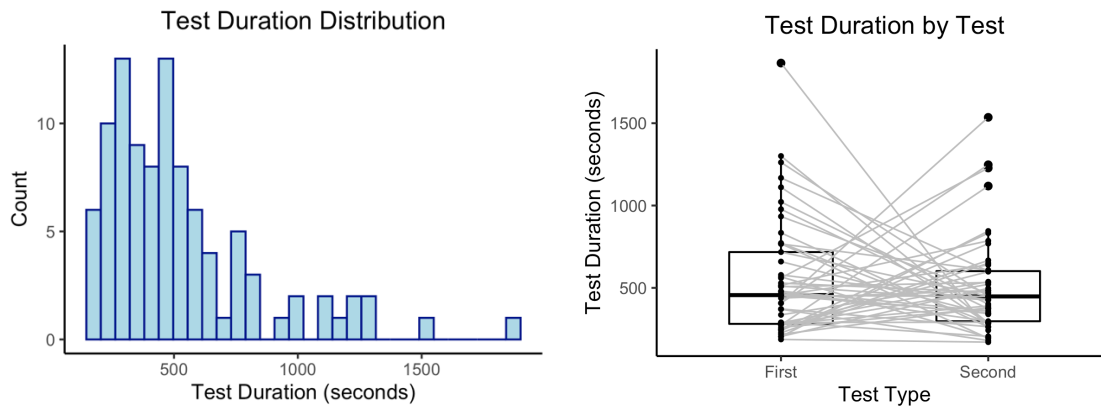


Figure 3-5 (Left) test duration distribution for all participants, (right) test duration separated into the first and second tests.

3.4.2.2 Response Time

Response time is defined as the time a participant took to respond to an individual trial.

Median response time was 7.3 IQR=8.4 [5th=2.34, 95th=31.4] (seconds).

There was a significant difference between the response times for a correct (median 6.3ms, IQR=7.3ms) and incorrect response (median=6.3ms, IQR=10.1ms) [Wilcoxon rank-sum test, $W=3559486$, $p<.0001$] (figure 3-6). Response time negatively correlated with trial difficulty [Spearman's rank correlation $\rho=-0.51$, $p<.0001$], the harder the trial, the longer a participant took to respond.

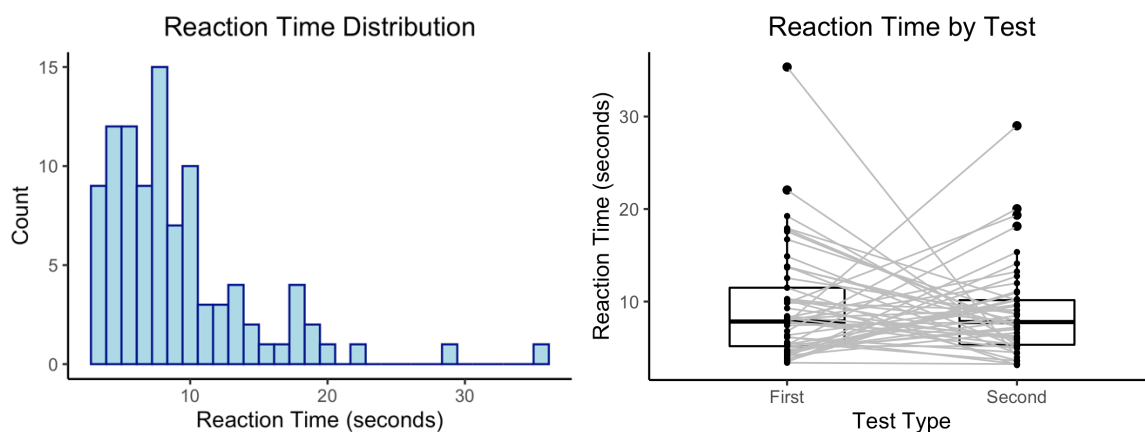


Figure 3-6 (Left) test reaction time distribution for all participants, (right) test reaction time separated into the first and second tests.

3.4.2.3 Orientation Bias

Within the experimental paradigm, the target is randomly assigned to a location; top, bottom, left or right. A small number of participants noted that they had a bias towards one of the locations.

The display used for the experiment was not calibrated, meaning that the distribution of pixel brightness is unknown. An artefact of LCD technology is that viewing angles can affect visual appearance. Therefore, I investigated if there was any bias with the location of the stimulus.

Table 3-1 Stimulus target locations

	Top	Bottom	Left	Right
<i>Target location</i>	1217 (24.8%)	1217 (24.8%)	1215 (24.8%)	1255 (25.6%)
<i>Responses</i>	1258 (25.7%)	1367 (27.9%)	1162 (23.7%)	1113 (22.7%)
<i>Correct</i>	744 (24.6%)	805 (26.6%)	749 (24.7%)	732(24.2%)
<i>Incorrect</i>	469(25.1%)	412(22.0%)	466(24.9%)	523(28.0%)

Table 3-1 presents the number of trials where the target was at one of the four locations.

Target location shows a uniform distribution, with a slight bias to the right – the choice was made with a random number generator in python. The total responses have a slight tendency away from the right and towards the bottom. When this is broken down into correct and incorrect responses, there is a slight bias or preference for the bottom location. i.e. Subjects were more often correct at this lower location. A potential reason is due to the configuration of the room and viewing angle; the participant was slightly closer to the bottom part of the screen. However, an analysis of variance of the intensities presented at

each location where a correct response was made showed no significant difference [f=0.066, $p=0.978$], and the differences are most likely noise.

Overall, these variances are small and do not contribute to any significant bias indicating that the experimental setup is reliable.

3.4.2.4 Repeatability

Repeatability was measured as the difference between the scores for the first and second tests and reported with the Bland-Altman analyses of retest repeatability (CoR_{95}), calculated as $1.96 \times$ the standard deviation (σ) of the difference between the two test scores. (figure 3.7)

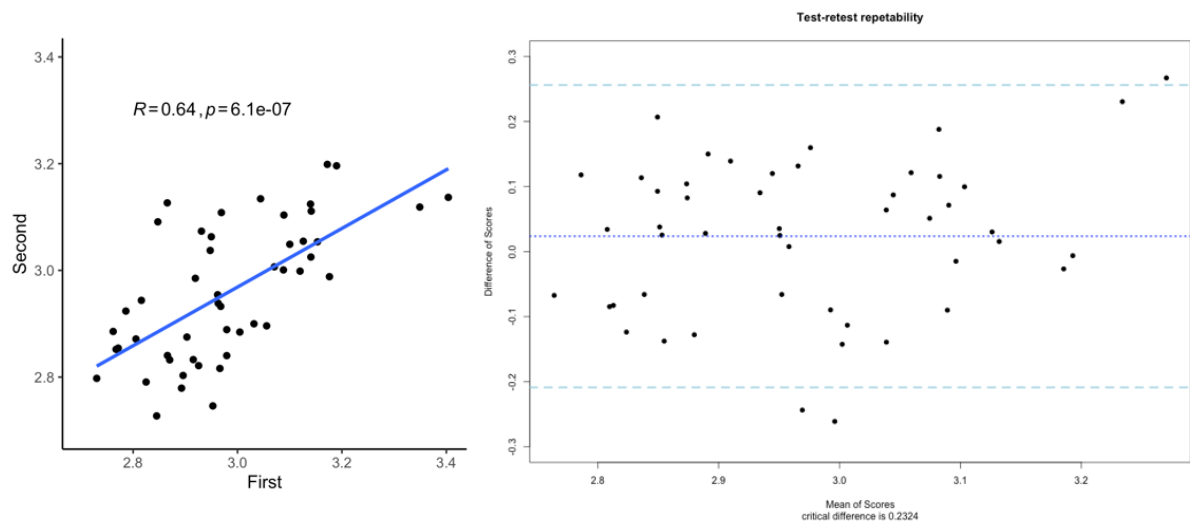


Figure 3-7 (Top) comparison of the two test scores, line of best fit presented and Pearson's product-moment correlation. (Bottom) Bland-Altman plot, dashed light-blue lines indicate Coefficient of Repeatability 95th percentile.

The duration of the test was set to be larger than potentially needed, 50 trials. Using CoR_{95} as a measure, I can potentially reduce the number of trials within the test to as low as 40 (figure 3.8).

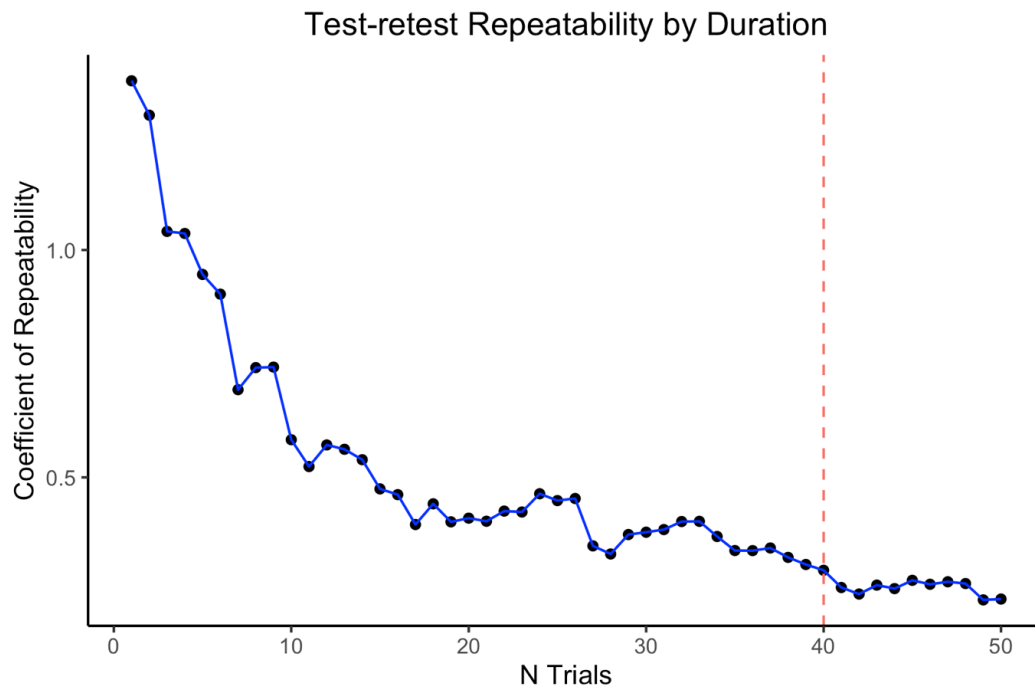


Figure 3-8 Plot shows how CoR95 changes with task length. The final value after 50 trials was 0.23. The value starts to plateau after 40 trials [CoR95=0.25].

3.4.2.5 Learning effect

I have defined metrics for repeatability, where scores after 50 trials can vary by 0.23. This is higher than that reported by Logan et al. This variability is further investigated by identifying any learning effects. i.e. do the participants get better at the test over time?

A further 18 participants performed the test 3 times, with 5-minute breaks between the tests.

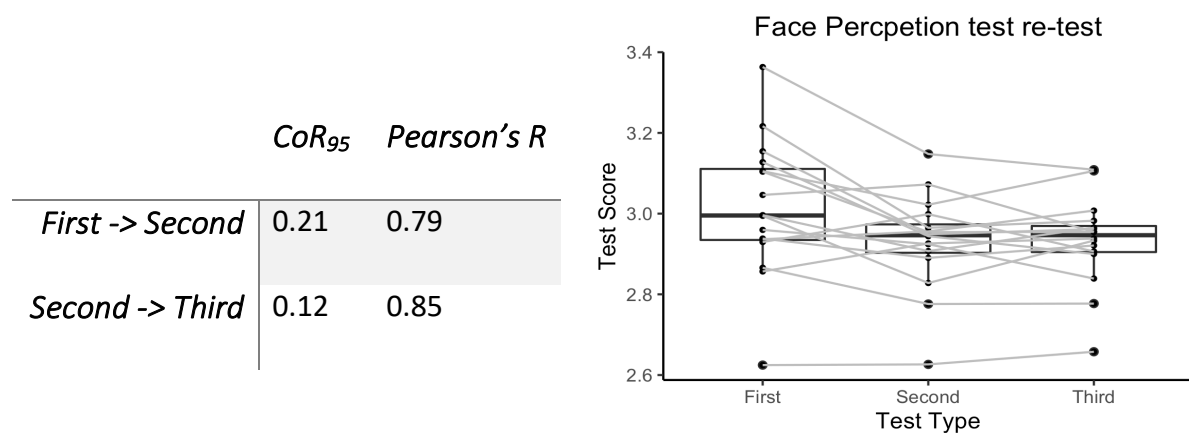


Figure 3-9 (left) Table presents the CoR95 and correlations between the tests. There is a notable reduction in variability from the first->second to the second->third. (right) Score box plots for each test.

These results show strong evidence for a learning effect, with the variability reducing once the participants had become familiar with the test (figure 3.9). Any implementation of the test must contain a significant training period for familiarisation.

3.4.2.6 *Stimulus effects*

It has been assumed that each face identity used for the stimulus generation produces the same difficulty when a similar amount of morphing is applied. i.e. the same difficulty value corresponds to a perceptually similar difficulty. Thus, any variability with this difficulty measure is a potential source of noise.

I investigated this potential noise source by collating all of the individual trial results and grouped them by stimulus rather than by participant. The idea is that each stimulus has been seen at least n times (where n is the number of participants) at various difficulty levels; therefore, I can calculate a population sensitivity threshold for each stimulus. The ideal data set would have an identical population threshold for all.

Figure 3.10 presents example trials from the two lowest {M1041 = 2.35, M1034 = 2.43} therefore easiest, and two highest scores {M1014 = 3.466, M1036 = 3.416}. All the trials are presented at similar difficulties (2.9 ~ the population median). While this is an objective measure, it may be apparent to the reader that the top two trials show more obvious differences between the target and distractors than the bottom two.

While the quantitative measurement shows significant variability (figure 3.10), it is not

possible to confidently isolate the source to inaccuracies within the difficulty metric alone. However, the qualitative comparison does indicate that the difficulty metric is undoubtedly a contributor to the variability.

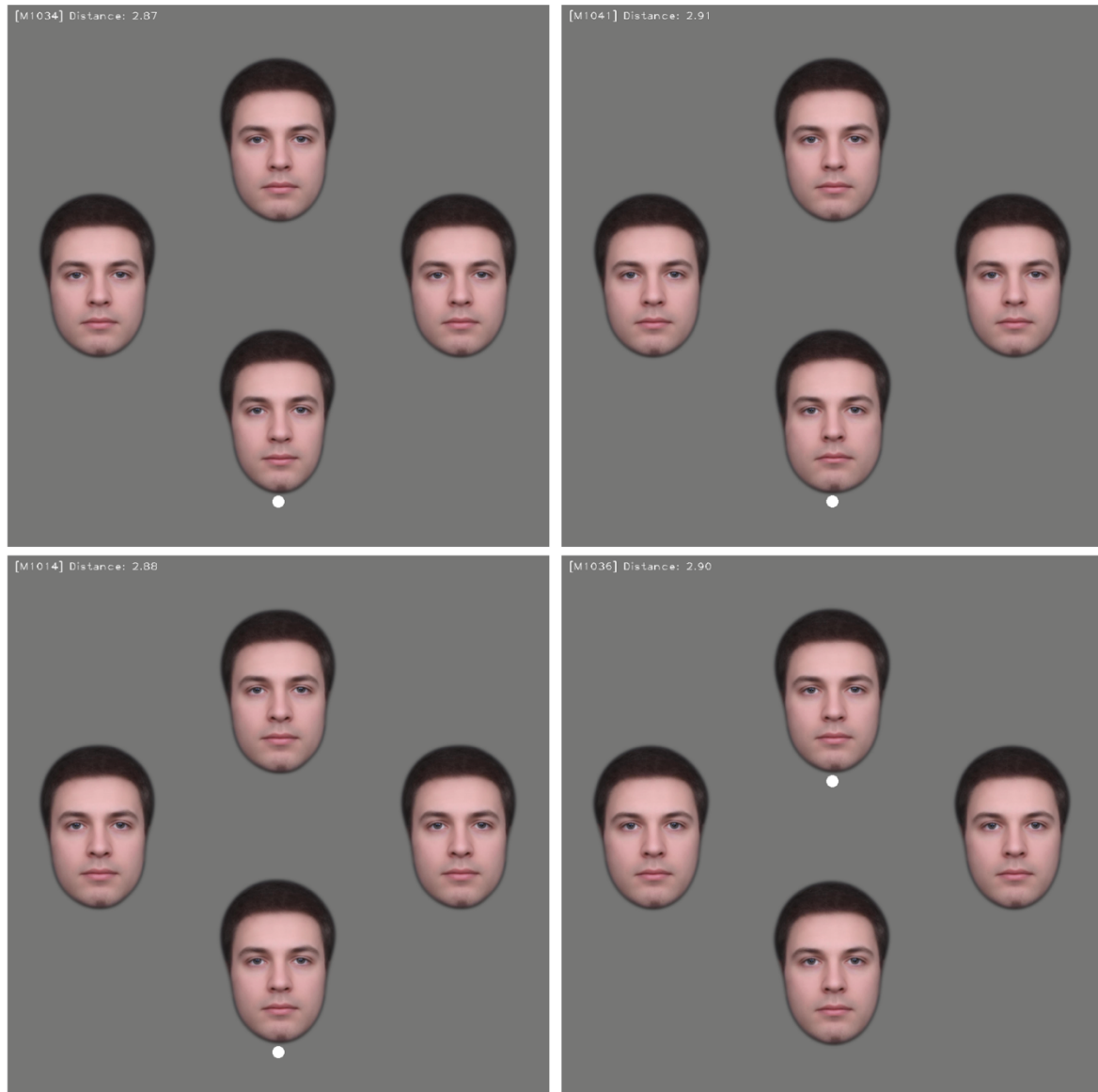


Figure 3-10 Four examples of the stimulus with similar difficulty, according to the distance metric. The top two (M1034, M1041) are potentially easier than the Bottom two (M1014, M1036). The white dot indicates the correct choice (target) for each trial.

3.4.3 Measuring difficulty

The difficulty measure used for the face discrimination task is the Euclidean distance of the face feature points. This was chosen as it replicates the method used by Logan et al.

However, I am using photographs that are more visually complex stimuli. The different face stimuli could explain the higher test-retest variability seen with this version of the test compared to the Caledonian face Test. This section describes an alternate method for determining the test difficulty measures.

3.4.3.1 *Computer Vision Face Recognition*

Face recognition algorithms have been an active area of research for some time and are now achieving results as accurate as expert humans (Phillips et al., 2018). The general approach is to extract facial features and aggregate them into a face feature vector or face descriptor. These face descriptors are then compared to each other using a distance measure, where this distance is proportional to face similarity. Finally, the face descriptors are grouped by the distance measure into identities.

This is a broad description and may not cover all approaches, but most algorithms fall within this framework. The most important part of this framework is the generation of face descriptors. Most current research has been on generating these face descriptors, with the best performing approaches using Convolutional Neural Networks (CNNs).

The method used with the face perception test, Euclidean distance of feature points, is a type of face descriptor. This method's disadvantages are the accuracy of the feature selection and the lack of ability to discriminate which features are more important (Jafri & Arabnia, 2009). The experiments carried out as part of this thesis have investigated whether

using a different face descriptor, with a performance more in line with the current state of the art face recognition systems, is a more accurate difficulty measure for the face discrimination task.

3.4.3.2 Face Descriptors

There have been a variety of methods for defining face descriptors, and they can be broadly categorised into the following types:

Feature-Based – these approaches extract distinct features, such as eyes, nose, and mouth, and then compute the geometric relationship between these points.

Holistic - these approaches use the entire image to define their descriptors rather than local features. Employing a variety of statistical methods, such as Principle Component Analysis (PCA), robust descriptors can be generated (Sirovich & Kirby, 1987; Zafaruddin & Fadewar, 2018). More recent developments using Convolutional Neural Networks (CNN) have extended the holistic approaches' performance.

3.4.3.2.1 FaceNet

One of the best-performing face recognition approaches is a method called “FaceNet” by Philbin et al. (Schroff, Kalenichenko, & Philbin, 2015). This method's core feature is the generation of a 128-bit face feature vector referred to as an “embedding”, where face similarity is proportional to the squared Euclidean distance between two embeddings. I have used their terminology for this thesis. Philbin et al. approach was to train a CNN to generate the embedding per image, using a triplet loss function. The training method loads triplets of pre-labelled faces, two of one identity (positive samples) and one of a different identity (negative sample), and the triplet function attempts to reduce the distance

between the two positives and increase the distance to the negative samples. This training step was repeated a large number of times >1million. The result of this is a classifier that generates a face feature descriptor that has a small squared Euclidean distance for the same identity, regardless of pose, scale or lighting.

The recognition step is straightforward, the generated feature vectors are compared to known identities, and the lowest value indicates the highest similarity and likelihood of a match.

The OpenFace project (Amos, Ludwiczuk, & Satyanarayanan, 2016) uses this as the basis of their approach. The Source code is available, and although the training step takes a large amount of time (>24 hours) to generate, it only needs to be performed once. The inference step, generating feature vectors from an image, is quick ~seconds.

3.4.3.2.2 Distance Method Comparison [Feature Points v FaceNet]

Distances based on the FaceNet embeddings and the Euclidean method were generated for each face stimulus within each identity, which provided the full range of difficulties for each identity, enabling the comparison of the two.

Figure 3.12 presents how the two methods relate to each other. When considering all, the Spearman rank correlation is [$\rho = 0.89$, $p < 0.0001$], indicating a good relationship.

However, the relationship is not entirely linear, and it can also be observed that the embeddings distance shows a larger range for smaller distances. This observation is better demonstrated when presenting the distances for a single step; that is the smallest difference available in the dataset. The correlation between the two is significantly reduced;

Spearman's rank correlation [$\rho = 0.154$, $p < 0.0001$].

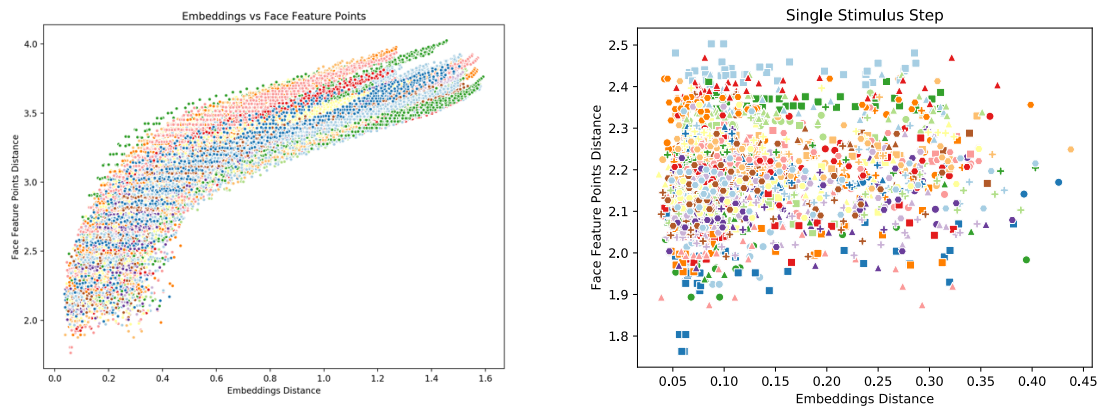


Figure 3-10 (left) Embeddings vs Face feature points distances for all faces – colours represent a stimulus identity. (right) Embeddings vs Face feature points distances for a single stimulus step, and colours represent stimulus identity.

3.4.3.2.3 FaceNet and Test-retest - Simulation

The implemented test records all of the details of every trial within its log file. This provides the ability to re-process all of the experiments with an alternate difficulty measure. FaceNet embeddings were generated for all of the stimuli, and new distances were calculated for each trial. The threshold scores were then re-calculated using the recorded responses.

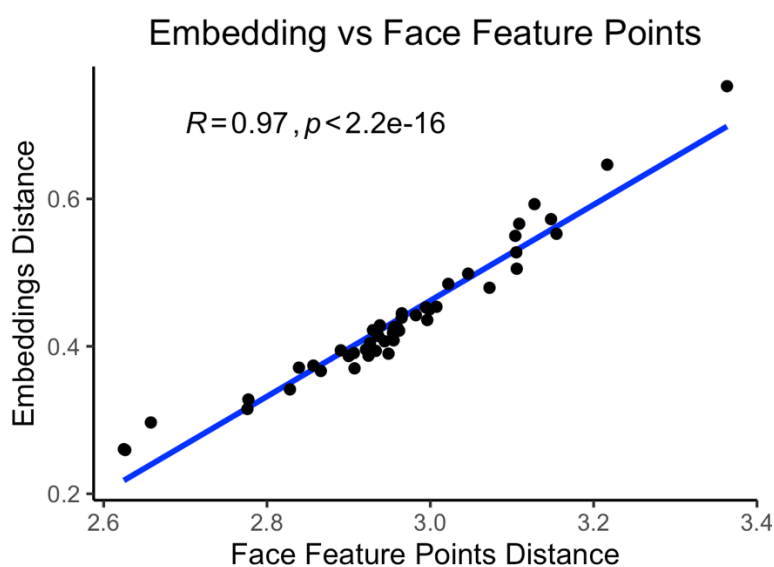


Figure 3-11 – Comparison of threshold score estimates. Pearson's correlation coefficient presented.

Figure 3.13 presents how well the two thresholds scores correlate after re-calculation, indicating that the embeddings do not impair the results.

However, when assessing the coefficient of repeatability, the two measures use different scales and are not easily comparable. To estimate the agreement, a correlation coefficient is used:

Embeddings: Pearson's $[R=0.73, p<0.0001]$

FP Euclid: Pearson's $[R=0.64, p<0.0001]$.

This stronger correlation indicates a reduction in the variability of the test when using the FaceNet embeddings distance measure.

3.4.3.3 FaceNet and Test-retest - Experiment

Using the FaceNet embeddings as a distance measure has indicated a reduction in variability; however, this has only been on re-calculated data. A further 16 participants performed the same face discrimination task, using the FaceNet generated difficulty measure. Each participant was asked to complete the test three times.

Test 1->Test 2 $CoR_{95} = 0.15$; Pearson's $[R = 0.76, p=0.0006]$

Test 2->Test 3 $CoR_{95} = 0.07$; Pearson's $[R = 0.87, p<0.0001]$

The correlation between Test 1 and Test 2 is stronger using the embedding than the Face Feature point (3.4.2.5) and similar to the simulated values (3.4.3.2.3). When comparing Tests 2 and 3, the correlations are comparable to the face feature points method.

3.5 Discussion

The aim was to develop a new face perception test that meets image enhancement assessment requirements in a visually impaired population. The test has been validated with a normally sighted population, and repeatability values ascertained.

The use of photorealistic stimuli has increased the test variability compared to the Caledonian face test by Logan et al. (Logan et al., 2016). This increase is likely due to the increased complexity of the stimuli used. For instance, the synthetic stimuli used by Logan et al. used identical internal features (eyes, nose and mouth); only the location of these features changed. While the newly generated stimulus showed more variability, this is more representative of the real world, and importantly can be used to assess image enhancement techniques.

A limitation was the inability to compare the OVAT with the original Caledonian face test. I contacted author A. Logan and asked for the version to use, but he was unable to share the test with me.

A learning effect was observed, which plateaued after the first test, reducing the test's variability. Logan et al. reported no learning effect, again, the likely reason being the more complex stimuli used here made the test harder, and participants developed more effective strategies as the test progressed. Anecdotally, participants stated that the test was much harder than expected, and they developed their own strategies, such as focusing on certain features in a particular order.

The difficulty of the test is supported by the correlation of test duration and score.

Implying attention, or lack of is important when performing the test. It should be noted that Logan found no significant differences in test scores when viewing time was restricted (apart from when viewing time was limited to 3 seconds).

The analysis of stimulus effects did indicate an influence of the variability of the stimuli on repeatability. This implies that some stimuli are easier than others, and the distance measure used is not fully representative of the task's perceptual difficulty.

A new distance measure was introduced to reduce the variability of the test and reduce it by a small amount; however, the important feature of the new distance measure is that it can easily be applied to any image of a face. A criticism of a number of the previously used Face discrimination tests (e.g. Barnes et al. (Barnes, De L'Aune, et al., 2011)) is the lack of control for the difficulty of the stimuli used. An approach is to have a large number of participants organise the faces by similarities. But this is time-consuming and must be performed every time the stimulus data set is changed. The use of the FaceNet Embeddings as a difficulty measure negates this and can be a powerful tool for future face perception tests, easily applied to any image of a face.

In the presented test, the mean face was used for all presentations; it is possible to use the original faces image with the Embeddings distance measure textures to provide an even more realistic test. This would remove the need to morph the faces, which adds complexity and can introduce image artefacts. Experimental work by Abudarhama et al. (Abudarham et

al., 2019) and their perceptually sensitive features assessed the same method used (FaceNet) and showed it is sensitive to the same features as their human participants.

The intended audience for this test is visually impaired, I am assuming that this will be translatable. However, a limitation of these results is no testing with any simulated impairment was performed. For example, blurring the stimulus to removing the higher spatial frequency components or applying an opacity filter to reduce the contrast (Thompson, Legge, Kersten, Shakespeare, & Lei, 2017). To simulate a central scotoma, one could use the results of work by A. Watson (A. B. Watson, 2014), who created a model for retinal ganglion cell receptive field density as a function of the visual field location. Watson showed how his model can be used to estimate the Nyquist frequency at any retinal location. This could be used to translate the extent, or radius, of a central scotoma to the maximum perceivable spatial frequencies. Which could then be applied as a specific level of blur to an image. While simulation is imperfect, this would enable an understanding of some of the limitations of the test when translating to a visually impaired population.

Challenges with recruitment meant that the age of the participants will be significantly different (younger) than the typical visually impaired patient population. If any age effects are identified in the visually impaired population, it will be hard to distinguish the cause from behavioural, or visual.

The next chapter presents the application of this test, with image enhancements and a visually impaired population.

Chapter 4:

Face Perception – Patients

My contributions to this chapter: I designed, wrote, and obtained ethical approval, including all required patient information and recruitment documentation. The majority of the data collection was performed by myself, with two months of support from students. I performed all data processing and statistical analysis using Python and R.

Other contributions: Caroline Justice aided with patient communication and room bookings. Professor Downes helped with patient identification. Abby Cave assisted with data collection in February 2018. Aran Johal assisted with data collection in September 2019. Professor Robert MacLaren and Professor Susan Downes assisted with proofreading.

Timeline: Data collection was performed between December 2017 and September 2019. Data analysis took three months.

4 Face Perception - Patients

The primary aim of the work described in this chapter is to understand if image enhancements impact face perception performance. The secondary goal is to identify any factors that may contribute to performance.

4.1 Materials and Methods

4.1.1 Participants

Patients with visual impairment were recruited from the Oxford Eye Hospital as part of an ethically approved trial [REC: 17/SC/0178: IRAS: 198017].

The inclusion criteria were; visual acuity of 6/24 (0.6 logMAR) or worse in the best eye and a self-reported difficulty with face recognition. Participants without conversant English or any signs of cognitive impairment were excluded.

Participants fulfilling the inclusion criteria were approached and provided with an information sheet containing information regarding the study, along with a verbal description. Willing participants were invited and recruited to attend the experimental sessions at the Oxford Eye Hospital within the John Radcliffe Hospital.

Controls

Twenty 'control' participants with normal or normal corrected vision and no history of eye problems were approached and provided with an information sheet containing information regarding the study, along with a verbal description. Willing participants were invited and recruited to attend the experimental sessions at the Oxford Eye Hospital within the John Radcliffe Hospital as part of an ethically approved study [R50271/RE004].

4.1.2 Experimental Protocol

Patient participants were invited to attend the Oxford Eye Hospital; the experimental procedure contained several discreet steps:

1. Informed consent
2. Quality of Life Questionnaires
3. Visual assessments
4. Face perception test

Control participants only performed informed consent and the Face perception test, using the same experimental setup as described in the previous chapter, while using the same stimuli and test as the patients.

4.1.2.1 *Informed consent*

All participants received an information sheet containing a study description [appendix A]. A verbal description of the study was delivered, allowing for the participant to ask any questions. Upon agreement, the participant completed a written consent.

4.1.2.2 *Questionnaires*

Questionnaires were used to obtain more information about the individual and how their visual impairment impacted their quality of life. Two questionnaires were used; Daily Living Tasks Dependent on Vision (DLTV) (Hart, Chakravarthy, Stevenson, & Jamison, 1999) and the Visual Functioning Questionnaire – 25 (VFQ-25) (Mangione et al., 2001). The questionnaires

were completed with the experimenter asking questions and the participant verbally answering.

4.1.2.3 Visual Assessments

All participants were asked to perform a battery of functional vision tests. All tests were performed binocularly, rather than treating each eye individually, which is common in clinical practice. This was to ascertain a measure of their real-world performance rather than measure any disease progression or treatment effect.

4.1.2.4 Binocular Visual Acuity

Best-corrected binocular visual acuity (BCVA) was measured using the Revised 2000 Series Early Treatment for Diabetic Retinopathy Standard (ETDRS) letter reading eye (Precision Vision, La Salle, Illinois, USA). BCVA was calculated as the lowest complete row read, subtracting 0.02 log units for every letter correctly identified beyond that point.

4.1.2.5 Reading Assessment

The Radner Reading test (Radner et al., 1998) was used as the basis for the reading assessments. The Radner test contains short standardised sentences, where a participant is asked to read aloud, binocularly with any habitual reading aid, at a fixed viewing distance. The participant is asked to keep the same reading distance, resisting the urge to bring the text closer as it decreases in size. The experimenter visually monitors the reading distance and prompts the participant if they noticeably move the text closer.

The test provides two measures, reading acuity and reading speed (words per minute).

Reading acuity is reported in terms of logRAD (Log Reading Acuity Determination). This is determined as the last sentence read, adjusted by the number of reading errors and viewing distance.

4.1.2.6 Contrast Sensitivity

Binocular contrast sensitivity was measured at five spatial frequencies, 1,2,4,8 and 10 cycles per degree. (Cambridge Research Systems, Cambridge, UK).

The Cambridge contrast sensitivity test (Cambridge Research Systems, Cambridge, UK) is a computerised test that enables the measurement of several different spatial frequencies.

Either vertical or horizontal lines, which fill the screen (figure 4.1), are presented on a calibrated display (Display++, Cambridge Research Systems), capable of a 10bit high dynamic range. The participant is asked to identify if the lines are horizontal or vertical, responding with a button box. Interleaved staircases modulate the contrast of the lines and estimate the participant's sensitivity at the five different spatial frequencies.

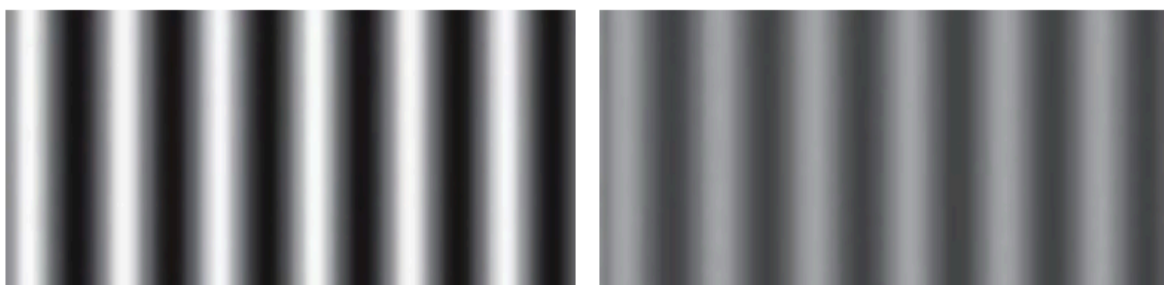


Figure 4-1 – Contrast sensitivity test example stimulus – (left) High contrast, (right) low contrast.

4.1.2.7 Colour Sensitivity

Binocular colour sensitivity was measured using a version of the Cambridge Colour Test (CCT) (Cambridge Research Systems, Cambridge, UK). The CCT is a trivector test optimised to test low vision, measuring L, M and S-cones in CIE 1976 L*u*v* colour space.

Four discs were displayed on the screen in a diamond formation. Three discs contained varying brightness of grey (distractors) and one with colour (target) (figure 4.2). Participants were asked to identify the target responding via a button box. Stimuli were presented for a maximum of 20 seconds. The colour intensities varied by decreasing the hue saturation, controlled by three interleaved staircases, one for each colour. Five negative responses aborted the test, setting the threshold at maximum saturation.

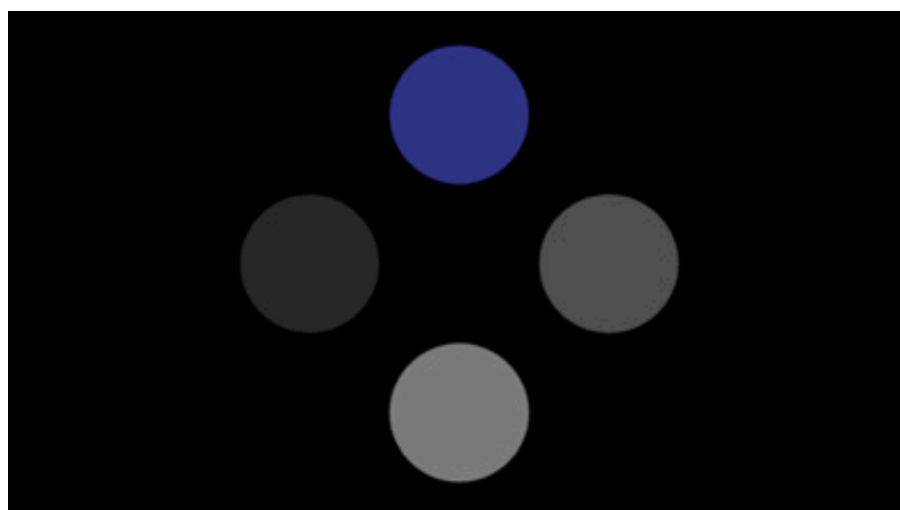


Figure 4-2 Cambridge Colour Test example stimulus. The top disc is the target, containing colour; the other three are the distractors

4.1.2.8 Face perception Test

All participants were asked to perform the face perception test (the previously described OVAT) three times. The three tests were performed with different presentations; no enhancement applied (Normal), adaptive filter (Peli) and Cartoonisation (Cartoon). Participants were given a practice round with the test with a total of 15 trials. The practice was to account for the learning effect observed with the normally sighted population. The full test of 50 trials followed the practice test. The order of the enhancements was

randomised for each participant.

All tests were presented on a Dell – 24-inch monitor at a distance of 1m. During the practice, if the participant could not see any differences in the stimulus, the viewing distance was decreased by 0.2-metre increments until they could see at least one of the stimuli correctly.

Responses were made via a custom-built button box – with large tactile buttons in the same diamond configuration as the stimuli.

The described data collection protocol amounted to a significant amount of time for the participants. The visual assessments are cognitively demanding, particularly the contrast and colour tests, which can take 15 minutes each.

This impact of the testing duration was notable for the older participants, where breaks were taken between tasks and refreshments were taken. In total, the average time was approximately 3 hours per participant.

4.2 Results

4.2.1 Control Participants

Twenty-two normally sighted participants were recruited (10 female), Mean age (SD) = 35 ± 18 , range of 22 to 77 years.

4.2.2 Patient Participants

Sixty-four patient participants were recruited; table 4.1 presents a summary. The majority of the participants had conditions typical of central vision loss, e.g. AMD, Stargardt (56%). The

patient cohort had a variety of visual disorders, where the common features comprised low visual acuity and difficulty with face perception; this heterogeneity is a representative sample of the local population.

Table 4-1 (left) summary of primary diagnosis, (right) age and visual acuity statistics.

Diagnosis	N = 64 [†]		
Condition		Characteristic	Mean (SD) [†]
AMD	20 (31)	Age, years	56 (22)
Stargardt	16 (25)	Visual Acuity, logMAR	0.87 (0.36)
Cone Dystrophy	5 (7.8)	Sex	
Macular Dystrophy	5 (7.8)	F	31 (48)
Retinitis Pigmentosa	4 (6.2)	M	33 (52)
Cone-Rod Dystrophy	2 (3.1)		
Congenital Cataracts	1 (1.6)		
CSCR	1 (1.6)		
Doyne retinal dystrophy	1 (1.6)		
Myopic Degeneration	1 (1.6)		
Nystagmus	1 (1.6)		
Optic Atrophy	1 (1.6)		
Optic Neuropathy	1 (1.6)		
Pattern Dystrophy	1 (1.6)		
Retinal Hemangioblastomas	1 (1.6)		
Retinitis Pigmentosa (Ushers)	1 (1.6)		
Sorsby Fundus Dystrophy	1 (1.6)		
Unknown	1 (1.6)		

[†] Statistics presented: n (%)

[†] Statistics presented: mean (SD); n (%)

4.2.3 Visual assessments

All patients performed the battery of visual assessments; BCVA, contrast sensitivity, and colour sensitivity test.

4.2.3.1 Visual Acuity

The binocular BCVA results are presented in Table 4.1 (right table), figure 4.3 shows the distribution, a Shapiro-Wilk normality test [$p=0.002$] indicates the data is not normally distributed – all statistical tests used will account for this non-normality. It should be noted that some participants had a better visual acuity than the recruitment criteria ($6/24 = 0.6$ logMAR). They were included in the study as they had a self-reported difficulty with face perception. For some participants, at the time of recruitment, their visual acuity was not known.

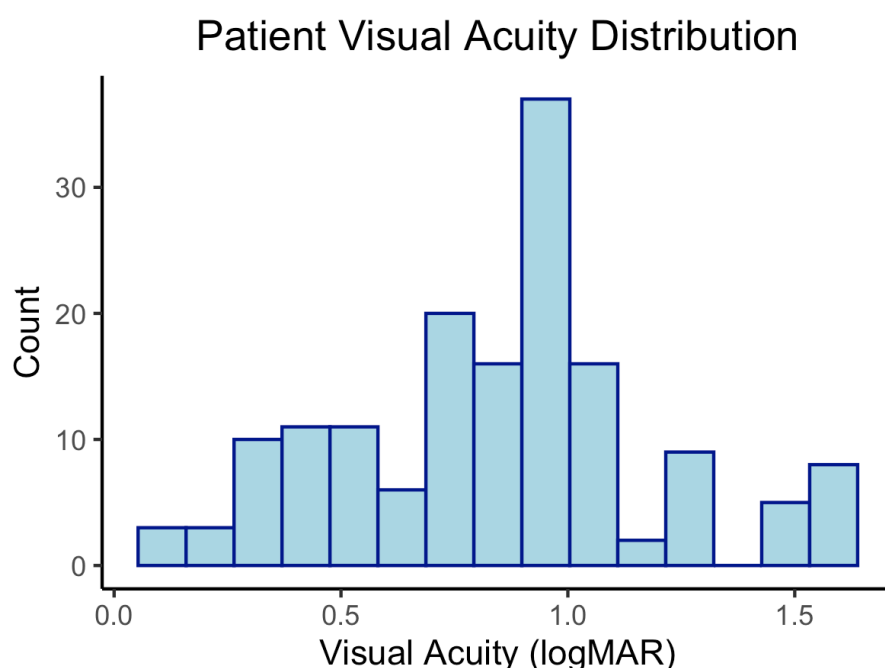


Figure 4-3 Patients Visual Acuity distribution

4.2.3.2 Contrast Sensitivity

The patients' contrast sensitivity results are presented in table 4.2 as average scores for each frequency. The normal group is for comparison, where the participants had no known

visual impairment. Twenty normally sighted participants performed the task binocularly (Age 36 ± 15). Note; that this normal group is not the same group used that performed the face perception test. The normal group is presented to highlight the significant deficit in all frequencies by the patient group, especially at the higher spatial frequencies (Figure 4.4). This is characteristic of central vision loss, where the higher frequencies are impacted earlier in disease.

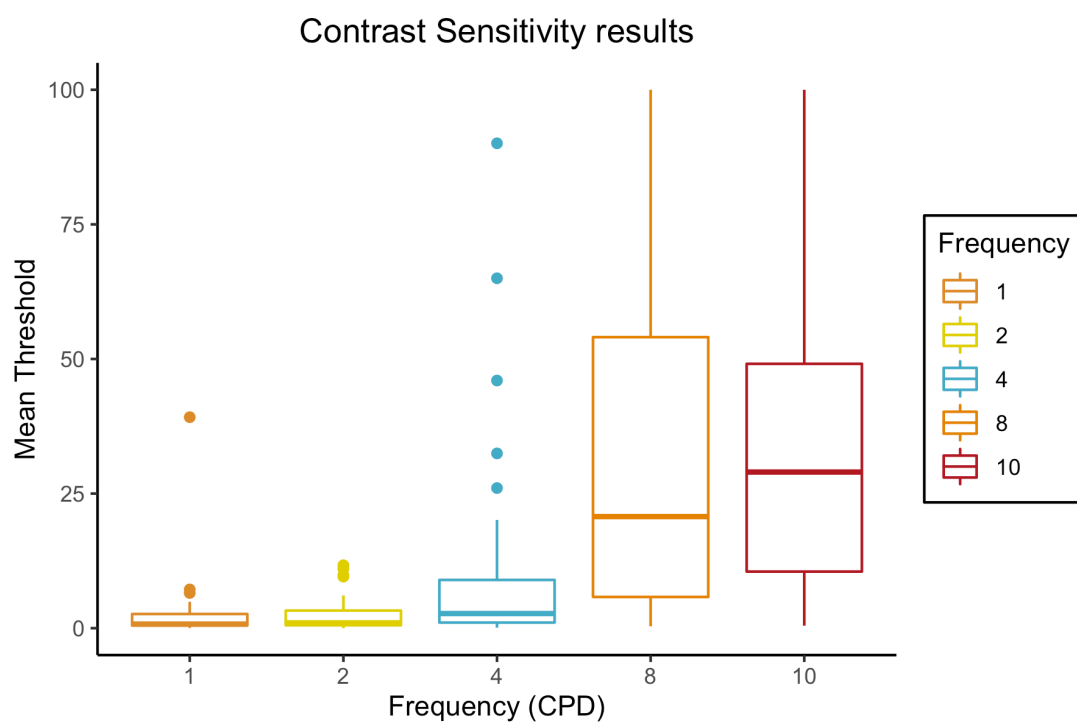


Figure 4-4 Patient Contrast sensitivity results. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.

Table 4-2 Contrast sensitivity results

Group	1 cpd	2 cpd	4 cpd	8 cpd	10 cpd
Normal	0.27 ± 0.15	0.30 ± 0.28	0.25 ± 0.18	0.70 ± 0.7	0.87 ± 1.0
Patients	2.33 ± 5.53	2.36 ± 2.9	9.23 ± 16.84	29.7 ± 28.61	33.46 ± 27.47

All values are mean $\pm$ standard deviation

cpd = Cycles per degree

4.2.3.3 Colour sensitivity

Results from the colour sensitivity results can be seen in figure 4.5, separated by deficit type. The tritan deficit is more prominent, likely due to the significant number of AMD patients. Figure 4.6 highlights the two largest primary diagnosis groups, AMD (20) and Stargardt (16) and how different the two conditions are when considering colour deficit. Table 4.3 presents a comparison of a normal population, highlighting the significant deficit when compared to the patient group.

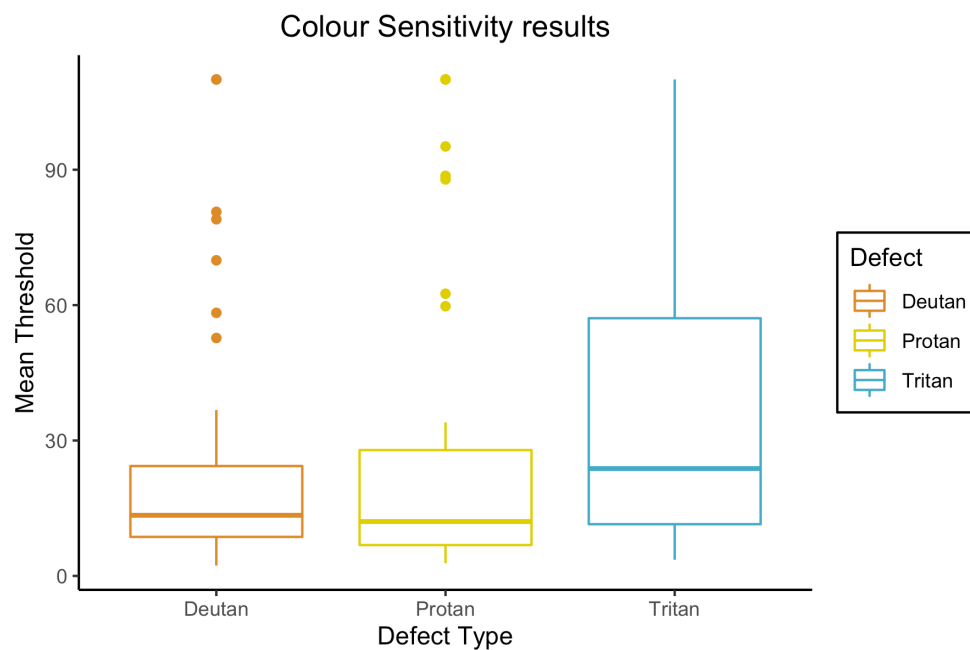


Figure 4-5 Patient colour sensitivity results. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.

Table 4-3 Colour Sensitivity results

Group	Protan	Deutan	Tritan
Normal	4.03±2.40	4.39±3.07	6.65±2.44
Patients	26.76±32.80	24.50±27.35	38.73±34.84

All values are mean±standard deviation

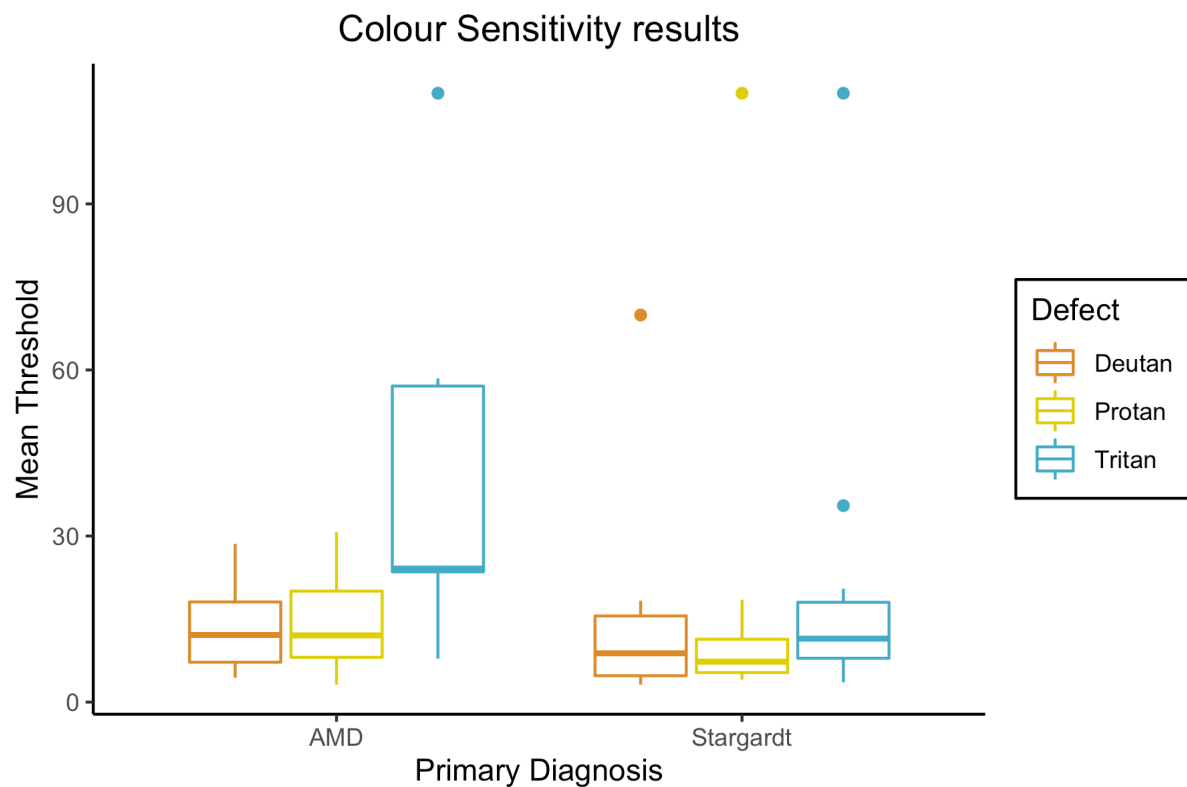


Figure 4-6 Patient colour sensitivity results – comparing AMD to Stargardt. The centre line of the boxes indicates the median threshold. The boxes boundaries the interquartile range, and the lines the range of 1.5 times the interquartile limits.

4.2.4 Face Perception Test

Of the 64 participants who took part, three were unable to perform the baseline face perception test. They failed to identify any correct faces even at the easiest setting. A summary of these individuals is presented in table 4.4.

Table 4-4 Patient participants who could not perform the baseline face perception test.

PID	VA	Condition	Age	Sex
SG006	1.20	AMD	59	M
SP012	1.16	AMD	81	F
SP013	1.60	Stargardt	36	M

Not all participants completed all versions of the enhancements. This is primarily due to time constraints enforced by a busy clinical environment. Tiredness was a factor, with some participants unable or unwilling to continue. Where time constraints were apparent, the Cartoon enhancement was chosen over Peli enhancement, hence a bias in the number of completed tests towards the cartoon enhancement.

4.2.4.1 Score validation

The face perception test uses an adaptive staircase strategy (QUEST) to estimate a threshold score efficiently. For the strategy to be valid, it must converge towards a result. Some participants could not perform the test due to a potential floor effect; thus, I needed to ensure that all those who did complete valid results. Figure 4.7 shows two examples that do and do not converge.

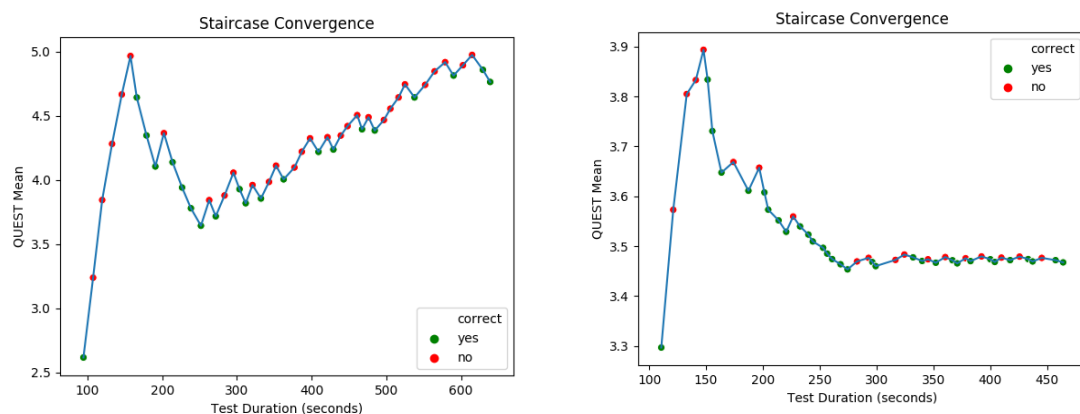


Figure 4-7 The left plot does not converge, with a consistently high score estimate. The right plot does, stabilising after approximately 300 seconds.

After an inspection of the data to identify results that did not converge, scores of greater than 4.0 were considered invalid and marked unsuccessful. This reduced the number of participants who completed the test from 61 to 57.

Eight participants had failures on at least one of the tests. Most of these 8 failed to converge on the normal. However, four failed on the cartoon, and three failed on the adaptive filter. One participant succeeded with the cartoon enhancement, and three failed under all conditions (see *Table 4.5*).

Table 4-5 Participant test failures; summarises the participants whose scores were marked as invalid; the trend is toward lower visual acuity.

Participants Test Failures				
Condition	VA	Age	Sex	Failed on
AMD	1.3	83	M	N,P
AMD	0.9	74	M	N,P,C
AMD	0.92	86	F	N
Stargardt	1.02	44	M	N,P,C
Retinitis Pigmentosa	1.5	31	M	N,C *
Macular Dystrophy	1.56	78	F	N
AMD	1.3	75	F	N,C *
AMD	0.68	65	F	N

N=normal, P=pele, C=cartoon

** did not attempt pele*

Normal Controls also exhibited some invalid results, but in the opposite direction. Three participants achieved very low scores (lower scores indicate a greater face discrimination ability) for the Cartoon enhancement, far lower than the population average (~2.9). After a discussion with the participants, it was understood that they could identify a unique feature, a pixel within the lips, that was visible even at the smallest step (figure 4.8).



Figure 4-8 Shows a lip region from the cartoon enhanced images. Left is the target, and right is the distractor. An example of a feature that some participants used to differentiate the stimulus is highlighted with a red circle.

4.2.4.2 Face Perception Scores - Patients

Each participant performed the face perception test first with no enhancement, then with the Peli and Cartoon enhancement applied. The grouped results for stimulus type are presented in table 4.6 (accounting for the lack of convergence). The results were normally distributed (figure 4.10).

Table 4-6 - Patient face perception scores by stimulus type.

Patient Face Perception Scores			
Stimulus Type	FP Score ¹	Duration (minutes) ²	Response Time (seconds) ²
cartoon	3.38 (0.21)	7.25 (7.86)	8.70 (9.49)
normal	3.39 (0.19)	8.25 (7.61)	9.90 (9.13)
pele	3.36 (0.18)	6.09 (4.12)	7.31 (4.94)
¹ Mean (SD)			
² Median (IQR)			

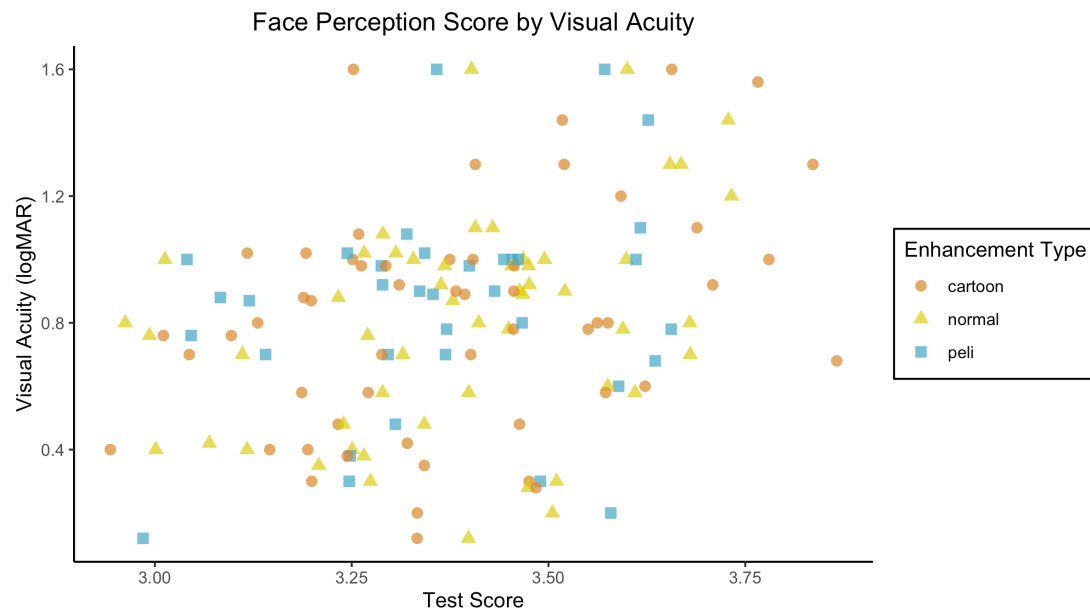


Figure 4-9 - All face perception scores by visual acuity. The coloured points represent the enhancement type; Yellow triangle = Normal, Red circle = Cartoon, and Blue square = Peli.

Visual acuity correlated positively with test scores, Pearson's product-moment correlation = 0.42, $p < 0.001$ for the normal enhancement, but there was a significant amount of variation, as seen in figure 4.9.

The primary aim of this study is to understand if image enhancements improve scores. Thus, the intra-subject rather than inter-subject scores are more relevant.

For intra-subject scoring, enhancement scores are subtracted from the baseline score. e.g. Cartoon – Normal, with a negative value indicating an improvement.

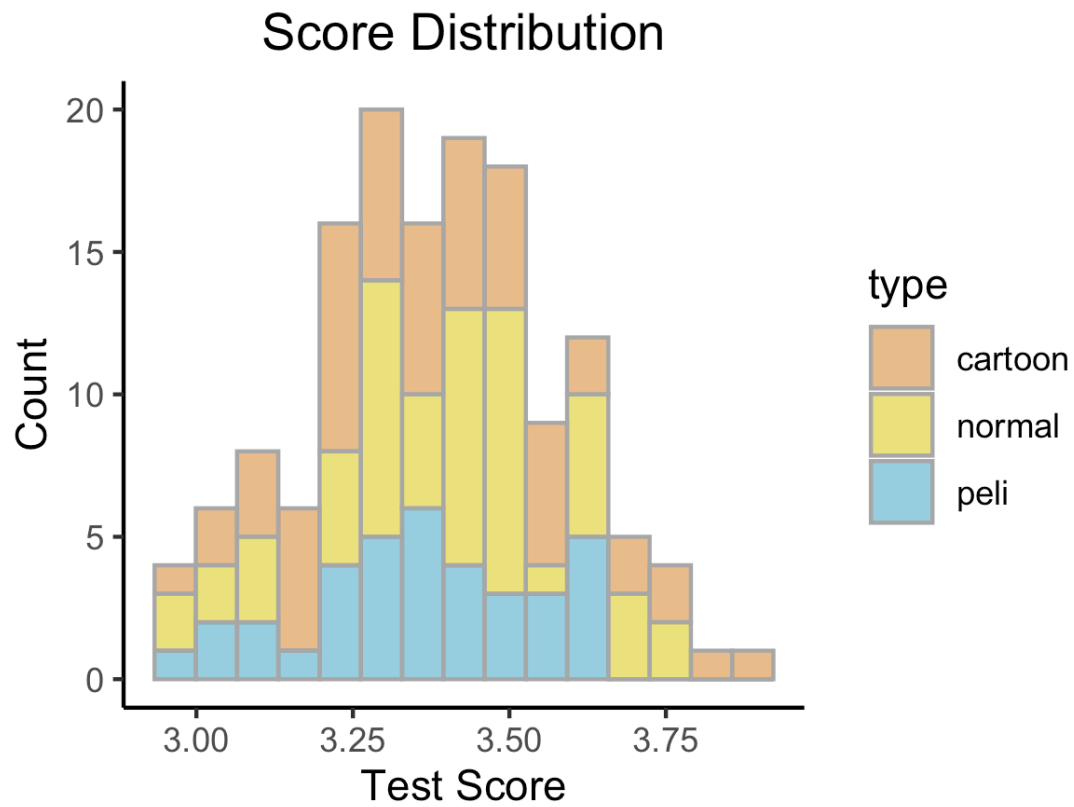


Figure 4-10 Test score distribution by stimulus type.

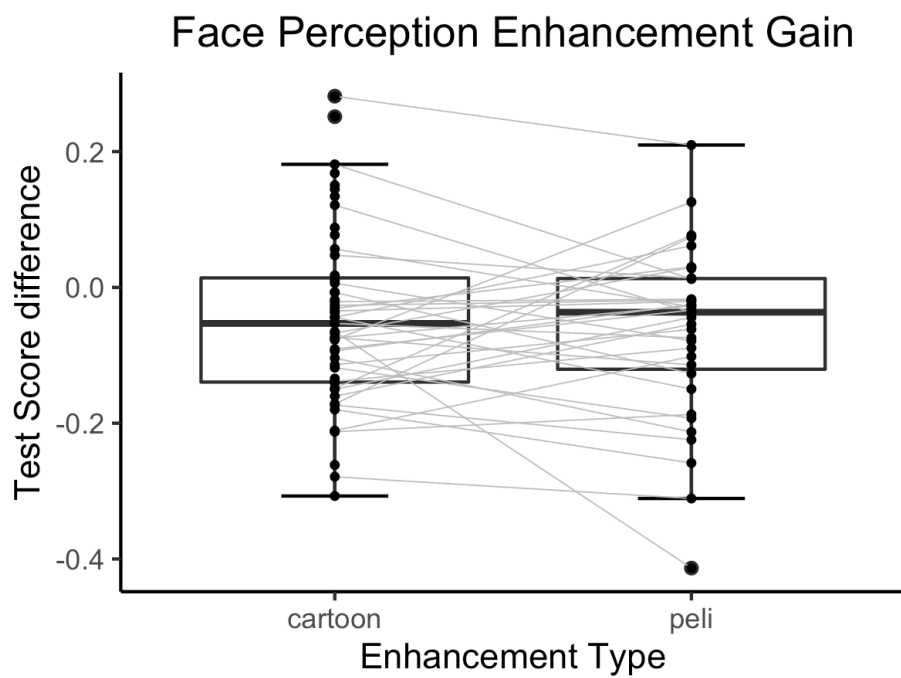


Figure 4-11 – Box plot of patient face perception scores relative to baseline, by enhancement type. A negative score indicates an improvement.

Figure 4.11 presents the relative performance gain. Grouped by type, cartoon enhancements; median= -0.04, IQR=0.172, max=0.33, min -0.31. Adaptive Filter enhancements; median=-0.03, IQR=0.12, max=0.21, min = -0.31.

As discussed in the previous chapter, the CoR₉₅ was defined as 0.12, and a change in a score at or above this value is considered a significant change in performance.

Using the CoR₉₅ as a threshold, I can define positive, negative and neutral improvers;

- Cartoon enhancement: 21 positive, 9 negative and 29 neutral. (total=59)
- Peli enhancement: 11 positive, 3 negative and 25 neutral. (total=39)

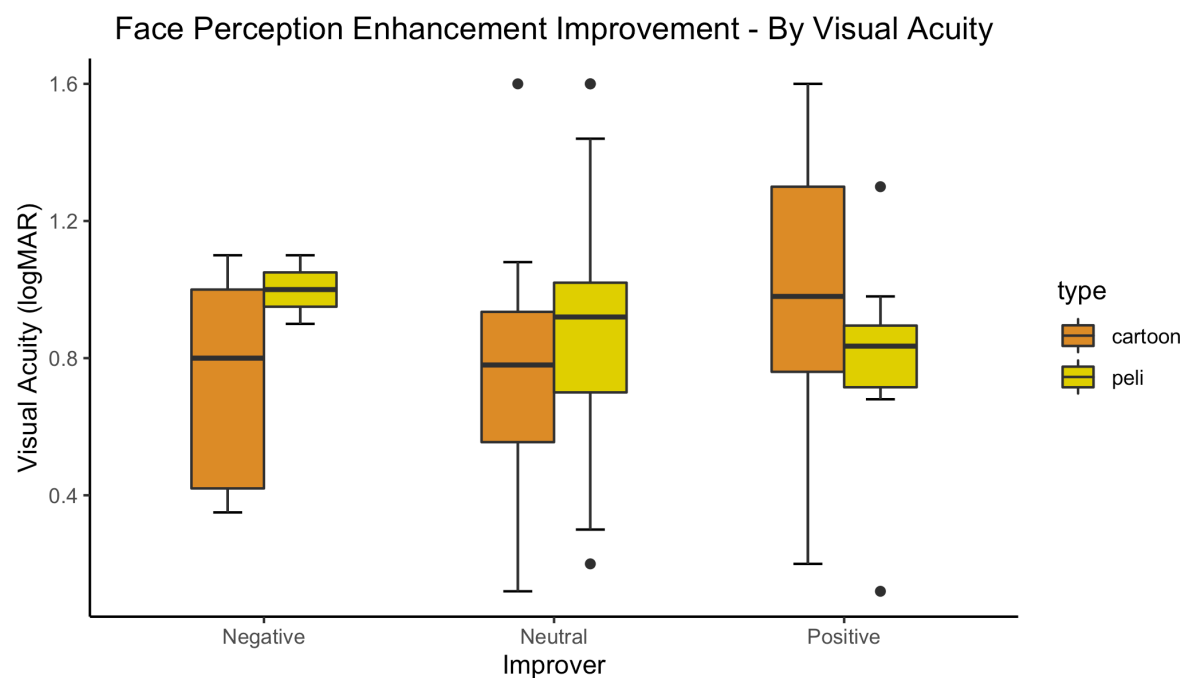


Figure 4-12 - face perception improvement by Visual Acuity. Colours indicate the enhancement type (orange=Cartoon, Yellow = Peli).

I grouped the participants into their improver category, and I then compared their visual acuity scores (figure 4.12). A one-way analysis of variance indicated a significant difference [F(2,48)=4.39, p=0.01] with the visual acuities between the groups for the Cartoon

enhancement. There were no differences for the Peli enhancement [$F(2,32)=0.235$, $p=0.8$].

A posthoc analysis, unpaired two-sided t-test, comparing the means of Positive->Neutral for the Cartoon enhancement presented a significant difference [$t = -2.4244$, $df = 19.537$, $p\text{-value} = 0.02518$].

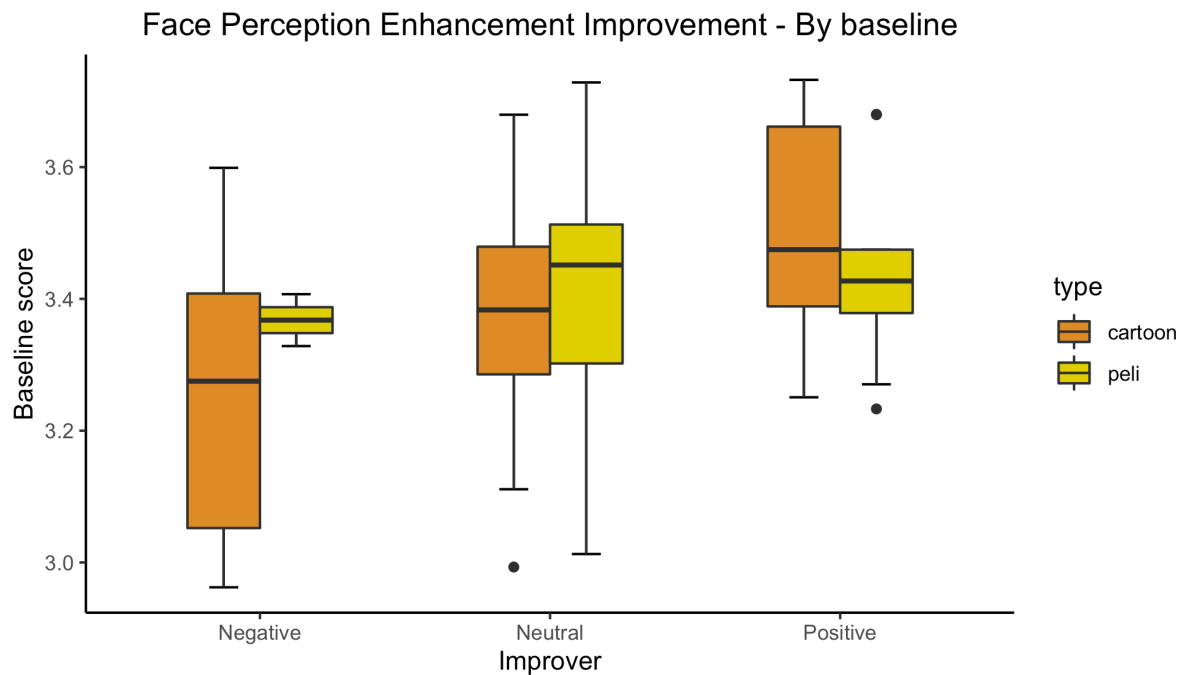


Figure 4-13 All performances grouped by their baseline face perception scores (unenhanced normal).

Using the same grouping, I investigated the relationship between improvers and their baseline scores (figure 4.13). A one-way analysis of variance showed a significant difference [$F(2,48)=5.076$, $p=0.01$]. An unpaired two-sided t-test comparing the means of Positive->Negative for the cartoon enhancement presented a significant difference [$t = -2.7527$, $df = 12.427$, $p\text{-value} = 0.017$].

This indication is that a lower visual acuity and baseline face perception scores are related to a positive improvement with the cartoon image enhancement. This differentiation is harder to make with the Peli enhancement; this could be due to the smaller group size, but

also the overall lower face perception performance when compared to the cartoon enhancement.

To summarise, 26 participants showed improvement with at least one type of enhancement, 11 of these with both Cartoon and Peli. 11 had a negative improvement with an enhancement, with only 1 participant having a negative improvement with both enhancements. Twenty-three had no improvement with either enhancement.

There is no intersection of the positive and negative improvers, meaning if a participant had any improvement, with any enhancement, they did not have a negative improvement.

4.2.4.3 Face Perception Scores - Controls

The control group is used to understand if there is a limitation to the enhancements. I.e. is there any inherent limitation of the enhancements that will limit any potential improvement.

The scores grouped by enhancement type for the control participants are presented in table 4.7. Using the same method as the patient data, enhancement gain is illustrated in figure 4.14.

When considering a significant score change, a difference greater than 0.12 (CoR₉₅ threshold):

- Cartoon enhancement: 8 positive, two negative and nine neutral
- Peli enhancement: 4 positive, two negative and 16 neutral.

Eleven participants had a positive improvement in at least one enhancement, and 3 had a negative improvement, with two in both enhancements. Eight people had no improvement with any enhancement.

There is an improvement trend with the cartoon enhancement; however, as shown in Score Validation, a number of the control participants performed much better than expected due to image artefacts introduced by the Cartoon enhancement (3 scores were removed from analysis). Thus, the results should be treated with caution.

Table 4-7 Control participants face perception score summary.

Control Face Perception Scores			
Stimulus Type	FP Score ¹	Duration (minutes) ²	Response Time (seconds) ²
cartoon	3.02 (0.16)	5.63 (2.58)	6.83 (2.78)
normal	3.08 (0.12)	5.56 (3.43)	6.67 (4.12)
pele	3.06 (0.11)	5.70 (2.70)	6.84 (3.25)
¹ Mean (SD)			
² Median (IQR)			

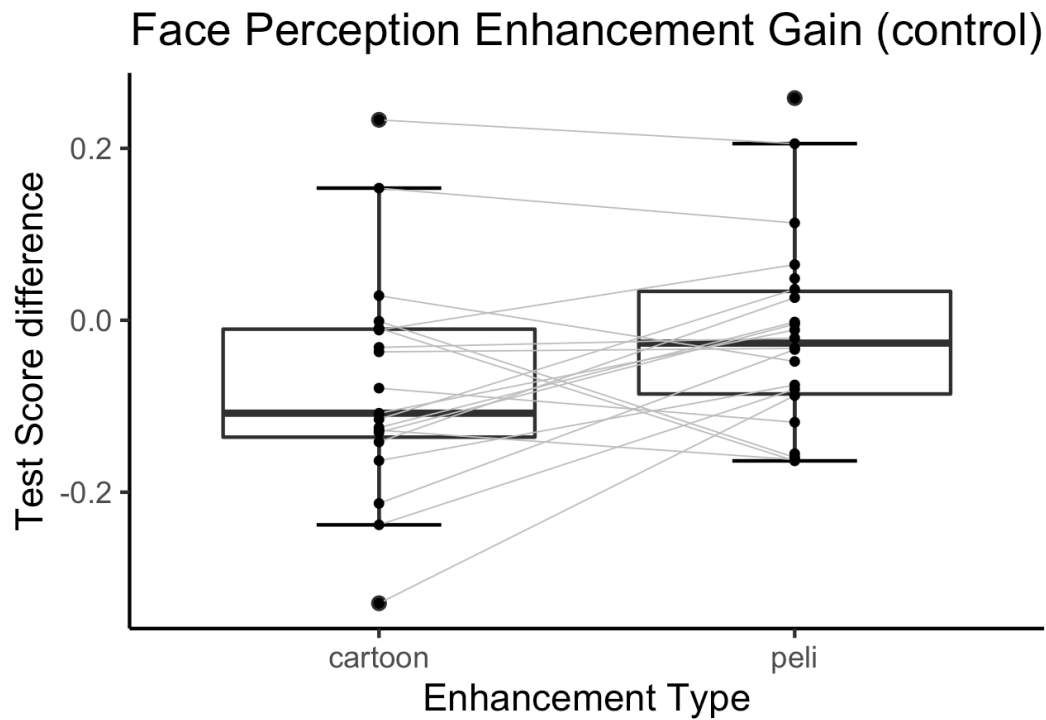


Figure 4-14 Control participant Face perception score relative to baseline. A negative score is an improvement in performance.

4.2.4.4 Test Duration

The total duration of each test is investigated to identify any interaction with enhancement type. Test duration was not normally distributed. For the patient group, Median = 7.16 IQR = 6.47 (minutes). Normal controls; Median=5.6 IQR=3.26 (minutes). Figures 4.15 present the duration distribution for the patient group.

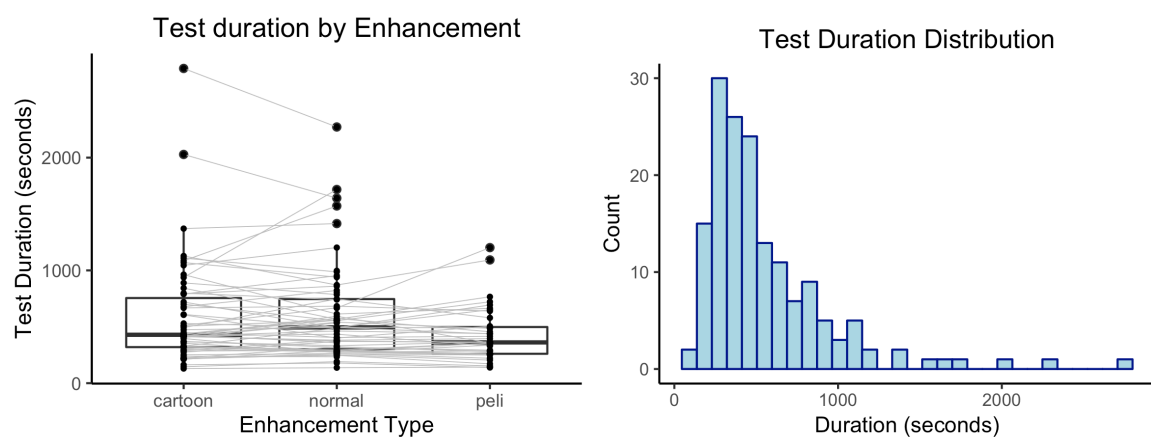


Figure 4-15 (left) Test duration by enhancement type. (right) Test duration distribution plot.

When comparing the three enhancement types' durations, a Kruskal-Wallis rank sum test found an indication of a trend $H(2)=4.85$, $p=0.053$]. However, there is an imbalance in the data where fewer participants performed the Peli enhancement. Accounting for this imbalance, by comparing participants who performed all tests presented in figure 4.15, the Kruskal-Wallis rank-sum test found no differences [$H(2) = 0.059$, $p=0.9$], indicating no difference in test duration by enhancement type. The control group had no differences between enhancement types.

For the patient group, age correlated with test duration, Spearman's rank correlation ($\rho = 0.35$, $p<0.001$). This is not observed in the control group ($\rho=0.01$, $p=0.9$), where the mean age is significantly lower than in the patient group.

4.2.4.5 Response Times

I investigated the response times to understand if there was any behavioural difference with the different enhancements that total scores may not capture.

Response time is defined as the time a participant took to respond from the first viewing of an individual trial. Patient group median response times; 8.23(seconds) IQR=6.6 (figure 4.16)

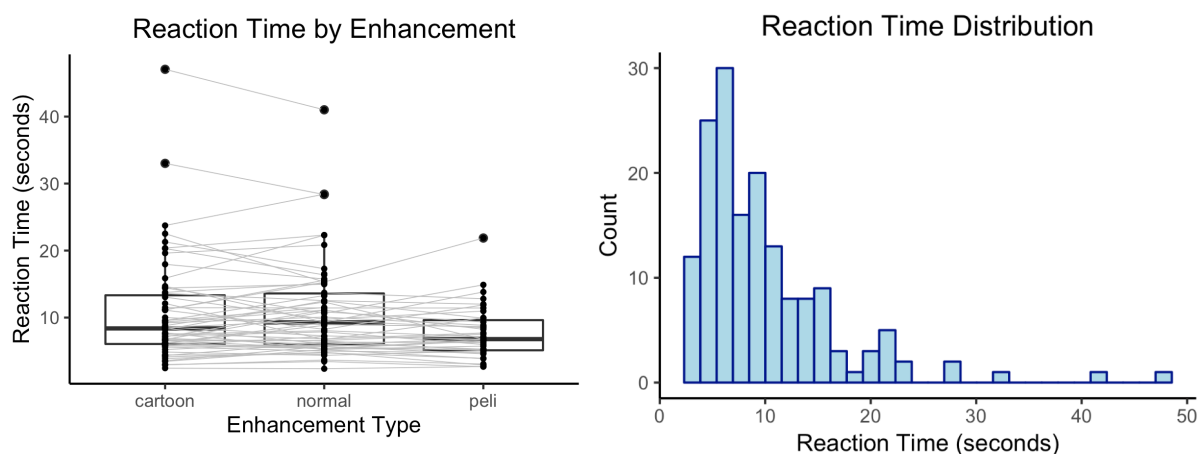


Figure 4-16 (left) Test reaction time by enhancement type. (right) Test reaction time distribution plot.

There was a significant difference between the response times for a correct (median 7.0ms, IQR=7.4ms) and incorrect response (median=9.2ms, IQR=9.3ms) [Wilcoxon rank-sum test, $W=9453818$, $p<.0001$]. There was no statistically significant difference in response time for the enhancement types (figure 4.16) Kruskal-Wallis rank-sum test [$H(2) = 0.07$, $p=0.9$].

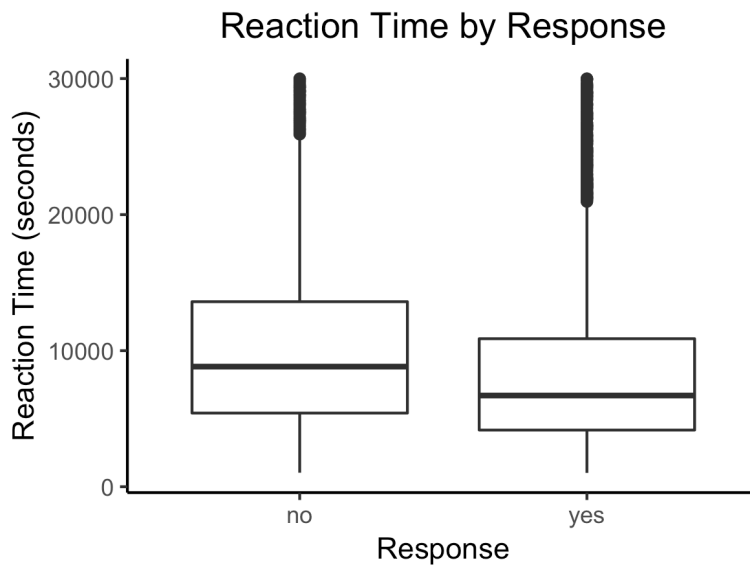


Figure 4-17 Participant reaction time by response type. Where the response is a correct or incorrect answer to an individual trial.

4.2.5 Spectral analysis

A Fourier transform was performed on all of the stimuli in the dataset and radially averaged to produce the plot in figure 4.18. This was performed for each enhancement; normal, Peli and cartoon. It is clear from figure 4.18 that the cartoon increased the amplitude of all the component frequencies the most, with the Peli enhancement second.

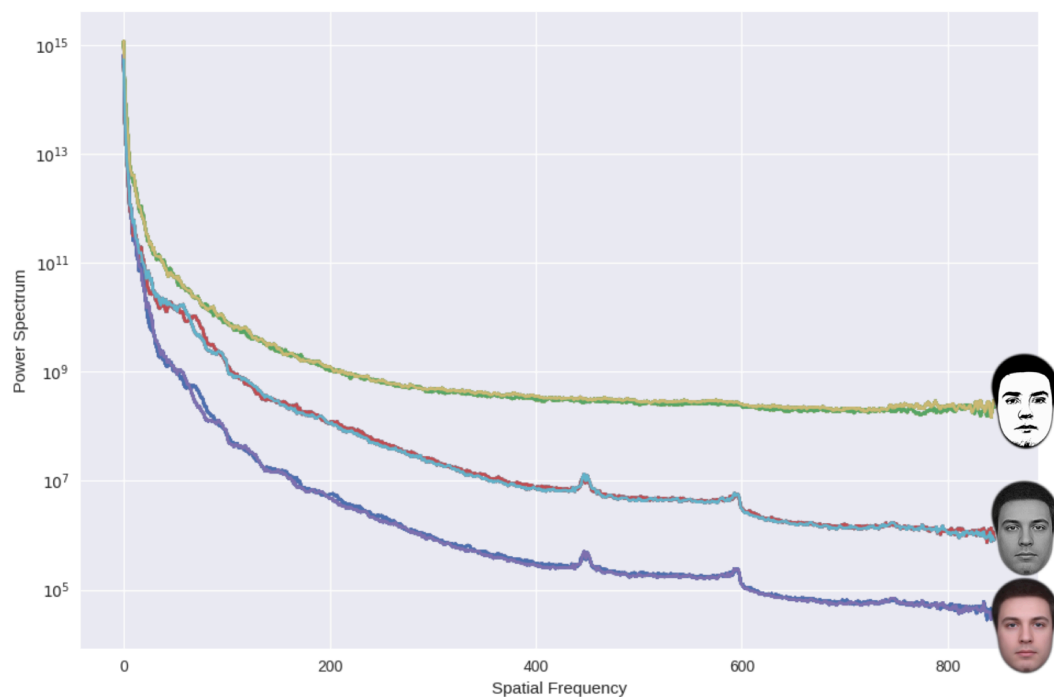


Figure 4-18 Spatial frequency analysis of the face stimulus. Three lines on the plot represent all of the stimuli with the three modalities; Cartoon (top plot), Peli (middle plot) and normal (bottom plot).

4.2.6 Factors relating to performance

I have collected a wealth of data from the participants relating to their functional vision. Can

I identify from this collection any factors that are predictive of face perception

performance?

4.2.6.1 Statistical Methods

The aim is to predict the face perception score, a dependent variable, from a selection of

parameters (independent variables). The parameters are a combination of vision

assessments taken from the participants, personal, age, sex and primary diagnosis.

Parameters

Visual acuity, age, contrast sensitivity, reading acuity, critical print size, colour sensitivity and primary diagnosis.

13 in total; contrast sensitivity and colour are separated into their separate parts (1,2,4,8, 10 cycles per degree) & (Protan, Deutan and Tritan)

Methods

The statistical method used is multiple linear regression, combining multiple parameters to improve the linear model's predictive power. While it is a simple approach, assuming a linear relationship provides for a more straightforward interpretation. The aim here is not so much prediction, although this will be used to assess the quality of the model, but to describe performance.

To interpret these results, I must describe how the regression models are working.

Multiple regression is an expansion of linear regression

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Where I am trying to predict a quantitative response (Y) based on a single predictor (X), assuming some error (ϵ). This is performed by estimating the variables β_0 and β_1 . The core assumption is that the relationship is linear.

The method of estimating is to choose values for β_0 and β_1 that minimise the residual sum of squares (RSS). The residual is the error, or difference, between the predicted value $\hat{y}_i$ and the actual value y_i .

$$RSS = \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

For multiple linear regression, multiple predictors are combined; suppose there are p predictors, the equation now takes the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Parameter Selection

It is unlikely that all of the parameters will be predictive of performance, so selecting the best combination is a necessary process. Using a step-wise best subset selection method (James, Witten, Hastie, & Tibshirani, 2013b), a linear model is fit to each of the possible permutations of the parameters, resulting in 2^{13} different models.

The model performance is assessed using Akaike information criterion (AIC). AIC estimates the prediction error and determines the relative quality of a set of models for a given dataset.

Model Validation

Secondary model validation is performed on the models selected using the step-wise selection method. For each of the models used, K-folds cross-validation is used to understand the model quality. This aids in detecting an overfitting mode; for example, if the error of a model is significantly smaller for the initial fit than the k-folds fit, then the model is likely overfitting to the training data and will not generalise well.

The resulting models were assessed for assumptions of homogeneity of variance, linearity and the residuals were approximately normally distributed.

Missing Data

Due to the occurrence of missing data, the regression methods require the removal of incomplete measurements. This reduces the total number of observations (participants).

Using the full model investigating performance on the normal condition, I have 33 participants, Cartoon 34 and Peli 19; this is a significant drop of near 40% and will negatively impact the predictive models' generalisation.

Due to the complexities of collecting large amounts of patient data, timing, equipment malfunction etc., missing data are difficult to avoid.

All statistical analysis was performed using the R statistical package (R Development Core Team, 2008).

4.2.6.2 Results

Initially, I attempted to predict the scores of the face perception test under the normal condition (unenhanced faces). A significant regression equation was found ($F(4,33) = 5.398$, $p=0.002$), with an R^2 of 0.32 and an RMSE of 0.168.

$$Y = VA\beta_0 + \text{age}\beta_1 + \text{CSF}_2\beta_2 + \text{CSF}_8\beta_3 + \epsilon$$

Where Y is the test score, estimates of the coefficients are presented in table 4.8.

Table 4-8 Linear regression model coefficients for normal enhancements

Characteristic	Beta	95% CI ¹	p-value
VA	0.391	0.198, 0.583	<0.001
Age	0.002	0.000, 0.005	0.073
csf_2_mean	0.033	0.009, 0.058	0.009
csf_8_mean	-0.003	-0.006, -0.001	0.018
¹ CI = Confidence Interval			

Next, I attempt to predict the cartoon enhancement scores.

A significant regression equation was found ($F(4,34) = 7.341$, $p = 0.0002$ with an $R^2 = 0.432$ and an RMSE of 0.153.

$$Y = VA\beta_0 + age\beta_1 + CSF_1\beta_2 + CSF_8\beta_3 + \epsilon$$

Where Y is the cartoon enhancement test score, estimates of the coefficients are presented in table 4.9.

Table 4-9 Linear regression model coefficients for cartoon enhancements

Characteristic	Beta	95% CI ¹	p-value
Age	0.003	0.001, 0.005	0.019
logRAD	0.395	0.194, 0.597	<0.001
csf_1_mean	0.038	0.008, 0.068	0.015
csf_8_mean	-0.002	-0.004, 0.000	0.044
¹ CI = Confidence Interval			

Finally, I attempt to predict the Peli enhancement scores.

A significant regression equation was found ($F(4,19) = 15.2$, $p < 0.0001$) with an $R^2 = 0.71$ and an RMSE of 0.104.

$$Y = age\beta_0 + LogRAD\beta_1 + CSF_2\beta_2 + CSF_8\beta_3 + \epsilon$$

Where Y is the Peli enhancement test score, estimates of the coefficients are presented in table 4.10.

Table 4-10 Linear regression model coefficients for Peli enhancements

Characteristic	Beta	95% CI ¹	p-value
Age	0.003	0.001, 0.006	0.018
logRAD	0.601	0.424, 0.777	<0.001
csf_2_mean	0.021	0.005, 0.036	0.012
csf_8_mean	-0.002	-0.004, 0.000	0.066
¹ CI = Confidence Interval			

The contrast sensitivity is clearly an important component and is consistently involved in the regression models. However, visual acuity is only included in the normal condition. I used the same methodology to attempt to predict visual acuity from contrast sensitivity.

Visual acuity in the patient group correlated with contrast sensitivity loss. Using step-wise multiple linear regression using all frequencies (1,2,4,8 and 10 cycles per degree), a significant regression equation was found ($F(3,126) = 22.54$, $p < 0.001$), with an R^2 of 0.35 and an RMSE of 0.29.

$$Y = CSF_2 \beta_0 + CSF_4 \beta_1 + CSF_8 \beta_2 + \epsilon$$

Where Y is the visual acuity, CSF_n is the frequency at n cycles per degree. The β_n is the estimate of the coefficients (table 4.11). The negative Beta value at two cycles per degree indicates that a loss of sensitivity is negatively contributing to the model. The sensitivity loss at 4 and 8 cycles degree positively contributes to the model predicting visual acuity.

Table 4-11 Estimates of the linear regression model Coefficients. The Characteristic column represents the frequencies tested (cycles per degree).

Characteristic	Beta (95% CI) ¹	p-value
2	-0.038 (-0.061 to -0.015)	0.001
4	0.009 (0.006 to 0.013)	<0.001
8	0.006 (0.004 to 0.008)	<0.001

¹ CI = Confidence Interval

4.2.6.3 Discussion

The results of the step-wise regression analysis can aid the understanding of factors related to face recognition performance. For the baseline condition, no enhancement; the performance was best described by a participant's visual acuity, age and contrast sensitivity at 2 and 8 cycles per degree (table 4.8). These findings agree with other studies (Barnes, De l'Aune, et al., 2011; Taylor, Smith, et al., 2018) while also indicating an age effect. With the inclusion of age in the normal enhancement model, the estimator beta is very close to zero and has a non-significant p-value (implying it is hard to distinguish from zero and may not contribute much to the model). Its removal reduces the model's fit (adj-R^2) and increases the AIC. (without age: $\text{AIC} = -20.98$, $\text{adj-R}^2 = 0.274$, with age $\text{AIC} = -22.75$, $\text{adj-R}^2 = 0.32$). Age as a correlating factor is most likely due to the high percentage of age-related macular degeneration (31%) driving the relationship rather than anything intrinsic. However, it cannot be confidently ruled out.

Age becomes a statistically significant factor when considering the image enhancement conditions, while the Beta value is similar to ~ 0.003 (table 4.9 & 4.10). Again, this correlation is likely due to the patient sample and the relatively small number of subjects.

Other observations are the logRAD beta for cartoon enhancement has a similar value to the VA beta for baseline condition (table 4.8). The Peli logRAD beta is much higher (table 2.3.8), implying it has more of an impact on predicting the results. It should be noted that the Peli dataset is smaller ($n=19$).

The results from predicting visual acuity with contrast sensitivity highlight an interesting difference. The coefficients of the regressions models for predicting face perception all include the lower spatial frequencies contributing positively, 1 and 2 cycles per degree, and

the higher negatively contributing (8 cpd), where the model is predicting poorer performance (a higher score is a poorer performance). The inverse is found with visual acuity; the lower spatial frequencies (2 cpd) negatively contribute and the higher positively (visual acuity is poorer with a higher value). The relevance of this is that while visual acuity is correlated to face perception performance, the intrinsic spatial frequencies are very different. As central vision loss starts with impacting the higher frequencies and progresses to lower, this could explain why the correlation of improvement and visual acuity is more apparent with poorer VA (higher numeric logMAR), and the correlations are less apparent with the better VA (figure 4.9).

In conclusion, the regression models have similar components with slightly different model coefficients. Baseline performance is mostly driven by lower spatial frequency loss. Both cartoon and Peli enhancement performance is driven by Reading Acuity and lower spatial frequency.

4.2.7 Questionnaires

As part of the data collection, I have collected quality of life information in the form of questionnaires. These are the DLTV and VFQ-25. They provide numerical scores, a single total score for the DLTV and ten sub-scores for the VFQ-25.

Questionnaires' while not perfect, are a simple method of collecting information. Testing in clinical is time-consuming and expensive. If any part of this process can be simplified or sped up, that will be a benefit.

In the context of wearable displays and image enhancements, from the information presented in this study, it is apparent that not everyone will obtain the same degree of benefit.

This presents an exciting hypothesis, can the questionnaire data be used to predict positive performers, i.e. those who are more likely to benefit from image enhancements and a wearable display.

4.2.7.1 Statistical methods

The goal is to use the questionnaire data to predict functional performance. Can a model predict a continuous variable, visual acuity, or face perception score? The methods used are similar to 4.2.6.1. However, this data requires some different approaches.

4.2.7.1.1 Model selection

Three distinct methods will be used. Firstly, multiple linear regression, whose advantage lies in interpretability, and the resulting models are simple to describe and understand. The disadvantage is in the parametric assumptions and model simplicity.

Secondly, a bagging (James, Witten, Hastie, & Tibshirani, 2013e) method is used, specifically a Random forest (RF) (Tin Kam Ho, 1995). Bagging works by combining the output of multiple models (a collection of decision trees). Predictions are made by averaging the output of the predictions of all the sub-trees. Advantages of a random forest; robust against overfitting of models, handles high-dimensionality well and handles non-parametric data. The disadvantage is it won't give precise values for regression as its prediction output is the mean of a collection of decision trees, and its interpretability is limited.

Finally, a boosting (James et al., 2013e) method, specifically gradient boosted trees (GBT) (Friedman, 2001). The idea of boosting is that it sequentially uses multiple models. Starting with a weaker classifier (a simple decision tree), further models are built and applied to output to attempt to correct the errors. The advantages of gradient boosted trees are the potential for higher accuracy than bagging methods, and they perform well with unbalanced data. The disadvantages are they are prone to overfitting and are slower to train.

Validation – for each of the models used, K-folds cross-validation is used to understand the model quality.

Independent Variables

The focus is on the questionnaires and the advantage they can provide. Thus, only data obtainable from them will be used. For the DLTV, there is a single score; the VFQ-25 has an overall score and ten sub-categories, producing 12 variables.

Furthermore, each question can be used, 24 for the DLTV and 25 for the VFQ-25. The aim is to select from a subset of individual questions that can be more predictive than the questionnaire totals and sub-totals.

4.2.7.1.2 Feature Selection

The selection of the independent variables has an impact on model performance. Features that have little importance can negatively impact the model fitting. Also, not all methods, specifically linear regression, are suitable for use with high-dimensional data (James et al., 2013b). As a result, limiting the number of features by removing those that have little to no impact on a model's performance is an important step.

There exist numerous strategies for feature selection; one method already used in this thesis is step-wise selection, where features are sequentially added or removed. This can perform well, but when the number of features increases, iterating through every combination of features is prohibitive. For example, testing every combination of questions in the DLTV questionnaire containing 24 questions will result in 2^{24} combinations equalling 16,777,216 permutations; this is not feasible to compute in a useful time frame.

Thus, alternate strategies are required. One option is to use lasso regression, a type of linear regression that includes a shrinkage penalty (James et al., 2013b). This has the effect of performing feature selection and regularisation and is much more applicable to high-dimensional data than linear regression. The method can be used in place of linear regression or as a tool for feature selection.

Random forests can also be used for feature selection; it has the built-in ability to calculate feature importance. Two metrics can be computed: the Gini Importance (or mean decrease impurity), a sum of the number of splits across all the trees that use a particular feature. Secondly, the mean decrease in accuracy is an estimate of the prediction strength of each feature (Hastie, Tibshirani, & Friedman, 2009b). The two measures tend to produce similar results but are not always identical. Experiments performed by Calle and Urrea (Calle & Urrea, 2011) on the stability of the two metrics found the Gini importance to be more stable to small perturbations in data. Thus, I shall use the Gini importance for feature selection.

Multicollinearity is an effect to be considered, especially so for the linear regression models. Multicollinearity is where features are correlated to each other and reduce the accuracy of

the estimate of the model coefficients (James, Witten, Hastie, & Tibshirani, 2013c). Thus, identifying and removing any features with high collinearity is essential and reduces the dimensionality of the data.

Process

For each dependent variable, multicollinearity is assessed, and relevant features are removed. Then a collection of features is chosen using the step-wise selection if the number of features is less than twelve. If higher, lasso regression and Random Forrest feature importance will be used. The RF feature importance approach selects the top 10 features, increasing until the model RMSE stops reducing.

Each selection of features is used with the three models, Linear Regression, Random Forrest and Gradient boosted trees. Validated with k-folds cross-validation, and the model with the lowest RMSE is selected as the best.

4.2.7.2 Results

A distribution plot of the totals and sub-totals is shown in figure 4.19. There is a significant variance in the types of destitutions.

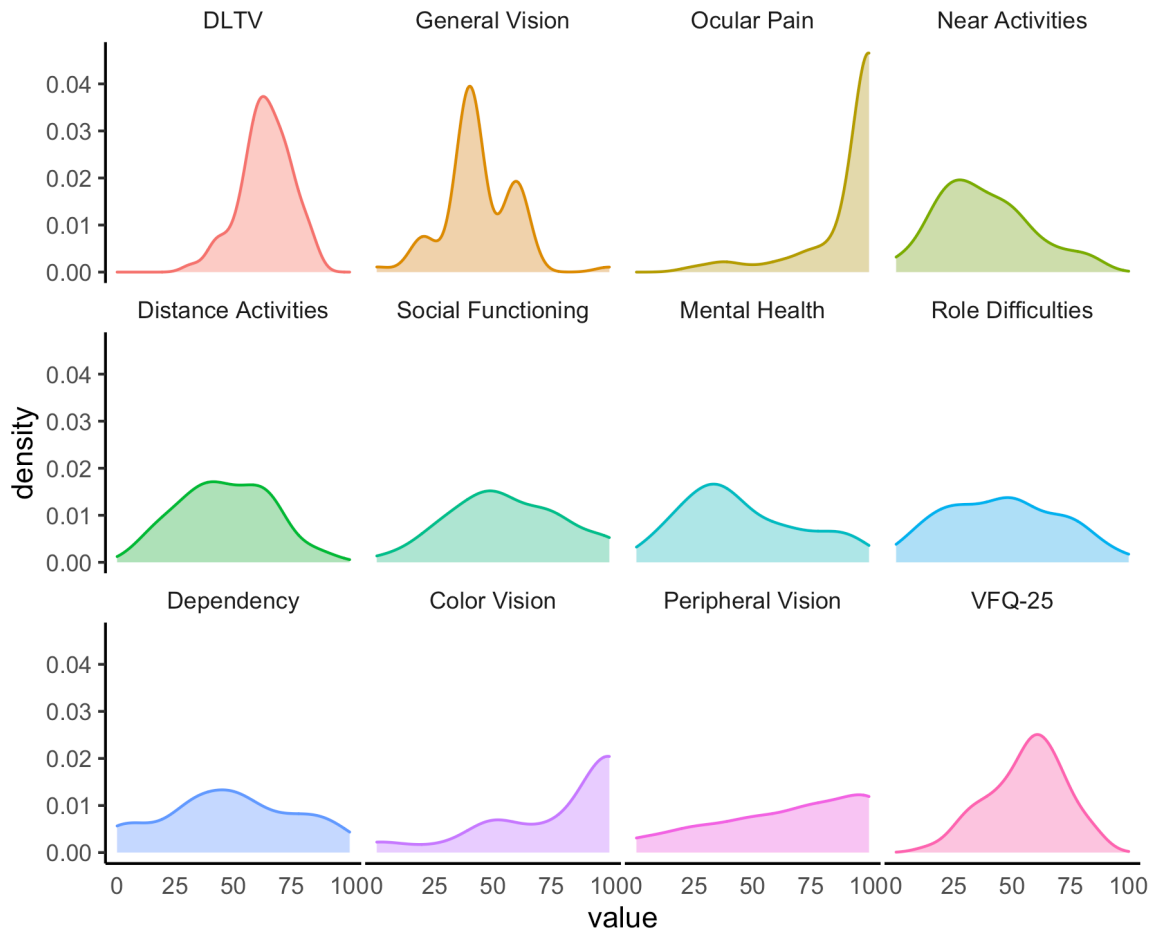


Figure 4-19 The overall distribution of scores for the questionnaires totals and subtotals. Each plot represents a total. DLTV and VFQ-25 are for the whole questionnaires; the rest are sub-totals from the VFQ-25.

I used the questionnaire totals and subtotals first; the results can be seen in table 4.12. The best model RMSE value is highlighted in bold. The RMSE values are similar to the scores found in section 4.2.6.2, where I tried to predict the same dependent variable using the functional data. On the whole, the linear regression models had the lowest RMSE, except for visual acuity. The variance across all the models RMSE is low, which indicates that when using more complex methods (RF and GBT) did not improve the error, there is likely little more predictive power in the data.

Table 4-12 Regression Results

Dependent Variable	Linear Regression	Random Forrest	Gradient Boosted Trees	Best Model
Visual Acuity	0.314	0.291	0.311	DLTV + Ocular Pain + Near Activities + Distance Activities + Social.Functioning
Face Perception - Normal	0.163	0.184	0.177	Ocular pain + Near.Activities
Face Perception - Cartoon	0.184	0.215	0.206	DLTV + General Vision +Ocular Pain +Distance Activities +Social Functioning +Mental Health +Role Difficulties+ Peripheral Vision
Face Perception - Peli	0.152	0.211	0.174	DLTV + Ocular Pain + Distance Activities + Social Functioning + Peripheral Vision

*All values are RMSE - **Bold** values are the best performing model scores*

I then used every question in both questionnaires as features for the regression models; Table 4.13 presents the results. The bold value is the lowest RMSE and is considered the best model. There is an overall reduction in RMSE for all the dependent variables. The best selection is now tending to the random forest; this could be due to its ability to handle non-parametric data better than linear regression. However, the variance in RMSE across all models is low, indicating I am at the limit of what this data can predict.

Table 4-13 Regression results - Individual Questions

Dependent Variable	Linear Regression	Random Forrest	Gradient Boosted Trees	Best Model
Visual Acuity	0.30	0.251	0.24	X1 + X8 + X9 + X11 + X16 + X17 + X19 + X20 + X21 + X22 + Y11 + Y.22 + X25 + Y7 + Y8
Face Perception - Normal	0.157	0.14	0.141	X9 + X19 + X20 + X21 + X22 + X23 + Y19 + Y3 + Y4
Face Perception - Cartoon	0.179	0.16	0.168	X2 + X9 + X18 + X19 + X20 + X22 + X23 + Y3 + Y5 + Y6
Face Perception - Peli	0.187	0.124	0.141	X2 + X4 + X9 + X18 + X19 + Y10 + Y4 + Y5

*All values are RMSE - **Bold** values are the best*
X prefix = DLTV question, Y prefix = VFQ-25 question.

The questions selected by the models can be seen in table 4.14. The complete questionnaires are available in appendix A. There are common questions surrounding difficulty with reading – which relate to the logRAD score correlating with face perception performance. Some less obvious, for example, X2 'noticing objects off to the side' this question likely relates to peripheral loss; an investigation into its contribution may be more

informative. However, the choice of models used makes this task difficult and is the trade-off for predictive power over interpretability. But this experiment aims to predict rather than explain.

Table 4-14 Selected questions from the DLTV and VFQ-25

	Question ID	Question prefix	Question	VA	Normal	Cartoon	Peli
DLTV	X1	How much difficulty do you have	Distinguishing a person's features across the room	X			
	X2		Noticing objects off to either side			X	X
	X4		Seeing steps and using them				X
	X8		Recognizing seasonal changes in the garden	X			
	X9		Distinguishing a person's features at arm's length		X	X	X
	X11		Cutting up food on your plate	X			
	X16	How confident do you feel in your ability to walk around	In your immediate neighbourhood	X			
	X17		Outside your immediate neighbourhood	X			
	X18	With your near glasses on how much difficulty do you have	Reading normal sized newspaper print			X	X
	X19		Reading newspaper headlines	X	X	X	X
	X20		Reading correspondence (e.g., bills, letters, cards)	X	X	X	
	X21		Signing documents (checks, etc.)	X	X		
	X22		Identifying money from purse or wallet	X	X	X	
	X23	How would you rate	Your overall distance vision	X	X	X	
VFQ-25	Y3		How much of the time do you worry about your eyesight?		X	X	
	Y4		How much pain or discomfort have you had in and around your eyes		X		X
	Y5		How much difficulty do you have reading ordinary print in newspapers?			X	X
	Y6		How much difficulty do you have doing work or hobbies that require you to see well up close, such as cooking, sewing, fixing things around the house, or using hand tools?			X	
	Y7		Because of your eyesight, how much difficulty do you have finding something on a crowded shelf?	X			
	Y8		How much difficulty do you have reading street signs or the names of stores?	X			
	Y10		Because of your eyesight, how much difficulty do you have Noticing objects off to the side while you are walking along?				X
	Y11		Because of your eyesight, how much difficulty do you have seeing how people react to things you say?	X			
	Y22		I have much less control over what I do, because of my eyesight	X			
		Please tell me if it is definitely true, mostly true, mostly false, or definitely false for you or you are not sure.					
Total number of questions				14	8	8	8

X denotes inclusion of the question in the model, predicting either VA or face perception scores from the three modes. (Normal, Cartoon and Peli)

4.2.7.3 Discussion

An essential feature of these results is how close the RMSE (0.14 in the best model) is to the CoR₉₅ threshold of 0.12 for the face perception tests. This indicates that the models are likely to predict an individual's score to a useable degree of certainty.

Suppose I take into consideration the results from the analysis of the improvers and non-improvers (4.2.4.2). A significant difference between positive and negative improvers was identified. The regression models could potentially predict who would benefit from the image enhancements if I predict a face perception (normal) score of 3.5 or greater from a

participant. This score will place that individual in the positive improvement group. Using the random forest model, which has an error of $0.14\pm$, the prediction is likely to be correct, considering the upper interquartile limit of the negative improver is 3.4. All from answering only nine questions.

The results are also similar in error (RMSE) to the regression analysis performed using functional visual data from the patients (4.2.6.2). There is a real potential to be able to use questionnaires to predict performance and provide recommendations. The significance of using a questionnaire is it can be performed almost anywhere, with minimal training, and at a low cost.

The limitation of these results is the sample sizes. Even with careful model training and validation, there is a chance that they will not generalise to a more significant population. However, the results are a good indication that this approach has merit. Future work extending the sample size would help to understand if this approach will work with the greater visually impaired population.

4.3 Discussion

This study aimed to understand if image enhancements improved face perception for the visually impaired.

The work carried out here has shown that there is a significant improvement in face recognition performance for some participants. Twenty-six participants (42%) showed an improvement with a type of enhancement. The majority of those with improvement were associated with Cartoon image enhancement. While there was a slight disconnect with the

performance of the different enhancements, the results show that when a participant gained an advantage with one enhancement, they did not have any negative impact on the other. Twelve participants (20%) had a significant negative improvement in face recognition with at least one of the enhancements. The remaining twenty-three (38%) participants had no measurable positive or negative improvement.

The criteria used to classify participants into the positive, neutral and negative improvement groups were derived from the variability of the test in a normally sighted cohort. It is likely that there is higher variability in a visually impaired population, as found with visual acuity tests (Patel, Chen, Rubin, & Tufail, 2008a).

For the majority of the participants, the image enhancements are not likely to provide any meaningful benefit to face perception. But for a minority, there is likely a significant benefit.

As the enhancements are similar in their approach, they can be generalised as contrast enhancement; it makes sense that contrast sensitivity correlates with performance. This is also observed with the spatial frequency analysis (figure 4.18). The Peli adaptive filter is an amplification of certain spatial frequencies, and no information is added or removed. The gain in the amplitude can be seen in figure 4.1.8. The effect of Cartoonisation on spatial frequencies is slightly different as the image processing used modifies the image significantly. A Gaussian blur is first applied that removes some higher spatial frequencies. Then a valley detector is applied, which focuses on useful facial features and creates the line image and combined with a binary thresholded image (D. E. Pearson & Robinson, 1985). The resulting image contains only black or white pixels, which has the effect of adding some high spatial frequencies (at borders and lines) and maximising all of the amplitudes. This amplitude maximisation is likely the reason for more participants gaining a benefit with the

Cartoonisation over the Peli adaptive filter, as those with lower visual acuity and reduced contrast sensitivity were more likely to benefit from the enhancements.

The control participants performed mostly the same with baseline and the enhancements, but sometimes better with the enhancements than without. Three participants performed worse, two with both enhancements. The participation numbers are quite small, so it is difficult to understand if the negative performances are intrinsic to the enhancements, testing conditions or participant attention. Performing further repeated tests are needed to be performed to understand the repeatability of these results. It should be noted that for the cartoon condition, some participants could perform better than expected due to some image artefacts. These artefacts effectively turned the task into a "spot the difference" game rather than a holistic face perception task. Thus, gains with the cartoon enhancement are possibly less 'real'. The control group's average age was lower than the patient group.

Face Perception Test

This study further validated the new face perception test. The test was well received by participants; the reaction times and duration of the test (for the baseline condition) were similar to the control groups (chapter 3). This study showed that the OVAT is useful in assessing visual enhancement devices in the visually impaired population.

There were a few failures; 3 participants were unable to perform the baseline test, and these individuals had low visual acuity (~ 1.0 logMAR). There is a limit to the maximum differences available within the stimulus used. Further refinement of the stimulus could be

performed to include these participants. This could be achieved by increasing the maximum morphing percentage used in the stimulus generation.

While tiredness was a factor, it should be noted that this was in a combination of the whole testing protocol, including multiple repeats of the test. However, for future testing, this should be taken into consideration. In isolation, the face perception test was efficient and not time-consuming and has the potential for use in the wider research community.

Factors relating to performance

The best factors that relate to performance through regression analysis are an individual's visual acuity and contrast sensitivity, supporting findings by Barnes, Taylor and Alexander (Alexander, 1988; Barnes, De l'Aune, et al., 2011; McCulloch et al., 2011; Taylor, Smith, et al., 2018). The number of participants and heterogeneity of the cohort make it difficult to make diagnosis-specific correlations. However, lower visual acuity and baseline face perception performance was a strong indicator that an individual would benefit from the image enhancements. Conversely, higher visual acuity and better baseline performance indicate a likely negative improvement. The cohort was recruited from an NHS clinic with as little bias as possible, other than the participant's willingness, thus representing the local population.

I also presented an analysis of questionnaire data and its potential predictive power. The results, while tentative with the small population, indicate that they can be used to predict the performance of the face perception test, and, therefore the usefulness of the image enhancements.

The criticism of previous work with image enhancement and face perception has been the subjective or poorly controlled assessment method (Leat & Mei, 2009; Leat et al., 2005; Mei & Leat, 2009; E Peli & Peli, 1984). None of these studies used a validated face perception test, and most results are stated in terms of preference (Leat et al., 2005; Mei & Leat, 2009) rather than any quantitative measure. The results and methods presented here are an improvement over these studies. Using the OVAT tool and the Cartoon enhancement, I have identified individuals who have a measurable improvement in face perception.

Overall, while the majority of participants (58%) did not gain a benefit from the image enhancements, a significant number of participants did. Thus, the image enhancements have potential value to a portion of the visually impaired population.

The next section presents the development and implementation of a real-time wearable face enhancement system.

Chapter 5:

Realtime Face Enhancement

My contributions to this chapter: I designed and wrote all of the software for real-time enhancement in C++. I designed and wrote all statistical testing tools and data analysis using Python and R. I performed all of the data collection with patients.

Other contributions: Caroline Justice aided with patient communication and room bookings. Professor Downes helped with patient identification. Professor MacLaren and Professor Downes assisted with proofreading.

Timeline: Software development took place from February 2019, taking approximately six months. Data collection was performed between July 2019 and September 2019.

5 Realtime Face Enhancement

As described in the previous chapter, image enhancements have the potential to improve a face's visibility for some visually impaired individuals. Not all participants benefitted, with only 42% showing an improvement. However, when extrapolated to the visually impaired population, a significant number of individuals could benefit from the ability to use the image enhancements in a real-life setting.

This chapter describes the development and implementation of a real-time face enhancement system.

5.1 Introduction

Several challenges must be overcome when creating a real-time face enhancement system. Globally applying the enhancements to an image has the potential to confuse by introducing clutter. Crowding is when object identification is impeded due to clutter (Wallace, Chung, & Tjan, 2017) and can potentially negate any advantage provided by the filter. Thus, the optimal application of the image enhancements would be best selectively applied to a face only. This introduces the primary challenge, automatic face segmentation.

Secondly, for a face enhancement system to be of any practical use, it must work in real-time. Real-time, in this context, means something that is virtually immediate; it happens immediately from the user's perspective. Considerable delays will become perceptually uncomfortable and limit any potential enhancements.

The target HMD is the Epson Moverio BT35E. [see appendix 2 for technical description]

5.1.1 Previous work

There is no published peer-reviewed research regarding the assessment of real-time face enhancements in use for the visually impaired at the time of writing. The field has several face-enhancement methods discussed in chapter 2, but none have presented a complete real-time solution. As stated in the introduction, several commercial products are reported to aid face recognition (NuEye, irisvision, esight); however, they provide only magnification and simple image filtering.

The most closely related work can be found in applications used for entertainment and social media, e.g. Snap Chat Filters (Snap Inc. Santa Monica, California, USA), where a user's face can be augmented for entertainment purposes on a mobile device. More complex solutions exist in the movie and computer game industries, offering real-time augmentation and character animation, typically requiring an expensive motion capture system, e.g., FaceWare (Faceware Technologies Inc. Austin, Texas, USA).

While examples of real-time face enhancement for the visually impaired may not exist in the current literature, the parts are commonplace.

5.1.2 Automatic Face Segmentation

Segmentation in computer vision is the process of separating a digital image into multiple segments. Typically used to simplify an image, highlight aspects, identify objects and boundaries. Faces follow a distinct form; they are mostly skin and have a relatively static structure. This information can be used to simplify the segmentation process.

Previous work in face segmentation has typically used the homogenous nature of faces to aid the segmentation process. Sobottka (Sobottka & Pitas, 1998) approached the problem by using colour and shape information. Colour information was defined in the HSV colour space and used fixed threshold values, calculated from Caucasian faces only. Garcia and Tziritas (Garcia & Tziritas, 1999) used a similar approach but employed two colour spaces, YCbCr and HSV, along with geometric constraints. Chai and Ngan (Chai & Ngan, 1999) used colour information only, using a fixed range in the YCrCb colour space, with contour extraction to aid segmentation. H Greenspan et al. (Greenspan, Goldberger, & Eshet, 2001) used a Gaussian mixture model (Dempster, Laird, & Rubin, 1977) to model the face-colour distribution. An approach called FaceSeg by H. Li et al. (H. Li, Ngan, & Liu, 2009) uses a face detector, filtered with skin colour YCbCr colour space; segmentation was performed using a graph cut method (Boykov & Jolly, 2001). The GrabCut segmentation method (Rother, Kolmogorov, & Blake, 2004) (based on graph cut) was used by L. Gui et al. (Gui, Lin, Wang, & Ding, 2011) for face segmentation; D.Khattab et al. used a modified GrabCut implementation (Khattab, Theobalt, Hussein, & Tolba, 2014). More recent approaches have employed CNNs. Jackson et al. (Jackson, Valstar, & Tzimiropoulos, 2016) developed a method to segment faces and facial features, guided by landmark detection. Nirkin et al. (Nirkin, Masi, Tuân, Hassner, & Medioni, 2018) designed an end-to-end face-swapping system with a landmark free face segmentation method.

While CNNs improve performance, they are computationally complex and require high-performance GPUs to achieve real-time speeds. The graph cut based methods (Gui et al., 2011; Khattab et al., 2014; H. Li et al., 2009) also provide good results but at the expense of computation time.

The non-CNN based skin-colour segmentation methods (Chai & Ngan, 1999; Garcia & Tziritas, 1999; Greenspan et al., 2001; H. Li & Ngan, 2008; H. Li et al., 2009; Sobottka & Pitas, 1998) all have the potential to achieve real-time performance and are suitable for this application. However, they all suffer from a similar failing; the use of skin colour models based on pre-existing datasets. These models are prone to bias depending on different skin tones, lighting and image capture methods within their training datasets.

The proposed approach uses a combination of face detection and skin detection methods to perform face segmentation. The following section introduces the two tasks of face and skin detection. They are assessed and discussed for suitability within this application.

5.1.3 Face Detection

Face detectors can be categorised into two types; feature-based and image-based. The feature-based approach tries to match features in an image against known facial features (e.g. eyes, nose)—the image-based methods use statistical tools to get the best match between training and testing images.

The most well-known feature-based approach is Viola-Jones (Viola & Jones, 2004) using simple features (weak classifiers) that are combined in a cascade architecture to create a more robust detector (strong classifier), typically using the AdaBoost algorithm (Freund & Schapire, 1997). The result is a computationally efficient face detector and has served as the basis for many modern face detection systems. The primary advanced is the speed of

detection and good accuracy. The disadvantages are the imperfect detection of non-frontal faces; faces view from the side have a high detection failure rate.

Image-based approaches include statistical methods like eigenvectors (Kirby & Sirovich, 1990), and principal component analysis (Sirovich & Kirby, 1987). More recent techniques use deep learning and convolutional neural networks (CNNs). Farfadi et al. (Farfadi, Saberian, & Li, 2015) developed a multi-view face-detector using CNNs. Liao et al. (S. Liao, Jain, & Li, 2016) proposed a method to address challenges in unconstrained face detection, such as pose variations and occlusions. The latter of the CNNs outperforms the Viola-Jones approach by a factor of 6 times and is much more robust to different face poses.

5.1.4 Face Tracking

While modern face detectors perform well under most conditions, detection failures are common. Failures can be caused by; occlusions (a hand in front of the face) and extreme pose changes (a subject turning away from the camera). A typical approach is to use an object tracker to support the face detector when it fails.

In computer vision, an object tracker is a tool that attempts to follow an object in a sequence of images. First, an object of interest is defined in an image, i.e. with a bounding box. Then the tracker tries to locate a visually similar object in the preceding frames. The field of object tracking in video images is too large to cover in total; only relevant approaches are discussed.

Previous work approaching the specific problem of face detection and tracking have used a variety of tracking approaches. A. Das (Das, Pukhrambam, & Saha, 2016) used an adaptation of the Lukas-Kanade tracker (Lucas & Kanade, 1981). The Lukas-Kanade is a well-established tracking method that identifies feature points in one image and matches them in another, generating a displacement vector. However, it suffers from drift, where the feature points can subtly change over time.

Paliy et al. (Paliy et al., 2011) and K.Huang (Huang & Trivedi, 2004) used a Kalman filter to perform the tracking. The Kalman filter is used to predict the coordinates of a face based on the history of known coordinates. Strictly the Kalman filter is not an object tracker. It provides an estimate of the state of a known system based on prior observations. This estimate can be updated periodically with measurements.

5.1.5 Skin Detection

Skin detection is the process of differentiating between the skin and non-skin regions in a digital image. The application of skin detection ranges from facial recognition, human body part detection and hand gesture analysis. The problem of skin detection is challenging, and a large amount of research has been dedicated to the area. There are several principal challenges with skin detection that must be overcome:

- Illumination variations; lighting changes
- Ethnicity
- Imaging conditions
- Complex background

A comprehensive survey of the current approaches to skin detection is covered by Naji et al. (Naji, Jalab, & Kareem, 2019). In the following section, strategies relevant to this application are discussed.

Skin detection can be broadly categorised into pixel-based and region-based methods.

5.1.5.1 Region-based

Region-based methods try to identify regions in an image with similar features without building a skin colour model. They rely on the fact that most objects have spatial connectivity. There are two principal approaches: region-growing and watershed. Region growing starts from a seed pixel in an image chosen by the user or prior information.

Neighbouring pixels are then included to grow the region based on a similarity to the seed; growth stops when a similarity threshold is met. The result is usually a good quality segmentation but is highly dependent on good seed pixel selection and is very computationally expensive (Cheng, Jiang, Sun, & Wang, 2001).

Watershed approaches use the watershed segmentation algorithm (Bieniek & Moga, 2000). The principle of the watershed algorithm is to treat a 2D image as a 3D surface or a topographic map, where the third dimension, or height, is the intensity of a pixel value. This map is then flooded with water, and the topological barriers, equate to segmentation boundaries in an image (Hajraoui & Sabri, 2014; Jiang, Yao, & Jiang, 2007; Yusuf, Mohamad, & Sufyanu, 2017). The limitation of the watershed approach is it usually results in an over-segmentation and is computationally expensive.

While the region base approaches perform well, they are too computationally expensive for consideration with this application.

5.1.5.2 *Pixel-Base*

The colour of skin depends on the concentration/activity of melanin and haemoglobin (Van Gemert, Jacques, Sterenborg, & Star, 1989). The distribution of skin colour across different ethnic groups is relatively compact (Cotton, 1996). Pixel-based methods attempt to assign a pixel in an image to be either skin or non-skin based on its colour without using any spatial information.

There are many different approaches to this problem, ranging from simple thresholding approaches to more complex statistical methods and artificial neural networks. A summary of the techniques used follows.

Thresholding Methods

Threshold-based methods are where rules are empirically constructed to define bounds (Chai & Ngan, 1999) in varying colour spaces that contain the colour of skin pixels. The advantage of the threshold methods is their simplicity; they are fast. However, they all assume the shape of the distributions, which negatively impacts their performance.

Parametric Statistical models

These approaches rely on the skin colour distribution that fits a known statistical distribution, i.e. a Gaussian Model. Several methods have used Gaussian Mixed Models to classify skin pixels. (Son Lam Phung, Bouzerdoun, & Chai, 2005)

Distance Methods - These approaches rely on the fact that similar colours appear closer to each other in the colour space. Suppose one defines a cluster of skin colours in a given colour space. A sample pixel is deemed inside this cluster and, therefore, a skin colour if the Mahalanobis distance to the cluster's centre is below a threshold value. (the Mahalanobis distance is used as the clusters are typically ellipsoid in shape). These approaches are simple but suffer from image artefacts and are computationally expensive (Terrillon, Shirazi, Fukamachi, & Akamatsu, 2000).

Statistical learning methods

A large number of tools exist that apply to the problem of pixel-based skin classification. e.g. Random forests (Khan, Hanbury, & Stoetinger, 2010) , Gaussian Naïve Bayes (Osman, Maarof, & Rohani, 2016) and Support Vector Machines (Han, Award, Sutherland, & Wu, 2006; Q. Zhu, Wu, Cheng, & Wu, 2004). All of these methods have shown good performance, and some evidence indicates that Random Forests are the best performers (Khan et al., 2010; Osman et al., 2016) when compared to other methods.

Artificial Neural Networks

A range of methods has been applied to skin classification (S. L. Phung, Chai, & Bouzerdoun, 2001)(Zuo, Fan, Blasch, & Ling, 2017) with good classification performance (Kim, Hwang, & Cho, 2017). However, they are computationally expensive and therefore not suitable for this application.

The statistical learning methods strike a balance between performance and computational complexity. They are straightforward to implement and available in many software libraries.

However, the variability in datasets and colour spaces used by the comparative studies (Khan, Hanbury, Stöttinger, & Bais, 2012; Osman et al., 2016; Son Lam Phung et al., 2005) makes it difficult to choose any particular method over another from the literature. For this real-time application, the decision of which to use is likely going to be driven by computational efficiency rather than classification performance.

5.1.5.3 Colour Spaces

For most of the aforementioned methods, many different colour spaces have been employed. Often as part of the development of the classifiers, several colour spaces are used. Chaves-Gonzalez et al. (Chaves-González, Vega-Rodríguez, Gómez-Pulido, & Sánchez-Pérez, 2010) compared ten colour spaces to determine which is best for skin detection. Using a k-means classifier, they found the HSV colour model to perform the best. Terrillon et al. (Terrillon et al., 2000) devised their colour space TSL, and in a comparison of several different colour models, using a neural network, they concluded TSL showed the best segmentation results. Al-Mohair et al. (Al-mohair, Mohamad-saleh, & Suandi, 2013) used a neural network classifier to compare colour models, concluding the TIQ colour space to be the best performer. A survey by Naji et al. (Naji et al., 2019) concluded; while there was no apparent single colour space to use, they believe the HSV colour space is superior.

In summary, there does not appear to be one colour space that consistently performs the best. Performance varies with classification techniques and datasets. Thus, any development of skin detection methods should include an assessment of multiple colour spaces.

5.1.6 Real-time performance

For this system to be successful, it must perform in real-time. To be more precise regarding the system's latency, the higher the latency, the less real-time it is. Latency is defined as the "photon to photon" delay, the time delay between a photon of light entering the camera and a photon of light entering the eye.

5.1.6.1 *How much latency is too much?*

Typically, designers of HMD systems aim for as low latencies as possible and rarely is there a defined acceptable value.

An ideal enhancement system would have a latency below that of human perception.

However, the amount of latency that can be perceived is context-dependent.

An experimental study by Stroud (Stroud, 1967) estimated that perceptual moments are about 100ms in duration. Stimuli, flashes of light presented within the same perceptual moment are perceived as occurring simultaneously. Efron (Efron, 1967) identified these perceptual movements to be 60 to 70ms. Work with head-mounted displays has shown that latencies of ~10ms are perceivable. These lower latencies are perceivable when more significant head movements are involved, and a static scene is more tolerant to higher latencies (Adelstein, Burns, Ellis, & Hill, 2005).

In the intended use case, enhancing face perception, it is less likely to be in a highly dynamic situation. The individual using the system is trying to identify a person either in front of them or nearby. Thus, the higher values of 60 to 70ms are an ideal target latency for this application.

5.1.6.2 Latency and head-mounted displays

There are several sources of latency within wearable displays, and they can be categorised as follows:

Image Capture delay

Image capture delay is the time is taken from a photon of light hitting the camera sensor, being digitised and sent to the processing unit as a package of binary information.

For most modern cameras, this will be defined mainly by the exposure time. The duration the camera sensor is capturing data. In more traditional camera terminology, this is the time between the camera aperture opening and closing.

Exposure time can be dynamic, depending on the lighting level (less light requires a more prolonged exposure). The maximum exposure time duration is defined by the camera's capture rate (how many frames per second it is collecting). E.g. the full exposure time for a 30fps camera is 33.3 milliseconds but could be as short as one millisecond.

Image Processing delay

This delay is defined as the time a captured image data package is received from the camera and is sent to the graphics processor. This will be the time it takes to perform all of the image processing.

Rendering Delay

This is measured as the time from when the graphics processor received an image to when it is drawn to the display. The factors involved here are post-processing steps performed by

the graphics rendering pipeline and the update rate of the display. e.g. typical computer monitors update rates range from 30 to 120Hz.

Display Delay

Display delay is the time from when a pixel update is requested from the graphics card to the time that pixel changes its displayed intensity by 50%, from its previous intensity to its new intensity. Display delay depends on the technology of the display hardware.

The hardware that I am targeting is the Epson Moverio BT-35E [Appendix 2.4], and I have limited control over the sources of latency. The camera has a dynamic exposure, with a maximum frame rate of 30Hz and the display hardware updates at 30Hz. The only user-controllable areas for optimisation is image processing and rendering delay.

Measurements with the target hardware have shown latencies of approximately 160ms. This is below the ideal latency of 60ms; however, this is a hardware limitation. Thus, it is imperative to ensure that no further increase of this latency is introduced.

5.2 Proposed System

The goal is to implement a real-time face enhancement system. Based on the review of existing methods and approaches, a combination of a face detector, tracker and pixel-based skin detector was chosen to achieve this. The target display hardware is the Epson BT-35E wearable display (appendix 5.1) controlled using a laptop computer (appendix 2.4.2).



Figure 5-1 Epson Moverio BT-25E headset

5.2.1 Face detection

The choice of face detector was driven by availability. The selection was made from open source libraries, with assessments made on performance, accuracy and speed. I chose to use the OpenCV DNN method; a description of the candidates can be found in appendix 2.7.

5.2.2 Face tracking

Choice of object tracker is also limited by availability. The tracker chosen is the Minimum Output Sum of Squared Error (MOSSE) OpenCV tracker based on Bole et al. (Bolme, Beveridge, Draper, & Lui, 2010). This tracker is robust to lighting scale and pose and operates at a high framerate, ideal for this application.

The tracker is used primarily when the face detector fails; however, it can also be used as an optimisation. Face detectors are computationally expensive compared to trackers. Running

the face detector at a lower frequency, e.g. 10Hz rather than 30Hz, and using the tracker to fill in the missed frames. This may be a requirement on lower power mobile hardware, where avoiding high computation levels is desirable to increase battery life.

The approach uses the detect and track principle, employing a Kalman filter to stabilise the output.

Kalman Filter

A Kalman filter is a simple but powerful tool. Simply, it is an algorithm that predicts a system's state over time using a series of measurements. If the distribution and magnitude of the noise (inaccuracy) of the measurements are known, the algorithm can produce estimates that are more accurate than a single measure alone.

Several steps perform estimation; firstly, the filter produces estimates of the current state variables, using a physical model and a measure of noise. Secondly, once the outcome of the subsequent measurement is received, the estimates can be updated. In the update, more weight is given to measurements of higher certainty. The algorithm is recursive and can run in real-time.

In the use case of face tracking, the system state is the pixel coordinates of a face's location in an image. The measurements are the output from the face detector (which has a high certainty) and the tracker's output (which has a lower certainty). Using a Kalman filter allows for an efficient combination of the two measurements.

Parameter selection.

Implementation of the Kalman filter requires estimation of some uncertainties; *Process Noise* and *Measurement Noise*.

Process Noise: Decides the accuracy and time lag in the estimated value. Higher values indicate that the higher gain, the more weight to the noisy measurements, and the estimation accuracy compromised. Lower values mean better estimation accuracy, and time lag may be introduced in the estimated value. Process noise is typically, and was with this use case, experimentally chosen with trial and error.

Measurement Noise: Represents electronic, random noise characteristics of the sensor or measurements. Lower values indicate more trust in the measurements, and higher values less trust. As with process noise, the measurement noise was chosen experimentally. The method I used was to point the camera at a static photograph. The face detector was then run on multiple captured images, and the resulting variance of the detector was used as a baseline for optimisation.

The combination of face detection and tracking provides a smooth, reliable localisation of a face.

5.2.3 Face Tracking persistence

Having chosen a face detection method and tracking technique, the face tracking system requires some control logic for the selected application. The reason for tracking is to handle failures in face detection, primarily from occlusion. With potentially multiple faces in a

scene, any implementation must gracefully handle faces leaving the scene permanently and new ones entering.

For each frame provided by the cameras, the face detector will list areas it detected a face. It gives no knowledge of the identity of the face. Successful tracking of multiple faces requires a tracker to be matched with each unique face.

Matching is performed in two steps, location and similarity. Location is where in the image the face was detected; if there is enough overlap, it can be assumed the face is a correct match to the tracker. This will be true for transient occlusions such as the target face turning away.

As the target system is wearable, the camera is attached to the wearer's head pose; relying on position alone for matching would be problematic. If the camera wearer moves their head, and therefore the camera, this position logic will fail to match. For this scenario, face similarity is used. There are numerous methods for determining face similarity, which falls into the area of face recognition. Simple similarity checks can be used as full recognition is not required. The aim is to differentiate one face from others within the scene. Histogram comparison methods are suitable in this case, and the Chi-Squared method is employed.

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

Where H_1 and H_2 are the two histograms, I is the value of the bin within the histogram. If computation cost was not a concern, more sophisticated techniques could be employed.

The FaceNet (Schroff et al., 2015) method discussed in chapter 3 would provide a more

robust matching, certainly in a complex dynamic scene, but this level of accuracy is not warranted.

Finally, a temporal limit is set on the tracker. If no face is matched to a tracker after some time, the tracker is invalidated.

Figure 5.2 presents an information flow chart for the face detection and tracking system.

The input is an image captured from a camera, and this image is passed to the face detector.

If a face is found and there are no valid trackers, a new tracker is created, and the face's region of interest (ROI) is returned. If there are valid trackers, they are assessed for a match.

If the detected face is deemed part of an existing tracker, that tracker is updated; if not, a new tracker is created. If initially no face is found by the face detector and there are valid trackers, the region of interest of that tracker is returned.

A valid tracker is defined by; if a face has been successfully associated with this tracker within the last 200ms. This validation allows for failures of the face detector, limited to a short period.

Face Detect and Track

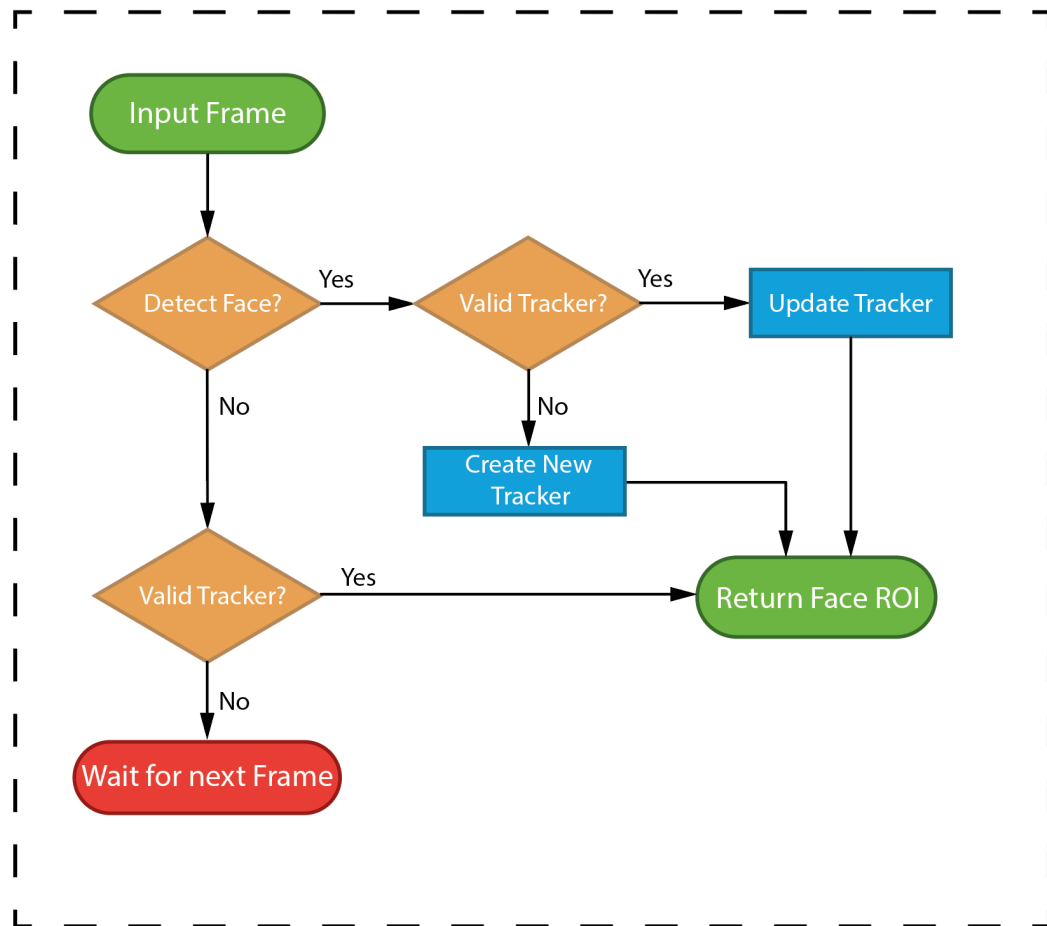


Figure 5-2 Face detect and track flowchart.

5.2.4 Skin Detection

Choosing the skin detection method is more complex compared to the face detector. The statistical learning methods provided the best speed/accuracy trade-off; however, they can suffer from training bias. An adaptive learning approach is used to overcome training bias. The principle uses the face detector's results to generate training data for a classifier, which continuously learns as the system runs. This has several advantages; it removes the requirement for an extensive and diverse training dataset, avoids skin tone bias, and avoids biases introduced by the camera.

Several approaches are empirically tested to decide which statistical learning classification method to use. This included methods known to perform well with skin classification (i.e. Random Forests) and other well-known classifiers. They were quantitatively and qualitatively assessed for classification performance, accuracy, and execution speed.

5.2.4.1 *Statistical Learning Methods*

Statistical Learning refers to a set of tools for understanding data. They can be classed as supervised or unsupervised. Supervised methods build statistical models from known input data for predicting output behaviour from an unknown input. Unsupervised models use no supervised output, no labelling, and describe the ability to derive relationships and structure from the data.

In this case, I am using supervised learning methods, the input is a colour pixel from an image, and the output is a classification of skin or not-skin. The model is trained using known data, which are pre-labelled pixels (skin or non-skin). This model is then used to classify or predict if a new unknown pixel is skin or not.

Four different classifiers with five colour spaces are assessed with existing skin datasets.

Decision Trees

Decision trees (James et al., 2013e) are a nonparametric supervised learning tool for regression or classification problems. They use a tree-like structure consisting of splitting rules, breaking data down into smaller and smaller subsets. The trees consist of nodes, where the top node is the root or the best predictor; subsequent nodes are called leaf nodes; each node can have two or more leaf nodes connected by branches. The advantage of decision trees, and hence their inclusion, is speed; they are fast to train and predict. The disadvantage is they suffer from high variance and are prone to overfit data.

Random Forest

As the name suggests, random forests (Tin Kam Ho, 1995) are a collection of decision trees developed to overcome the problems of high variance and overfitting with decision trees. This is performed firstly by bootstrapping (James, Witten, Hastie, & Tibshirani, 2013d) the training data (n) times, training (n) trees and aggregating the results. Secondly, when each split is considered, a random sample of predictors is chosen as split candidates out of the complete set of predictors. The product is less biased towards any single predictor, thus reducing the overall variance and likelihood of overfitting. The downside is a sacrifice of speed for improved classification performance.

Logistic Regression

Logistic regression (James, Witten, Hastie, & Tibshirani, 2013a) is a variation of linear regression, rather than modelling the response directly; it models the probability of belonging to a specific category. Logistic regression is a very efficient, simple, and widely used technique. The disadvantages are also that of linear regression; it cannot be used in nonparametric problems.

Gaussian Naïve Bayes

Naïve Bayes (Hastie, Tibshirani, & Friedman, 2009a) is a family of probabilistic classifiers based on Bayes' theorem, assuming that all the features are independent. The Gaussian variant assumes the features are normally distributed when working with continuous data. The advantages of Naïve Bayes methods are their simplicity and do not require large amounts of training data. Their disadvantage is their assumption of independence of predictors.

5.2.4.2 Training Data

As supervised learning strategies are applied so, some labelled training data is required. Several datasets are available; A. Lumini (Lumini & Nanni, 2018) provides an overview of their scale and quality. Two datasets were chosen; Pratheepan (Tan, Chan, Yogarajah, & Condell, 2012) and ECU (Son Lam Phung et al., 2005). These two were selected due to their high quality, representation of faces, and availability. Table 5.1 summarises the data.

Table 5-1 Labelled face data sets summary

Dataset	Files	Pixel count	Ratio of non-skin to skin	Balanced Pixel count
ECU	4000	1,110,708,015	81:19	418,783,838
Pratheepan	76	18,415,073	83:17	3,118,018

The balanced pixel count is the number of pixels in total when accounting for an even distribution of non-skin to skin pixels (50:50).

5.2.4.3 Methods

A test framework was implemented in python, using the Scikit-learn (Varoquaux et al., 2015) machine learning python API and OpenCV (Itseez, 2015).

As the training images used are real-world images, the ratio of the types of pixels, i.e. skin to non-skin, is heavily weighted to the non-skin, with an average of approximately 80:20. This imbalance needs to be addressed, and a subsample of the overrepresented class is taken, leaving a 50:50 ratio (Haibo He & Garcia, 2009).

All datasets were combined and then split into training and test sets, using a ratio of 70:30. The four classification methods were then trained and tested using five different colour spaces.

The performance of the classifiers is reported by precision and recall. Precision is the fraction of correct classifications out of all the returned results. Recall is the fraction of correct classifications out of all the results that should have been returned.

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN}$$

Where TP = true positive, FP = false positive and FN = false negative. To simplify comparisons, I also use the f-score, which combines precision and recall into a single value:

$$fscore = 2 * \frac{precision * recall}{precision + recall}$$

5.2.4.4 Results

Each classifier was trained on the combined training dataset (approximately 2800 files, almost 300 million pixels) and tested on the test data set (about 1200 files, around 100 million pixels). The f-score was reported for each file, which ranges from 0.0 to 1.0, with 1.0 a perfect performance where all pixels were labelled correctly.

This process was repeated for each colour space; the training and test data were converted to the respective colour space and repeated. This produced 20 individual classification results for comparison.

The overall classification performance, reported in f-score for each classifier and colour space, is presented in figure 5.3.

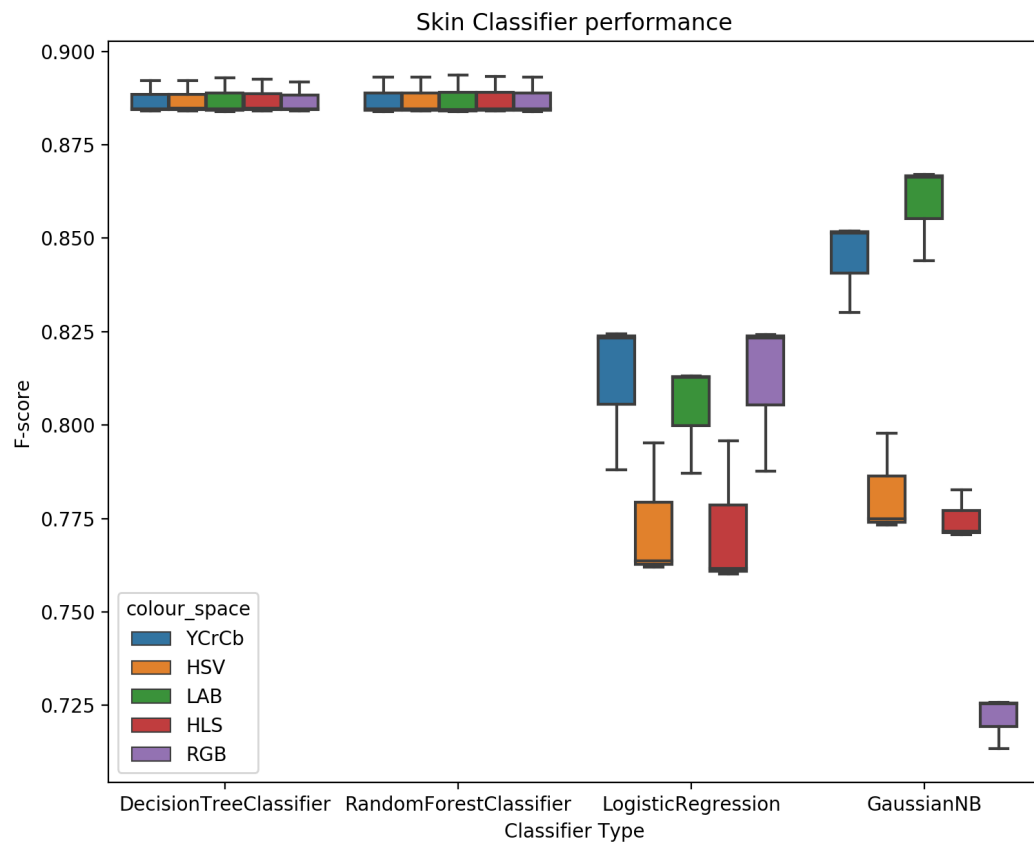


Figure 5-3 Skin classifier results – The plot shows four classifiers (grouped on the x-axis) and the five colour spaces. The colour of the box plots represents the colour space used. E.g. the left hand five boxes are the results for the colour spaces used with the Decision Tree classifier. The upper and lower boundaries of the coloured boxes are the interquartile range, and the error bars are the limits at 1.5 times the Interquartile range.

Table 5.2 presents the results of the best performing colour space for each classifier.

The Random Forest and Decision Tree classifiers performed consistently well across colour spaces with median f-scores of 0.88. The logistic regression was more variable, performing best with the YCrCb colour space (f-score=0.82). Gaussian Naïve Bayes performed best with the LAB colour space (f-score =0.87).

Table 5-2 Skin classification results – best scores for each colour space. AUC = area under the curve. The training time is the total time taken to train the classifier. The results are from classifiers trained and tested on the combined datasets.

Skin Classification Results - best scores						
Classifier Type	AUC	Training Time (seconds)	Colour Space	f-score	Precision	Recall
RandomForestClassifier	0.95	1043.8	LAB	0.88	0.88	0.88
DecisionTreeClassifier	0.95	528.9	LAB	0.88	0.88	0.88
LogisticRegression	0.87	403.9	YCrCb	0.82	0.82	0.82
GaussianNB	0.94	24.7	LAB	0.87	0.87	0.87

The training duration and prediction times are presented in *Table 5.2*. For most classifiers, the training time increases with the number of samples, with the Random Forest taking the longest time to train and Gaussian Naïve Bayes the shortest to train.

Figure 5.4 presents a random sample of images from the ECU dataset (Son Lam Phung et al., 2005). The classifiers perform well with high levels of recall (skin was usually always labelled correctly). The failings have high levels of false positives in scenes that contain colours similar to skin colour; note the bride's flowers and the baby's chair.



Figure 5-4 A sample of images from the ECU dataset (Son Lam Phung et al., 2005) and the output from the classifiers. Columns left to right are; Original colour image, Ground truth labels (white=skin, black=non-skin), Decision Tree labels, Random Forest labels, Logistic Regression labels and Gaussian Naïve Bayes labels.

The best performing classifiers were Random Forrest and Decision tree, with little to distinguish between them, and they were not significantly impacted by choice of colour space. When assessed by speed, the fastest was the Gaussian Naïve Bayes, by a significant margin. The Gaussian Naïve Bayes approach also had good classification performance with the Lab colour space.

The Decision Tree classifier is the best when considering speed (computation time) and performance (f-score), with Gaussian Naïve Bayes providing a balance between speed and performance.

5.2.4.4.1 Real-world factors

Real-world environments add factors that are not necessarily expressed within most datasets. Different cameras and lighting conditions add a significant variability, and there is a risk with relying on classifiers trained offline on poorly representative datasets.

Figure 5.5 presents three images of the same subject taken with three different cameras of varying quality. The variability in colour and brightness can be seen in the original pictures.

When one of the trained skin classifiers is applied (Random Forest - Lab colour space), this variability is also expressed in the resulting skin mask.

Camera Comparision

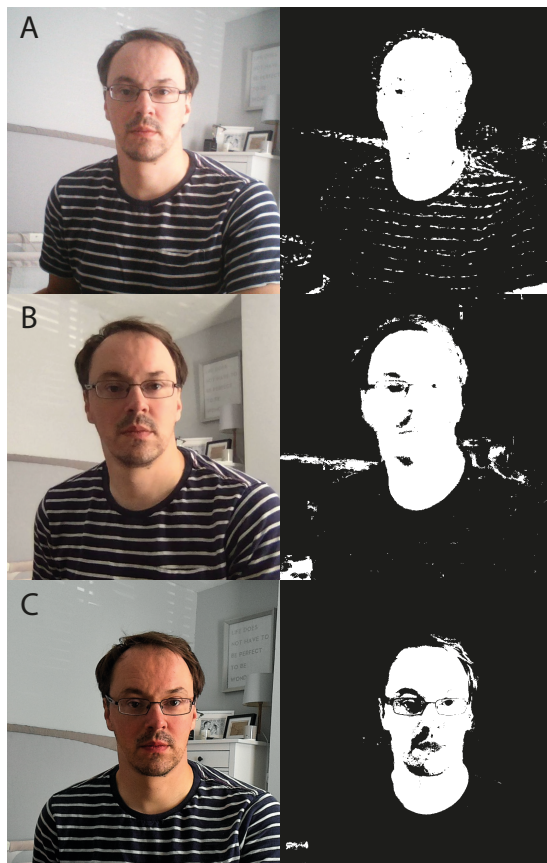


Figure 5-5 Camera Comparison of three cameras, original image left column, and the skin classification, right column (Random Forest – Lab colour space). Rows are labelled; A = Epson Bt-35E camera, B = Apple MacBook pro webcam, C = Econ Systems See3Cam_130

A solution could be to colour-calibrate each camera, assuming one had knowledge of the camera used within the dataset (which is rare). The calibration process is time-consuming and must be performed for each different type of camera used and perhaps every instance of each camera if manufacturing variances are large. This still does not account for lighting variances. Some of which can be negated using a different colour space (one that separates the luminance from colour, e.g. HSV), but this is only effective for small lighting changes.

5.2.4.5 Online learning

An alternate approach is to use an online learning method where a system generates its own training data, inspired by Zhu et al. (Q. Zhu & Wu, 2004). Training data is generated from the scene being imaged and can reduce the impact of cameras and lighting variances.

Simply, the method uses the output of the face detector along with a slower, more robust segmentation method to generate data for training a faster skin classifier. The slower method does not need to be calculated on every frame. It can be run periodically, continually updating the faster pixel-based classifier, which will be run on every frame.

The slower GrabCut algorithm (Rother et al., 2004) has shown good performance with face segmentation (Gui et al., 2011; Khattab et al., 2014) and is ideal for more robust segmentation. The GrabCut algorithm works by the user defining a region of interest containing the target object. The algorithm then estimates the colour distribution of the foreground and background using a Gaussian Mixture Model. Next, a graph is created from the distributions over each pixel, where the connections prefer pixels with the same label (from the mixture model). Finally, a min-cut algorithm (Kolmogorov & Zabih, 2002) is used

to segment the graph. The results can be further improved by manually labelling areas in the image as foreground and background, providing the mixture model with more information, resulting in an improved segmentation.

In this context, the output of the face detector provides the initial region of interest. To perform the manual labelling, the output of the face detector is not accurate enough alone, as only a bounding box containing a face is provided. A facial landmark detector (Kazemi & Sullivan, 2014) is used to generate labels for the foreground(face) and background(non-face) labels. The implementation in the DLib API (King, 2009) can detect the location of 64 features in approximately one millisecond, enabling accurate foreground and background labels to be generated. Thus, improving the GrabCut segmentation, producing robust, albeit slow, face segmentation results.

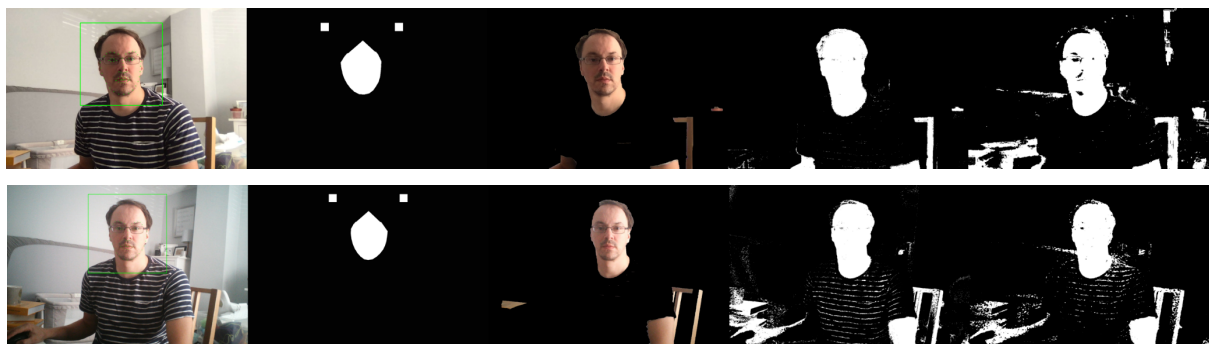


Figure 5-6 The online learning process. Columns: (First from Left) – original image with a green box produced by the face detector. (Second) – the masks used for labelling, the small squares are non-face labels, and the centre mask is for face labels. (Third) - results from GrabCut showing segmented skin. (Fourth) – skin segmented using Random Forest trained with GrabCut data. (Fifth) – skin segmented using Random Forest trained with offline data. Rows represent different cameras: (top) Econ Systems See3Cam_130, (bottom) – Epson BT300.

Figure 5.6 visualises the online learning process and compares the results to an offline trained classifier. The performance of the online trained classifier has a qualitatively improved performance over the offline method. The face area is well labelled, with little

background noise. The GrabCut method does select some incorrect areas (the chair in Figure 5.6) and is reflected in the classifier results. However, these artefacts can be easily removed. A final filter is applied that removes any labels too far from the face.

I further validated the approach against the ECU dataset, comparing the results to the offline trained classifiers. Figure 5.7 shows the results of this comparison. The online method outperforms the offline for all classifiers tested.

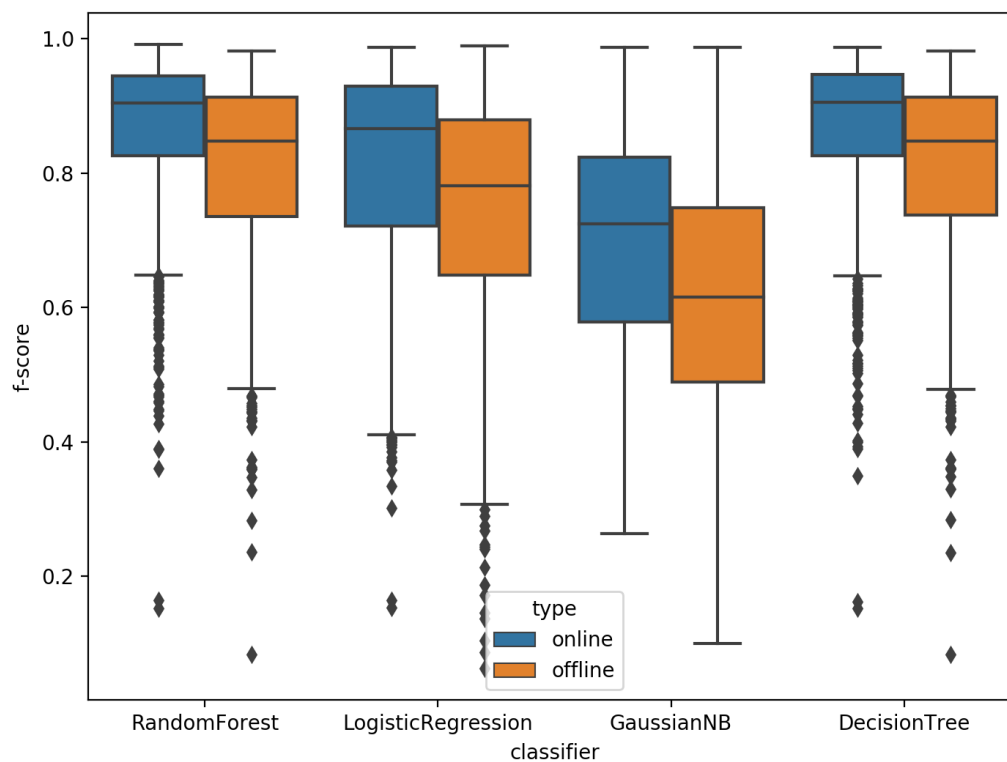


Figure 5-7 shows the f-scores (y-axis) of the online (blue) vs offline (orange) for each classifier (x-axis) applied to a sample of 1700 images from the ECU dataset.

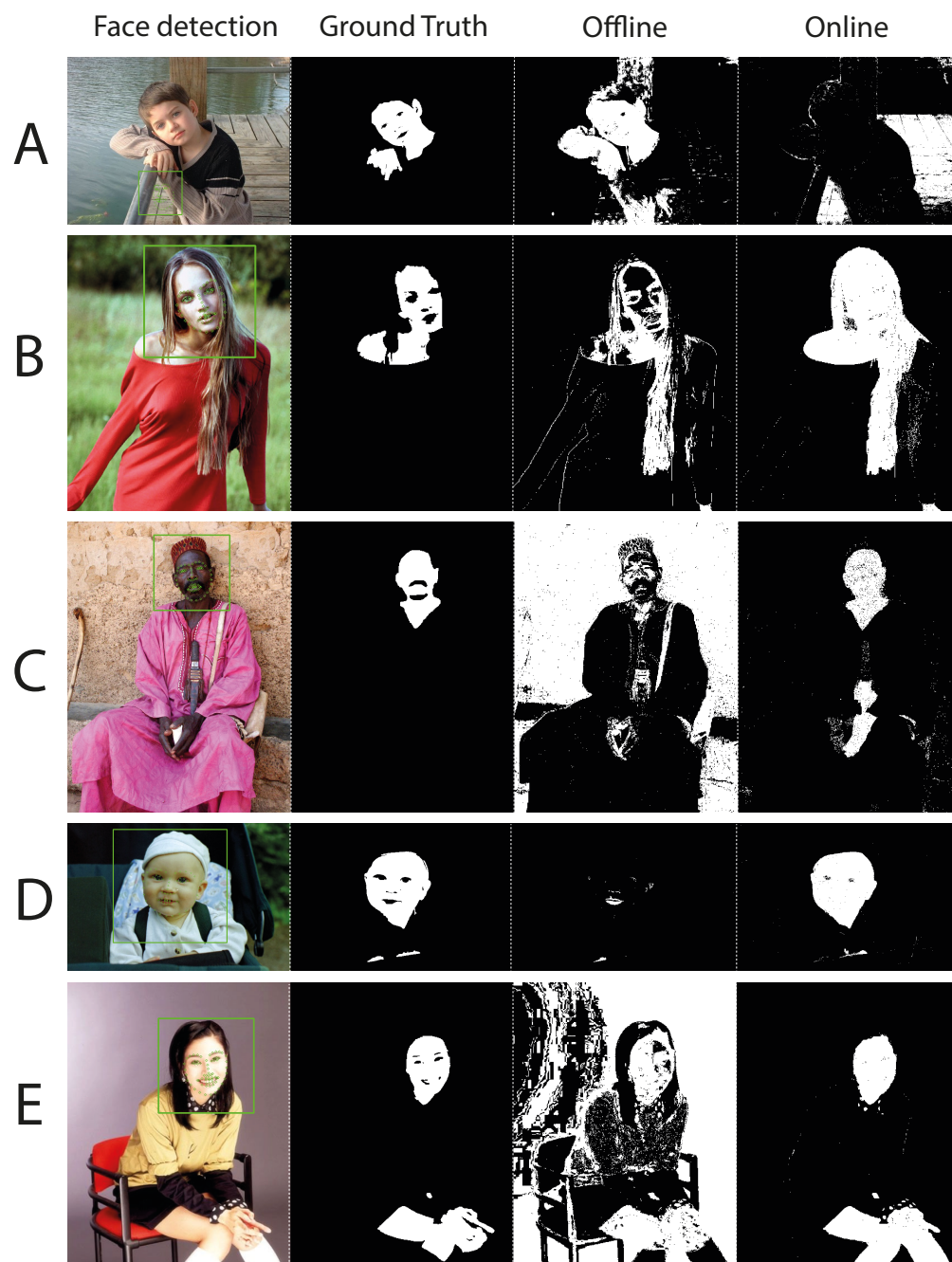


Figure 5-8 Examples of segmentation failures. Columns; Original image (overlaid face detector results), Ground Truth, Offline Segmentation, Online Segmentation. Rows A & B present failure of the online classifier A is a failure in the face detector, and B the hair has been included. Rows C to E present failures with the offline classifier. C shows a skin colour not represented in the offline dataset. D shows poor camera colour balance, and E shows the background colour too similar to skin.

Figure 5.8 presents examples that show when the offline method fails and the online succeeds. The advantage of the online training method is an improvement in classification performance, avoiding dataset biases (figure 5.8. row C & E) and being more resilient to

lighting and colour invariances (figure 5.8 row D). The drawback to the online method is the reliance on the GrabCut algorithm. If this performs poorly, the classifier will be trained with incorrect data (figure 5.8 row A). Labelling the GrabCut method with the output of the face detector can cause artefacts, i.e. the inclusion of hair (figure 5.8 row B).

In the presented results, the scores are calculated from all skin within the dataset. The online method is training just on the face area, leading the segmentation to bias away from non-face skin. However, as the application is focused on the segmentation of faces, this is not of concern.

A significant strength of this technique is with video, utilising several keyframes (figure 5.9), I can create a robust skin classifier that, once trained, will perform in real-time.

5.2.4.6 Data quality

The data quality used during training is vital for the performance of the skin classification. When generating the training data, two factors need to be considered in maintaining data quality: the face detector output and the Grab-Cut segmentation.

5.2.4.6.1 Face Detector Quality

False-positive face detections can negatively impact quality (Figure 5.12), especially when only a few keyframes are used. The images used as keyframes are validated using an offline trained skin classifier. If the number of detected skin pixels is below a threshold, then the image is rejected. A threshold value of 40% was experimentally found to reject the majority

of false positives. While there are biases in the offline classifier, using it as a validation step did not have any significant negative impact.



Figure 5-9 Example of good quality face keyframes segmented with GrabCut



Figure 5-10 - Example of mixed quality face keyframes segmented with GrabCut. Notice the non-face keyframes (a portion of a door in the room).

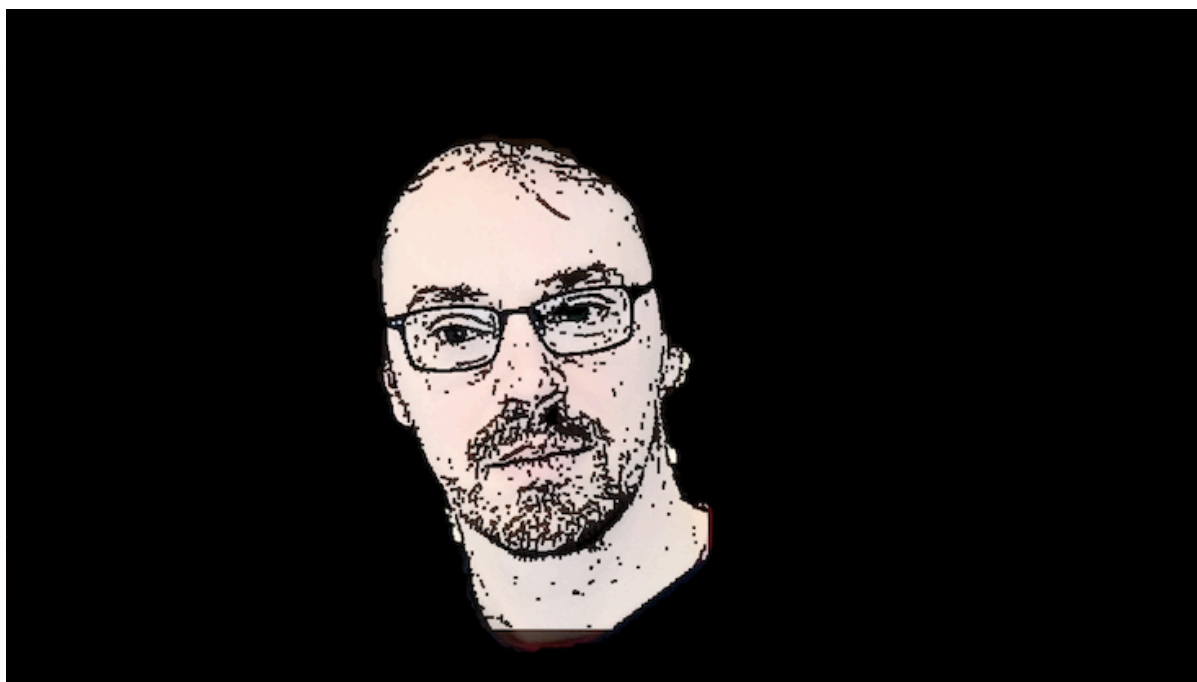


Figure 5-11 Example face mask created with good quality keyframes (figure 5.9) – with Cartoon enhancement applied.



Figure 5-12 Example face mask with mixed quality keyframes (figure 5.10) – with Cartoon enhancement applied – notice the holes in the mask on the left portion of the image.

5.2.4.6.2 GrabCut segmentation quality

GrabCut segmentation, while generally good, does not always produce an ideal segmentation. To minimise the number of bad keyframes where the GrabCut segmentation has performed poorly, I can take advantage of the relatively static shape of a face. The shape of the result face masks can be used to provide a measure of success.

A collection of 225 faces were manually labelled as 'good' or 'bad' keyframe candidates. 90 (42%) were labelled 'bad'. Figure 5.13 shows an averaged mask for each label; a colour map was applied to highlight the shape. It can be seen that the good labels have a more consistent uniform shape.

The approach is to train a classifier to distinguish if a mask was of 'good' or 'bad' quality based on information about its shape. Four shape metrics were calculated for describing the shape of the masks.

1. Circularity = $\frac{4 \pi Area}{(perimeter)^2}$ The more circular the mask the closer to 1.
2. Convexity = $\frac{Area}{Convex\ Hull\ Area}$ The convex hull is the smallest polygon that encapsulated the mask, with no corners bent inward. The closer to 1, the more convex the mask.
3. Inertia ratio = a measure of how elongated the mask is. This is calculated by fitting an ellipse to the mask and calculating the minor and major axis ratio. A circle will have a value of 1 and a line of 0.
4. Centroid = coordinates of the centre of mass of the mask.

The labelled data and calculated shape metrics were used to train and compare several binary classifiers; Logistic Regression, Gaussian Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbour.

The classifiers were optimised to maximise precision, as the goal is to reduce the number of bad keyframes, the false positive rate. False negatives are less of a concern; throwing away a good keyframe will not impact the classifier, but including a bad one will.

$$\text{Precision} = \frac{\sum TP}{\sum TP + FP}, \text{ TP} = \text{True Positive, FP} = \text{False positive.}$$

All of the classifiers tested had a precision of greater than 0.8, with Gaussian Naïve Bayes producing a precision of 0.92 and a runtime of less than one millisecond.

The occurrence of the bad keyframes during usage was roughly 5%. Utilising the mask shape classifier can reduce this significantly further, removing approximately 90% of the bad keyframes without substantially adding to the execution time.

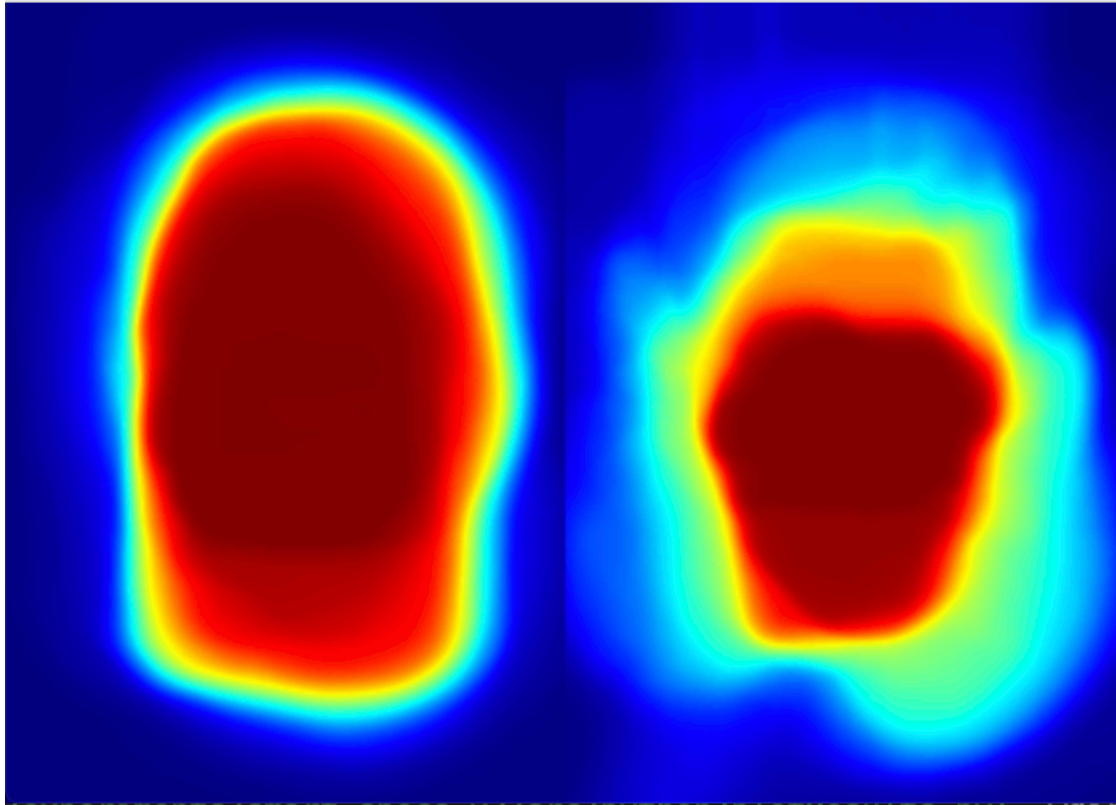


Figure 5-13 Mask comparison of 'good' (left) and 'bad' (right) grab cut segmentation of a face. All labelled masks' pixels were averaged and normalised, with a colour lookup applied; blue = low occurrence, red = high occurrence.

5.2.4.7 Implementation overview

The proposed method of skin detection relies on continuous online training. Figure 5.14 presents a flow chart the provides an overview of the method. When a face is detected, the cropped face image is sent to the skin training module. The skin training module manages the data generation and training of the skin classifier.

Each frame that is passed to the skin training module is referred to as a keyframe. For a successful classifier, a large amount of varied training data is required. This must be balanced with the time taken to train. The more training data, the better the classifier, but the longer it takes to train. Thus, choosing the number and size of keyframes is essential for a satisfactory user experience.

The strategy chosen is to define a minimum and a maximum number of keyframes. This was selected as two frames for the minimum and ten for the maximum through experimentation. The frame size was set to be 50% of the captured image (960x540), a maximum of 5,184,000 pixels for training. However, the actual number of pixels used for training is typically lower; the keyframe is generated from only the face containing part of the image captured by the camera.

The minimum keyframe number enables fast initialisation; the skin detection method is useable in approximately 2 seconds. As time progresses, more keyframes are added, and the training time increases to a maximum of roughly 10 seconds. While the initial classifier used is trained on a smaller number of keyframes, it is useable. As time progresses and more keyframes are added, the performance of the classifier improves.

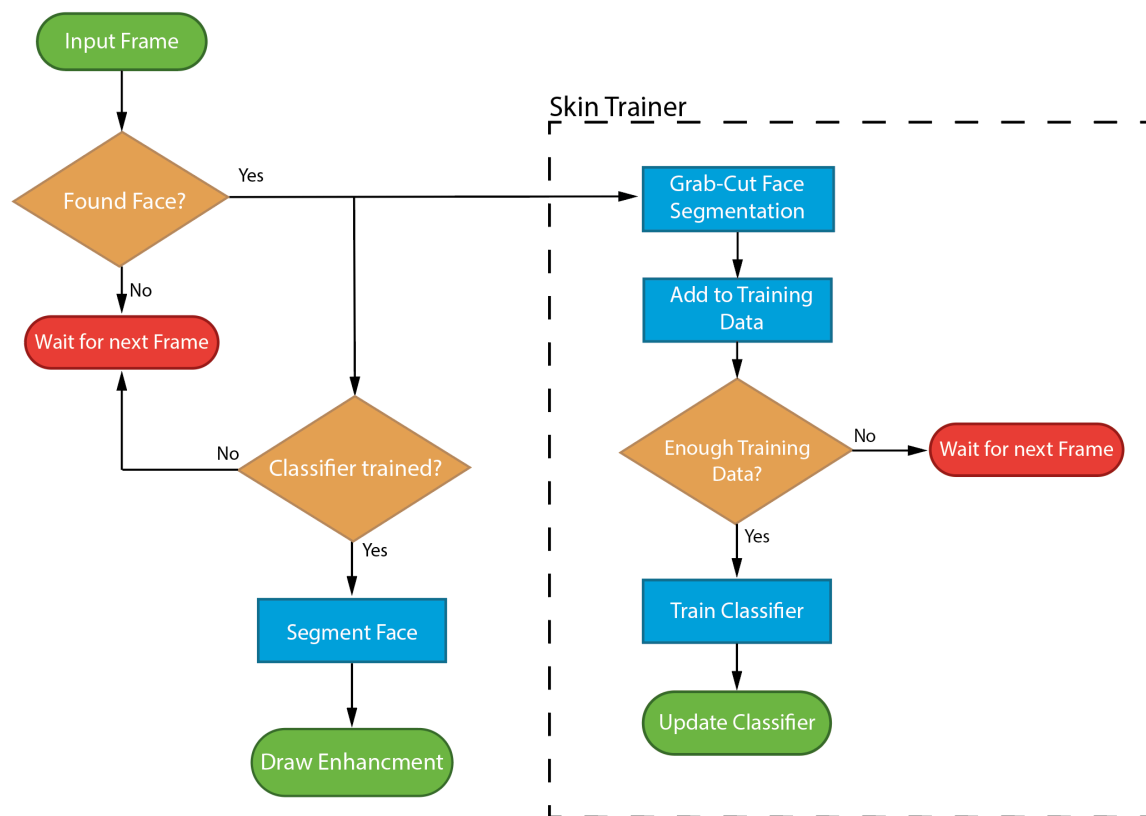


Figure 5-14 Processing flow chart for the skin segmentation method.

5.2.4.8 Execution time

The skin classification method presented shows good performance in terms of classification accuracy, and therefore face segmentation. It must also ensure that its execution time meets the real-time requirements (33ms). The components must be arranged in a multithreaded manner to achieve this execution time, and further optimisations are required.

Multithreading is a feature of modern operating systems, where it provides the ability to process multiple operations concurrently. If the face segmentation method is executed consecutively in a single thread, this will take between 1.5 and 4.5 seconds. Table 5.3 lists

the maximum duration all the components take to execute. While all of these components are not run on every frame, the Face Detection and Skin Detection must be, and their combined 2.5 seconds is far from the target of 33ms

Table 5-3 Speed of execution

Component	Maximum Duration
Face Detection	30
GrabCut Segmentation	500
Classifier Training	2500
Skin Detection	1400

All values are milliseconds

The thread organisation can be seen in Figure 5.14; the left-hand side of the flowchart is the main thread run on every frame, the right-hand side within the dotted rectangle is the background processing thread.

The skin detection is performed by passing each pixel in an image through the skin classifier. This is a computationally expensive task with approximately 5 million pixels needing classification, so it needs to be optimised further.

There is a finite range of colours that the camera can capture and digitise; about 16 million, each pixel is represented by three 8bit numbers (0->255). The classification for each of these values can be precomputed, stored in a lookup table, and used instead of running the classifier. This reduces the skin detection time from greater than 1 second to less than one millisecond.

The final method takes approximately 31 milliseconds to complete, within the target time of 33 milliseconds (Apple MacBook Pro, 15-inch 2017). There is no perceptual lag in the user experience.

The resulting skin mask generated by the implemented system allows for simple selective application of the image enhancements. There is a delay of approximately 1 second before the enhancement is applied to the segmented face from initialising the enhancement.

5.2.5 Real-time image enhancement system

During development, visually impaired users were shown versions of the system and asked for their opinion. The main feedback was the low quality of the integrated camera. Image degradation when zooming diminished any benefit. The integrated camera within the Epson BT series (figure 5-1) has no optical zoom or focus. Also, the video resolution is lower than the specification's state. The sensor can capture still images at 1920x1080; however, it is only able to capture video in 1024x768.

Also noted by participants was the camera offset, which is mounted on one of the display arms, which confuses when trying to direct the camera and focus on a person's face.

A new camera module was integrated with the displays, which is capable of 4096x2160 at 30hz and mounted with the camera sensor aligned over the bridge of the nose [details in appendix 2]. An automatic face tracking feature was added. This uses the face detector and tracker to localise a target face, aiding the user to keep the face within view.

The final enhancement system has the following features:

- 1) Apply cartoon enhancement to any face within a scene. (figure 5.16)
- 2) Remove background, leaving only the face. (figures 5.17 & 5.18)
- 3) Zoom in / out control
- 4) Automatic face tracking (figure 5.19)

All of these features can be combined, enabled or disabled as the user desires. Following are a selection of screen captures from the real-time face segmentation tool. Although these images are taken from the processing laptop, they are representative of what the wearer sees. The only difference is the displays are see-through, so any black areas will not project any light, and the user will see through to the real world. This frequently occurs when background removal is applied (figures 5.17 & 5.18).



Figure 5-15 Original unprocessed image – screen capture from processing laptop



Figure 5-16 Cartoon enhancement overlay, blended with the underlying image

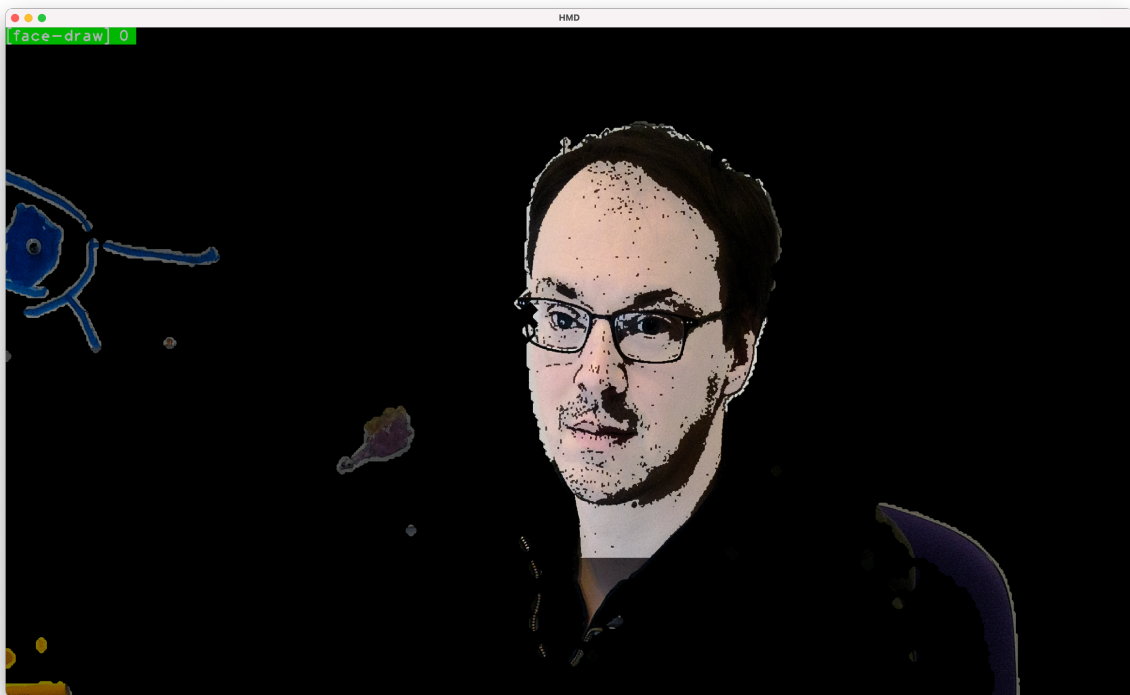


Figure 5-17 Background removed using a segmented face mask. Cartoon enhancement is blended with the underlying image/ Note the artefacts in the background as the raw mask is used (figure 5.18 uses a cleaned mask).

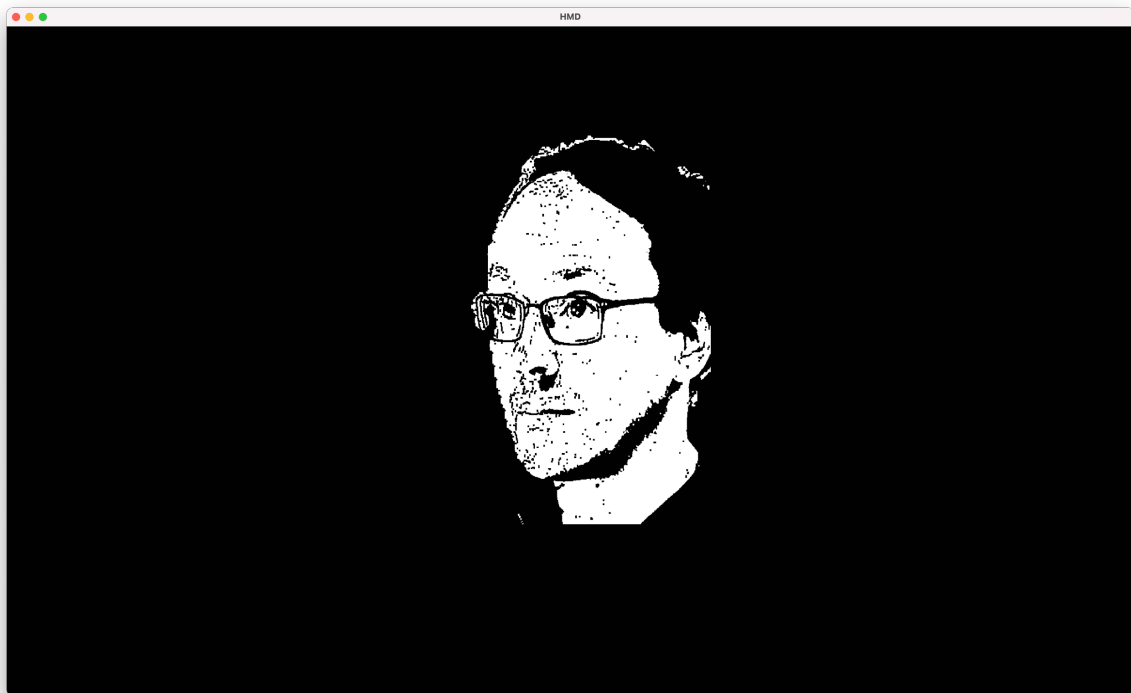


Figure 5-18 Background removed and cleaned, with cartoon enhancement applied (no blending of underlying image is used)

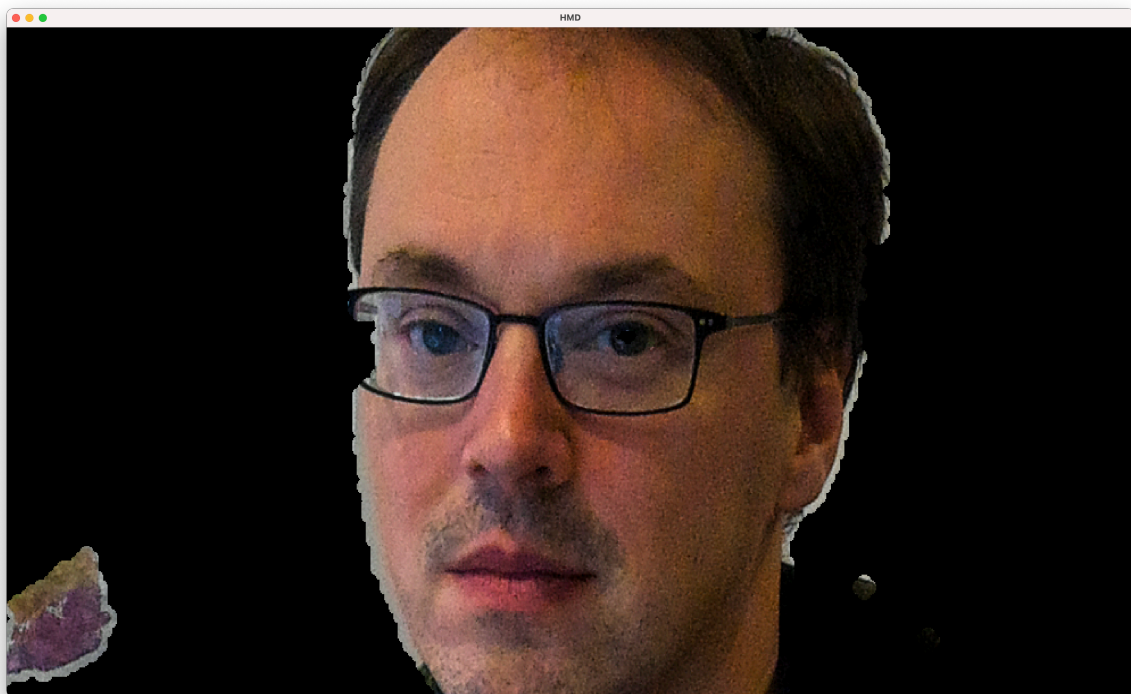


Figure 5-19 Face tracking enabled along with the background removal - the zoom and crop is fully automated and will follow the face around the image.

5.3 Discussion

The primary goal was to develop and build a real-time face enhancement system. This was achieved by combining and optimising existing methods (face detection and tracking) and the development of a novel technique (online skin learning). All of which was integrated in a considered manner that achieved real-time performance. The system achieved the performance goals, and aspects were well received by visually impaired participants. To the author's knowledge, no such system has been developed prior to this to perform face enhancement on a wearable display in real-time.

The system does have some limitations. The skin segmentation is sensitive to bad keyframes. Methods were developed to reduce the occurrence of the bad keyframes, but they are not 100% effective. However, the impact is limited, as the online learning method uses a circular buffer, and bad keyframes will be removed after a short period of time.

The current implementation uses a laptop computer, which, while mobile in principle, is not practical. The laptop chosen is from late 2017; with modern computing's progression still following Moore's law², it is not unreasonable to expect similar performance on state-of-the-art mobile phones. Modern smartphones are increasingly containing hardware that is dedicated to computer vision and deep learning, such as the Intel Myriad X³, and Apple Computing's Neural Engine⁴

² Moore's law states that the number of transistors that can be placed on an integrated circuit doubles every two years.

³ <https://www.intel.co.uk/content/www/uk/en/products/processors/movidius-vpu.html>

⁴ <https://www.apple.com/uk/newsroom/2020/11/apple-unleashes-m1/>

The image enhancement tested was the Cartoonisation method, but the system is not limited to this. The core of the implementation is a real-time face segmentation tool that produces a binary mask. This mask can be used for any other form of image enhancement or relevant investigation. The modular design presents a useful tool for future research into alternate image enhancement methods.

5.3.1 Further work and improvements

The online learning skin segmentation method has the ability to learn more about the user's environment. The current implementation only uses data captured within the current session. This can be extended to persist of multiple sessions and provide higher quality and more relevant statistical model. A log of all the keyframes collected during use is currently made, implemented for testing and debugging reasons, but this data can be used for higher quality training and refinement of the statistical models used for skin detection.

The best-performing methods for face segmentation use convolutional neural nets (deep learning). They were rejected for this application due to the computational expense; however, their use will become viable with advances in processing power. Typically, the training side of deep learning is computationally costly relative to the inference. If the same methodology is used with online training, this training can be offloaded to a cloud computing service where the inference is performed on the local device.

5.4 Experimental results

5.4.1 Introduction

Previous work using similar Epson displays had shown an objective benefit for people with peripheral vision loss (Joram Jacob Van Rheede et al., 2016). During this study, some individuals with central vision loss did show an objective benefit, but most did not. The poor performance was hypothesised to be the limited displays field of view and their incompatibility with central vision loss. Although the study had a significant mobility focus, central vision loss sufferers typically do not have significant mobility problems.

This experiment investigates the Epson BT350 wearable displays usefulness primarily for face recognition and is the limited field of view a factor. There is a secondary investigation into how the real-time face enhancement system performs.

Due to the nature of the testing environments and clinical setting, it was decided to perform qualitative tests as a semi-structured discussion. While a successful face perception test has been developed (chapter 3), it was difficult to use with the current displays. The camera and the computer display interacted negatively. The camera's auto-gain feature (standard on most modern cameras) struggled with the significant lighting differences of the computer monitor and room. While it was possible to use the smart glasses with the computer display, it took significantly more time to perform the face perception test, making it unrealistic to use.

5.4.2 Materials and Methods

Two versions of the smart-glasses were used:

- 1) A video passthrough version using the Epson BT350 built-in camera (see appendix 2.1 for technical details).
- 2) Secondly, a modified real-time version was used, consisting of a high-resolution video camera mounted centrally on the glasses' bridge (see appendix 2.2).

Fifty-three visually impaired participants tried the video passthrough version, and the participant could control the zoom level using the supplied handset (appendix 2.1). The participant cohort was the same as described in chapter 3.

Participants were introduced to the wearable displays and provided instructions on how to control the zoom feature. They were allowed 5 minutes to familiarise themselves with the displays. A semi-structured discussion was then performed with the experimenter; they were asked the following questions.

- 1) Do you like the smart-glasses? Would they be useful in their daily life?
- 2) What did you think of the display size?
- 3) How is viewing a face with the glasses (better with or without)?
- 4) Any other comments?

A selection of participants also tried the new real-time enhancement system; it was explained and introduced, then it was discussed with the experimenter.

Participants were asked if they liked each mode:

- 1) Normal
- 2) Segmented Normal
- 3) Cartoon-Face – with background
- 4) Cartoon-Face – without background
- 5) Any other comments?

All the modes were available with face tracking, which was enabled / disabled to demonstrate its effectiveness.

Participants were allowed to wear any prescription spectacles, where possible. The wearable displays can be worn on top of most spectacles.

5.4.3 Results

5.4.3.1 Passthrough version

Fifty-three participants tried the video passthrough version of the glasses. Participants were seated in a testing room in the Oxford Eye Hospital, where they were able to visually explore their environment.

Do you like the smart-glasses?

Thirty of them said they liked them (57%), and all find that viewing a face was significantly better using the displays than with their normal vision.

What do you think of the display size?

Fifteen participants commented that they would like a larger display. Of those, five stated that they did not like the glasses as a whole, and only two participants indicated that the display's size was a determining factor. Both had visual acuity of 1.6 logMAR, stating that they had to view the display peripherally.

SP013 (RE 1.6, LE LP) Male 36, Stargardt.

SP033 (RE 1.5, LE CF) Female 63, Cone-Rod dystrophy.

The commonality with the two participants was their lack of vision in one eye, which perhaps contributed.

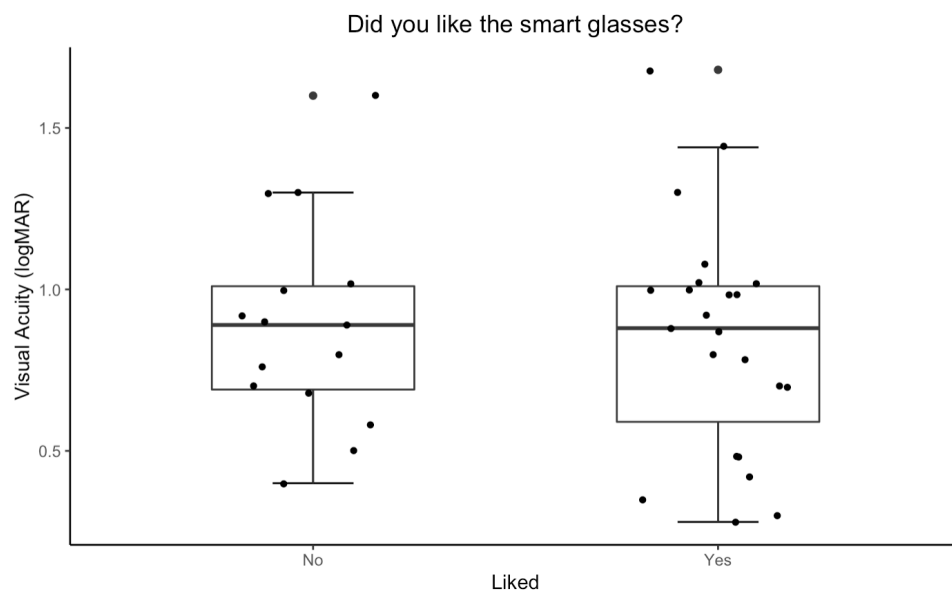


Figure 5-20 boxplot of visual acuity vs a participant liking the smart glasses

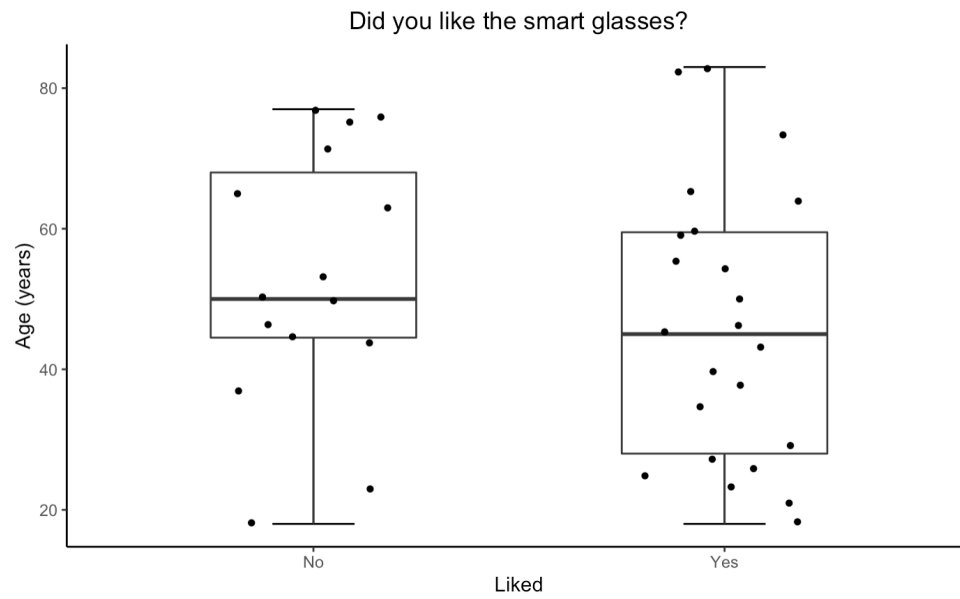


Figure 5-21 2 boxplot of age vs a participant liking the smart glasses

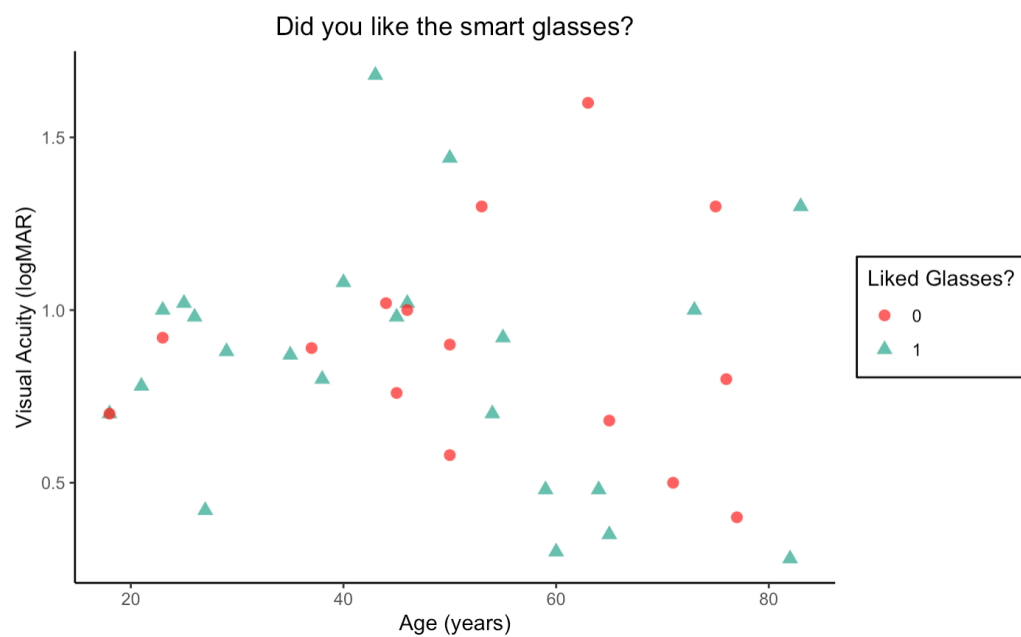


Figure 5-22 Scatter plot showing the relationship between age and visual acuity to preference. Red circles represent patients who did not like the smart glasses; green triangles are those who did.

5.4.3.1.1 Factors relating to performance

A wealth of information was collected from the participants. Is it possible to identify from this collection any factors that are predictive of their stated preference?

Liking the Smart Glasses

VA or age did not have any significant impact on preference (figures 5.20, 5.21 and 5.22).

Further investigation was performed, including all of the information collected about the participants (chapter 4). This included contrast sensitivity, colour sensitivity, reading acuity, gender and primary diagnosis.

A step-wise logistic regression was performed using AIC as an optimisation parameter. The best predictor was when a subject had a decrease in contrast sensitivity at two cycles per degree; they were more likely to like the smart glasses. This had an accuracy of 68%, showing a sensitivity of 0.3 and a specificity of 0.96.

Larger Display

Logistic regressions also show that a reduction in contrast sensitivity at 2 CPD was the strongest predictor of wanting a larger display. ~ 72% accuracy; Sensitivity of 0.96 and specificity of 0.2.

The difference of sensitivity and specificity implies that the "Liking the Smart Glasses" model is more reliable at predicting who will not like the smart glasses than who will. Conversely, the "Larger Display" model is better at predicting who will want a larger display than not. This result is likely indicative of the small population. However, what is clear is that the loss of lower spatial frequency does have a role to play.

5.4.3.1.2 Other comments

The participants' discussion was semi-structured, and they were free to raise any comments and make observations enabling qualitative assessment of the device. Several common themes were presented.

Forward viewing

Several participants commented that they found themselves viewing straight forward instead of using their peripheral vision. These were individuals with significant central scotomas and typically viewed objects peripherally. They commented that it could be socially awkward when interacting with people unfamiliar with their visual impairment when they peripherally view someone during a conversation. So, they found this aspect of the device very helpful.

Display location

Three participants commented that they would like the display presented in their periphery rather than centrally. All three participants had significant vision loss ($VA > 1.0$) and significant central scotomas.

Camera Offset

The camera is attached to the glasses frame's side, which means that the user must move their head to change their gaze direction. While most adapted quickly to this, some found it difficult, notably more so with elderly participants. The effect was inversely related to distance, with closer objects having a larger perceived offset.

Latency:

One participant reported significant motion sickness; although they commented that they are susceptible to motion sickness, the glasses amplified it.

Another participant commented on the latency, the delay introduced from the camera and processing. This individual had a hearing impairment and found lip-reading difficult with introduced latency. Otherwise, the high latency was not commented on.

5.4.3.2 Real-time enhancement version

Ten participants tried the RT system, and these participants also tried the passthrough version.

Cartoon Face

Only one participant liked the cartoon enhancement; the rest of the participants disliked the presentation of faces in this manner.

Background removal

Three participants did like the masking, removal of the background, finding it helpful with removing clutter. This was without the cartoon enhancement.

Face Tracking

Three participants liked the face tracking feature; this significantly helped them follow a face in the scene, where they had previously stated difficulty.

New camera.

The new camera was very positively received. The higher resolution and quality significantly improved the experience—the higher resolution allowed for much higher zoom levels. The camera's location was an improvement; even if the participant found the offset camera acceptable or did not negatively comment, they commented on how much improved it was.

5.4.4 Discussion

The passthrough version of the smart glasses was generally positively received. With 57% of participants finding view faces easier than with their unaided vision. There was no evident trend when comparing age and visual acuity. Age has been shown to be negatively correlated with technology acceptance (Hauk, Hüffmeier, & Krumm, 2018) but the observations here do not support this hypothesis. The lack of correlation may be due to the ease of use and direct potential benefit. The negative correlations shown by Hauk et al. (Hauk et al., 2018) were related to perceived usefulness and perceived ease of use.

The logistic regression results indicated a correlation with contrast sensitivity loss at two cycles per degree to liking the smart glasses. This was the same factor that had predictive power of performance in the face perception test (4.2.6.1). This makes sense, as the smart glasses were presented as a tool for improving face perception. The participant's positive evaluation of the device is likely to have been based on how it improved face perception. While there were no enhancements applied in the passthrough version, the auto-gain feature of the camera would have increased the brightness, and zoom increasing the size of the viewed face, making it easier to see.

The size of the display (the field of view) was not the limiting factor it was thought to be, with the vast majority of participants comfortable with the size of the image. Only two participants stated it was a significantly negative factor, both of whom had light perception only in one eye. However, the Epson headset used is a 'one size fits all' design; thus, the physical shape of the display could have contributed to this. It can be challenging for individuals with significantly smaller heads than the average to correctly align their eyes with the display.

A significant increase in the display size would also raise the risk of increasing the visual-vestibular conflict. The latency of both displays used was not insignificant (measured to be approximately 160ms). The current displays have a diagonal field of view of 23 degrees, leaving significant space in the periphery where the real world can be seen. Reducing this free space would likely increase the number of participants with induced motion sickness and the smart glasses' overall usefulness. Only one out of 53 reported any motion sickness.

The latency of the displays was a concern as it is a hardware limitation. There were very few comments from the participants regarding this and was likely due to the static nature of the test. Testing it in a dynamic environment, for example, in walking participants, would provide more information. However, this was not in the approved protocol due to safety concerns in this population with the testing environment available. Of note, a participant with a hearing impairment highlighted the latency when trying to lip-read. This indicated the glasses' potential usefulness in assisting with lip-reading.

The most commented-on limiting factor with the passthrough version was the camera. The low resolution limited the usefulness. When higher zoom levels were used for viewing distant faces, the image became highly pixelated. This was likely to have impacted and limited the acceptance of the smart glasses for several participants.

The real-time enhancements system performed well; however, the enhancements were not well received. This is likely due to the unnatural presentation of the cartoon enhancements. It has been shown in chapter 4 that the cartoon enhancement does improve face perception in some participants. Participants only experienced the enhancements in a static clinical setting. It may be the case that some will find that the enhancements aid them in everyday life with extended external use. The background removal was found to be helpful by a few participants, along with face tracking.

The new camera's inclusion was significant; all participants noted the higher resolution, new location (over the bridge of the nose), and image quality. The improved resolution allowed for higher zoom levels. However, this had a slightly negative impact as with increasing zoom levels; the image was affected by any head movement, making it very difficult to maintain a stable image. In this situation, the face tracking feature was crucial.

A limitation of the real-time enhancement testing was the number of participants. The technical development was performed throughout the study, and the decision making was driven by visually impaired participant feedback—notably, the improved camera and face tracking system. The methodology used is likely positively biased, the metrics I used are subjective and participants could be answering to please the experimenter. However, the

primary aim was to provide a technical solution, including features that overcame the simpler passthrough version's limitations. Aspects of the system showed promise, and the principles derived from this will inform future work.

All participants within the face perception study (chapter 4) tried the passthrough version. Their interaction and discussions helped guide this further development. The inclusion of the new, centrally aligned camera was patient-led, along with the automatic face tracking method.

5.5 Conclusion

The aim of this chapter was to be able to apply image enhancements selectively to a face in real-time. I have presented a novel method of real-time face segmentation and used it in conjunction with image enhancements.

The visually impaired participants did not enjoy the image enhancements, but they did find the features of the face segmentation system an improvement over the existing method.

Chapter 6:

Visual Acuity with a head-mounted Display

My contributions to this chapter: I designed and wrote all software for the visual acuity test in Java. I designed and wrote all statistical testing tools and data analysis using Python and R. I performed all of the data collection with patients.

Other contributions: Caroline Justice aided with patient communication and room bookings. Professor Downes helped with patient identification. Professor MacLaren and Professor Downes assisted with proofreading.

Timeline: Data collection was performed between July 2019 and September 2019. Software development and data analysis took one month.

6 Visual Acuity with a Head-mounted Display

This chapter investigates using a head-mounted display to perform visual acuity tests on a visually impaired population.

6.1 Introduction

Patients who participated in the face perception study (chapter 3) tried the Epson BT350 displays. The integrated displays are of high quality, designed to match standard visual acuity, where one pixel optically subtends 5 minutes of arc at 1 meter (hardware details are described in Appendix 2.1).

The primary goal is to understand if the displays of the Epson BT350 are suitable for a reading study, and how visually impaired participants can view text on the wearable display. This is implemented as a visual acuity test for direct comparison with a standard ophthalmic examination.

6.2 Methods

6.2.1 Binocular Visual Acuity

Best-corrected binocular visual acuity (BCVA) was measured using the Revised 2000 Series Early Treatment for Diabetic Retinopathy Standard (ETDRS) letter reading eye (Precision Vision, La Salle, Illinois, USA). BCVA was calculated as the lowest complete row read, subtracting 0.02 log units for every letter correctly identified beyond that point.

6.2.2 Smart Glasses Visual Acuity

A software application was developed for participants to perform visual acuity testing whilst wearing smart glasses, herein referred to as SGVA.

Digital visual acuity tests are valuable tools that have an established performance (Beck et al., 2003)(Bach, 2007) and work well with the visually impaired population (Jolly et al., 2020). Advantages of digital tests include the speed of execution and reliability.

The use of these tests with participants wearing a display is novel and comes with added challenges. Firstly, the experimenter is unable to see what the participant is viewing. This is important as the experimenter must know which letters the participant is attempting to read. Secondly, the calibration of the display is complex. Knowing the physical font size is required to calculate the visual angle and, in turn, estimate visual acuity.

6.2.2.1 Smart Glasses Visual Acuity Test Design

The design goal is to replicate the ETDRS chart behaviour in a digital format. The Epson BT300 series devices use the Android operating system on the control unit. An Android app was developed using the Java programming language and the Android Studio software IDE (Google Inc, California, USA).

The font used on the ETDRS chart is described as the Sloan font. A version of the font is publicly available; <https://github.com/denispelli/Eye-Chart-Fonts/>.

The displays are in a landscape orientation; the maximum font size is limited by the height of the display, which is 720 pixels.

The Android operations system provides information of the display density, dots per inch.

The density of the Epson BT350 display is 87.397, converting to mm is 3.44 dots per mm.

When rendering text to the display, the font size can be set by pixels. With the height of one Snellen M equal to 1.454 mm (see appendix 3.1 for calculations), a conversion can be performed from Snellen M to pixels.

$$px = M * 1.454 * p_{mm}$$

Where px is the height of a font in pixels, M = Snellen M and p_{mm} pixels per mm.

The display can show a maximum Snellen M of 144; however, this fills the entire height of the display. The maximum value is limited in the software to a Snellen M of 100; this allowed for a comfortable border around the font and is equivalent to a logMAR of 2.0.

At the largest M size, only one character can be displayed; this increases as the M value decreases, stopping at a maximum of 5 characters. Table 6.1 shows the relationship of M to the maximum number of displayable characters. The lowest M value displayable is 1, equal to 0.0 logMAR. The largest M is 100 and equivalent to a logMAR of 2.0.

Table 6-1 Character and acuity limits of the BT-350 display

Maximum M	Number of characters	LogMAR
27	5	1.4
36	4	1.6
51	3	1.7
85	2	1.9
100	1	2.0

M is the Snellen font size.

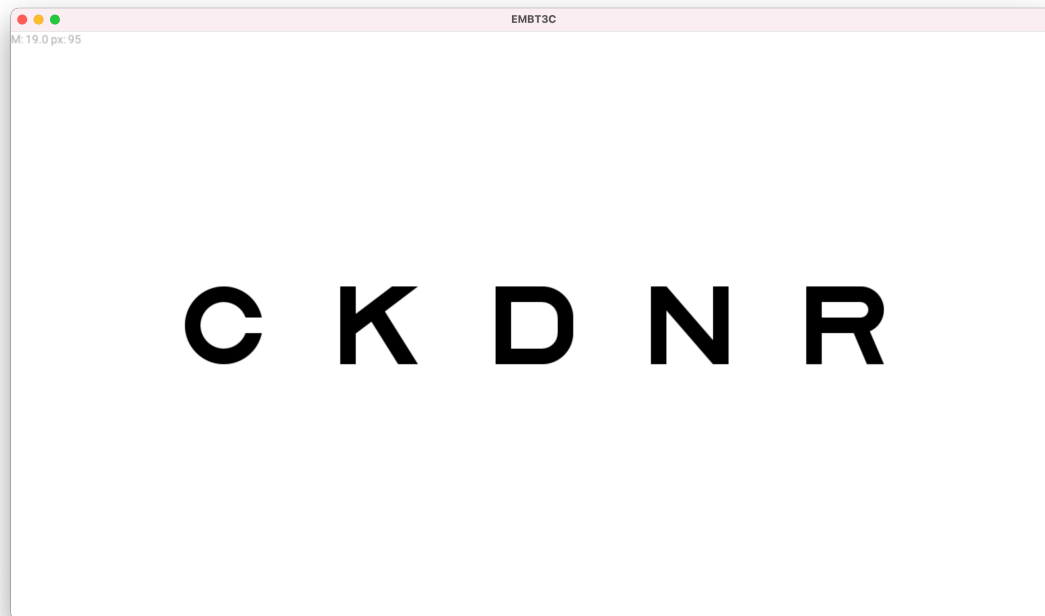


Figure 6-1 Screen capture from the Epson BRT350 display. Snellen M = 19.

6.2.2.2 Remote Control

Control of the display app was via remote control. This was implemented as an Android application running on the experimenter's mobile phone.

The app searched for the SGVA app running on the glasses, connecting wirelessly. The experimenter could change the font size (in Snellen M units) and the displayed text (Figure 6.2).

The experimental protocol allowed for cycling through a list of lines of optotypes. The order of M and characters was identical to the ETDRS charts.

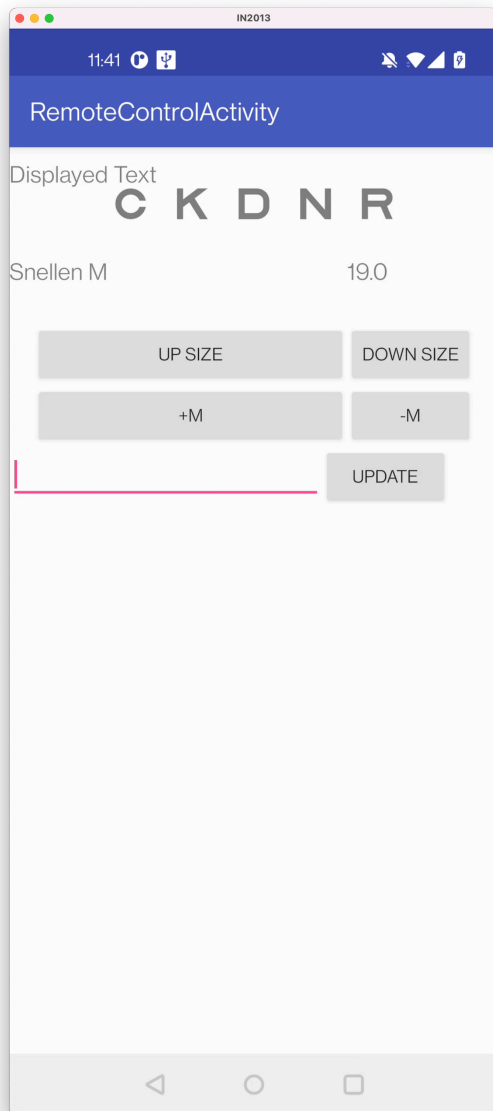


Figure 6-2 Screen capture from the remote-control application. The "UP SIZE" and "DOWN SIZE" buttons controlled which ETDRS line was displayed, automatically changing the M size. Manual control of the M is set with the "+M" and "-M" buttons.

The accompanying remote-control application allowed the experimenter to control and view a single line of Sloan letters, which was presented to the participant binocularly. The visual angle of the rendered text was calculated using a calibrated external camera. A description of this process and visual angle calculation is described in appendix 3.2.

6.3 Results

Fifty visual impaired participants performed the two visual acuity tests.

The two tests' average scores were: BCVA= Median 0.9 IQR 0.48, and the SGVA Mean 0.83 IQR 0.32. The difference in medians, 0.07 logMAR less than one line of letters. A Wilcoxon rank-sum test provided a p-value of 0.78, accepting the null hypothesis that the two groups' distributions are equal.

Linear regression produces a model of; intercept=0.071 and slope= 0.93, $R^2=0.69$. The model was assessed for assumptions of homogeneity of variance, linearity and the residuals were approximately normally distributed. Spearman's rank correlation coefficient rho was 0.84, indicating a good level of agreement.

An interclass correlation coefficient between the two measurements was 0.83 (95% CI 0.74- >0.89).

Figure 6.3 shows a Bland-Altman plot comparing the two measurements. The mean difference is -0.011 logMAR. There are a few individuals with significant differences. Positive points on the Y-axis indicate participants who performed better (better VA) with the smart glasses; negative on the Y-axis they performed worse.

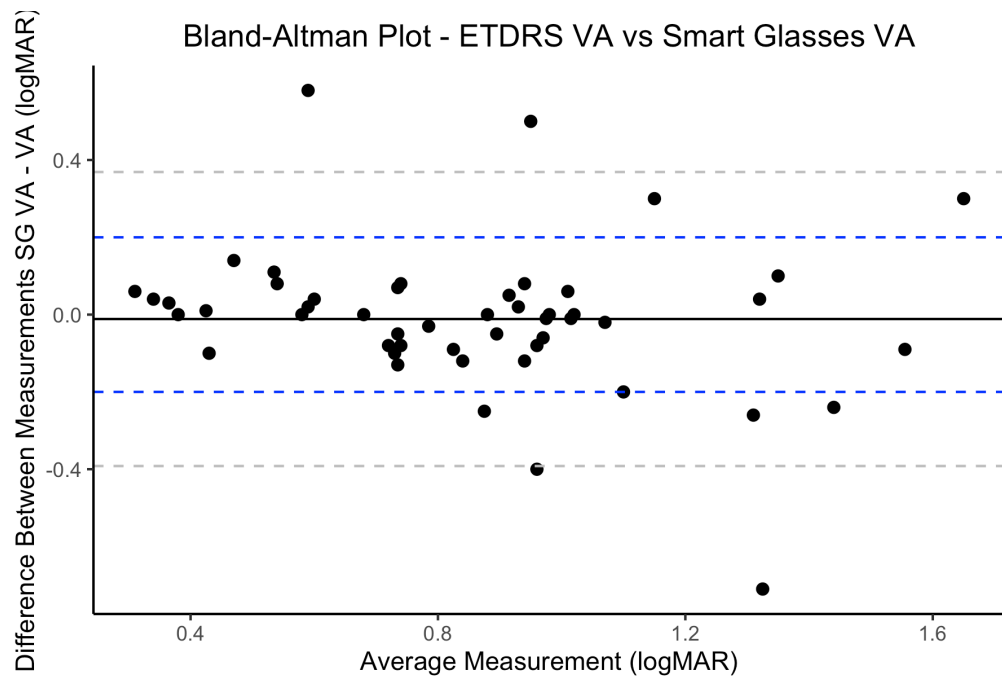


Figure 6-3 Bland-Altman plot of ETDRS VA against Smart Glasses VA. The black line represents the average difference in between the two measurements, the two grey dashed lines represent the 95% confidence interval limits for the average difference. The two blue dashed lines represent ± 0.2 logMAR a clinically relevant change.

Outliers

Some participants performed significantly better or worse using the SG VA test.

Regression analysis of the differences between the two scores did not identify any valid predictors; using contrast sensitivity (CS), colour, age, or primary diagnosis.

A clinically relevant change is considered as 0.2 logMAR (two lines of letters) (Rosser, Cousens, Murdoch, Fitzke, & Laidlaw, 2003). Participants were categorised into positive improvers, non-improvers and negative improvers. 41 (82%) participants were classed as neutral, 5 (10%) positive and 4 (8%) negative.

Table 6.2 lists the outlier participants labelled as positive or negative improvers. Of the four negative improvers, two participants (SP031 and SP035) had difficulty with the display; one experienced double vision induced by the backlit display. The other suffered from photophobia and struggled with the brightness of the displays. Of the five positive

improvers, they generally had lower visual acuity than the other two groups, median 1.44 IQR 0.4 (figure 6.4).

Table 6-2 Patients with significant score differences

Patient	Score difference	ETDRS VA	Primary Diagnosis
SP008	-0.3	1.5	Retinitis Pigmentosa
SP029	-0.3	1.0	Stargardt disease
SP031	-0.5	0.7	Macular Dystrophy
SP035	-0.58	0.3	Undiagnosed
SP009	0.24	1.56	Macular Dystrophy
SP012	0.4	1.16	AMD
SP028	0.25	1.0	Doyne retinal dystrophy
SP036	0.26	1.44	Macular Dystrophy
SP056	0.71	1.68	Stargardt disease

Score difference is the SGVA - ETDRS VA

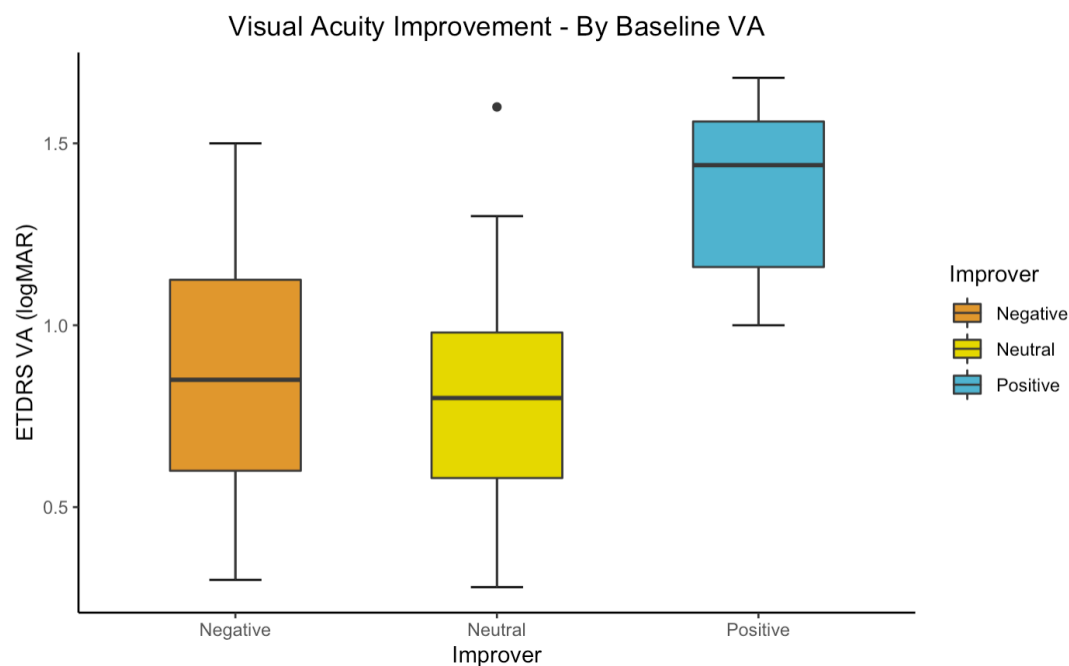


Figure 6-4 Visual acuity by improvement group. The X-axis is the group of negative (orange), neutral (yellow) and positive (turquoise) improvers. The upper and lower boundaries of the coloured boxes are the inter-quartile range. The median is the horizontal line within the box. The error bars represent 1.5 times the Interquartile range. Figure 2.6.4 Visual acuity by improvement group. The X-axis is the group of negative (orange), neutral (yellow) and positive (turquoise) improvers. The upper and lower boundaries of the coloured boxes are the inter-quartile range. The median is the horizontal line within the box. The error bars represent 1.5 times the Interquartile range.

6.4 Discussion

Comparing the two visual acuity tests showed good statistical agreement. Most participants' score difference between the two tests was within natural variation and would not be considered clinically significant (Rosser et al., 2003).

While assessing visual acuity using wearable display use is novel, it is not without comparison. The display technology is similar to that of a tablet computer or mobile phone display. Comparing an iPad tablet computer and the ETDRS chart under binocular viewing by J.M Black et al. (Black et al., 2013) found an interclass correlation coefficient of 0.77, similar to the results found here. Other work validating a smartphone-based test by Bastawrous et al. (Bastawrous et al., 2015) had higher correlations (Pearson's Correlation Coefficient=0.93, compared to 0.83 in this experiment.). It should be noted their population's median VA was 0.23 logMAR, significantly lower than this experiment's patient cohort.

Some participants did perform worse with the SGVA test; two had significant difficulty with the display and made up the most considerable score differences (table 6.3). It should also be noted that visual acuity variability can increase with worsening vision. This has been observed in retinitis pigmentosa (Bittner, Ibrahim, Haythornthwaite, Diener-West, & Dagnelie, 2011) and AMD patients (Patel, Chen, Rubin, & Tufail, 2008b). Patel found a coefficient of repeatability to be 12 letters (ranging from 9 letters to 29). Thus, for this experiment's cohort, a significance threshold of 0.2 logMAR (10 letters) is likely too conservative. It should also be noted that all participants who performed significantly worse with the Smart Glasses stated that they liked them and could potentially find them helpful.

A potential reason for this is some participants found the smart glasses useful for face perception, which has less requirement on the higher spatial frequencies than visual acuity. The improvement of some participants' visual acuity test results could also be related to this cohort's inherent high test-retest variability. The group that improved with the SGVA also had lower visual acuity. The high contrast and brightness of the display could contribute to higher scores; however, to make any significant claims to the source of this improvement, further experimentation and repeated assessments would be required.

One advantage that the wearable display provides when measuring visual acuity is the viewing distance is optically fixed. There is no requirement for a researcher to control for this, as it is impossible for the participant to vary this. This is an added advantage to digital visual acuity tests. However, a drawback is the fixed interpupillary distance. For some participants, careful alignment of the displays with their eyes was required during the experiment, which may be time-consuming as it is bespoke for the individual. Although for the majority, this was not a concern.

This experiment's limitations are the number of tests performed; ideally, multiple repeats of the same tests would have been performed on the same day and several days apart. However, this experiment was conducted as part of a more extensive testing protocol. More repeats would significantly increase the time required to complete or negatively impact participants' performances within other aspects of the experimental protocol.

6.5 Future work

This experiment has highlighted the usefulness of HMD for vision testing. More intensive assessments of the visual acuity performance could be carried out in conjunction with at-home testing.

6.6 Conclusion

This experiment aimed to understand how individuals with vision loss can view text on and the wearable displays. The findings demonstrate that the Epson BT350 displays can easily be used by a range of visually impaired participants.

A secondary finding is that wearable displays can enable visual acuity measurements to a standard similar to a standard clinical method. The impact of this is the potential for a reliable and straightforward at-home visual acuity monitoring tool

Chapter 7

Reading

My contributions to this chapter: I designed, wrote, and obtained ethical approval, including all required patient information and recruitment documentation. I designed and wrote all virtual reality software using Unity in C#. The majority of the data collection was performed by myself, with one month of support from Aran Johal. I performed all data processing and statistical analysis using Python and R.

Other contributions: Caroline Justice aided with patient communication and room bookings. Professor Downes helped with patient identification. Aran Johal assisted with data collection in September 2019. Professor MacLaren and Professor Downes assisted with proofreading.

Timeline: Data collection was performed between March 2017 and September 2019. Data analysis and software development took three months.

7 Reading

This section of this thesis focuses on reading and how a wearable display can potentially enhance this for people with visual impairment.

7.1 Introduction

Reading is a task that is central to modern life. Patients with bilateral central vision loss face numerous difficulties. One of the most significant is increased difficulty with reading, leading to a lack of independence and a considerable impact on quality of life (Mitchell & Bradley, 2006).

7.1.1 Reading and central vision loss

Difficulties with reading are quantified by measuring reading acuity, the smallest text size, and reading speed. Early-stage central vision loss (CVL) sufferers may not have significantly diminished visual acuity but can reduce reading speed. As diseases progress, both are significantly affected. A deficit in performance for either metric can increase the individual's cognitive strain, leading to discomfort, stress, and quality of life issues (Murro et al., 2017).

7.1.2 How is reading measured

Within ophthalmology, there exist several calibrated reading tests; W. Radner provides a comprehensive review of historical and modern methods (Radner, 2017). Of the contemporary techniques, Bailey–Lovie Word Reading Chart (Bailey & Lovie, 1980), MNREAD Charts (Calabrèse, Owsley, McGwin, & Legge, 2016), SKread Charts (MacKeben, Nair, Walker, & Fletcher, 2015), and the RADNER Reading Charts (Radner et al., 1998) all provide measures of reading acuity and speed. Bailey–Lovie uses short lists of unrelated words of four, seven- and ten-character lengths, descending in logarithmic steps. Similarly, the SKread test uses random unrelated words designed to estimate the location of

scotomas. MNREAD uses a similar method to Bailey-Lovie, but with complete sentences of 60 characters and ten standard-length words. The RADNER reading test introduces the concept of sentence optotypes; they generated a series of sentences that are quantifiably comparable. These tests all provide quantifiable measures of reading ability and have established test-retest repeatability with the MNREAD and RADNER tests most commonly used in research.

The metrics generated by these charts to describe reading performance include reading acuity (RA), the smallest size text one can read. Maximum reading speed (MRS) is the maximum reading speed one can achieve, and the critical print size (CPS) is the minimum size of text required to achieve a maximum reading speed. The relationship between text size and reading speed has been covered by Legge and Bigelow (G. E. Legge & Bigelow, 2011), where text size increases, so does reading speed up to a specific size – the critical print size. Speed then levels off at the maximum reading speed.

7.1.3 Why is reading hard for CVL?

Age-related macular degeneration and associated difficulties with reading have been the subject of several well-observed studies, with quantifiable difficulties reported by numerous studies. The reasons for reading being impacted by CVL is multifactorial.

As CVL progresses, the maximum visual acuity of an individual is reduced. Individuals start to use a different fixation area with bilateral vision loss, known as a preferred retinal locus (PRL) (Whittaker et al., 1988). Moving eccentrically from the fovea, the retinal ganglion cells' receptive field increases (Christine A. Curcio et al., 1990), reducing spatial resolution

(Westheimer, 1979). Chung et al. (Chung, Mansfield, & Legge, 1998) found that reading speed improved with text size when viewed in the periphery, in normal eyes. However, even with magnification, S. Chung (Chung, 2020) found that AMD patients reading speed showed less improvement than normally sighted volunteers. Legge et al. (Gordon E. Legge, Rubin, Pelli, & Schleske, 1985) also concluded that CVL sufferers had a lower reading speed even with magnification than people with normal vision. This evidence suggests that more factors play a part than just the lower spatial resolution typically found in CVL patients.

Contrast Sensitivity loss is expected with CVL patients, commonly observed in AMD patients (Chung & Legge, 2016; Elliott, 1987). Rubin et al. found that reading speed was lower than in control participants when compensating for this contrast sensitivity loss, implying that contrast alone is not the sole contributing factor.

Crowding is the effect of increased difficulty with object identification when it is surrounded by other objects (clutter) (Wallace et al., 2017). As mentioned in chapter 2.4, it impacts face recognition and is also known to affect letter acuity (Bauma, 1970) and reading. Crowding effects increase with eccentricity. However, while line spacing improved reading speed S. Chung (Chung, 2002) found increasing the letter spacing did not lead to an increased reading speed in CVL.

During reading, the eye moves in a sequence of steps (saccades) and fixations. Visual span is the number of letters identified in a single fixation, and perceptual span is the number of letters influencing saccades and fixation timings. Visual span has been hypothesised to be a limiting factor in reading speed (Gordon E. Legge et al., 2007). The consequence of a reduced visual span is more saccades during reading, which limits reading speed. Legge et al. (Gordon E. Legge, Mansfield, & Chung, 2001) showed that the visual span size decreases from at least ten letters in central vision to 1.7 letters at 15° eccentricity. AMD patients

show a reduced visual span (Cheong, Legge, Lawrence, Cheung, & Ruff, 2008) compared to normal central vision.

Fixation stability has been observed to be reduced in CVL (Bethlehem et al., 2014), and has been shown to correlate with reading performance (Amore et al., 2013; Crossland, Culham, & Rubin, 2004; Giacomelli et al., 2013). The impact of poor oculomotor control was assessed by Rubin et al. (Rubin & Turano, 1992) by rapid serial visual presentation (RSVP), where text is presented in the same location on a screen serially and removes the need for saccadic eye movements. They found that while reading speed increased for CVL participants; the gain was smaller than normal participants.

It should be clear from all of these investigations that a single factor does not fully describe reading performance. Reading is a complex task involving many functions, and interference in any of these functions can impact performance.

7.1.4 How can reading be improved?

7.1.4.1 *Training*

Eccentric viewing training is a widely adopted approach used by low vision rehabilitation services. The method teaches an individual to use a PRL for reading and provide training exercises and support. Exact methodologies vary, but the technique is effective. Palmer et al. (Palmer, Logan, Nabili, & Dutton, 2010) evaluated a four-year programme and identified that most patients improved in one or more vision-related tasks after this implementation. An analysis of training on individuals with absolute central scotomas by Howe (Howe, 2012) concluded that methods vary, and no one method was obviously better, but overall the results were positive.

An alternate approach is called perceptual learning, defined by E Gibson (Gibson, 1969) as "relatively permanent and consistent change in the perception of a stimulus array, following practice or experience with this array". The approach has been applied to vision rehabilitation, where typically participants repeat a visual task. S. Chung (Chung, 2011) found that CVL patients showed improved RSVP (rapid serial visual presentation) reading performance after six weekly sessions. They concluded that the improvement was not related to PRL location or stability and was likely to be due to the inherent plasticity of the visual system. Similar results were found by (Tita-Nistor, González, Mkowitz, & Steinbach, 2009) however, they reported a different PRL after training and improved fixation stability. Yu et al. (Yu, Cheung, Legge, & Chung, 2010) used repeated visual-span measurements as a training task and found reading speed improvements in the peripheral vision of older participants and an increase in visual span. The advantage of the perceptual learning method is its simplicity; participants are asked to repeat a task, some of which can easily be performed at home (Yu, Legge, Park, Gage, & Chung, 2010).

While reading training approaches have been successful, they demand a significant effort from the participant. Eccentric viewing training also requires trained technicians to support this implementation. Perceptual learning has the potential to be performed unsupervised, but it is yet to become widely adopted.

7.1.4.2 Presentation

Modifying the presentation of text can lead to reading improvements. As mentioned earlier, RSVP (Rubin & Turano, 1992) is a text presentation method that removes reading reliance on saccadic eye movements. This alternate presentation can and has been used to improve

reading speed. Also, attempts have been made to design fonts to improve readability for CVL. Xiong et al. (Xiong, Lorscheid, Mansfield, Bigelow, & Legge, 2018) assessed two fonts, one designed to maximise the shape difference between letters and another bold and wide. They found that while the two fonts improved acuity and reading speed compared with Helvetica and Times, they had no advantage over Courier. Xiong et al. hypothesised the reason for Courier performing best is likely familiarity due to its prevalence. Aries and Cho (Arditi & Cho, 2005) compared serif and sans-serif fonts, finding that serif fonts may increase legibility on a small font. They attributed this increase to the larger inter-letter spacing. However, they only assessed two visually impaired participants. Other work by Moret-Tatay and Perea on serif fonts (Moret-Tatay & Perea, 2011) found the opposite, with sans-serif fonts showing a slight advantage in word recognition response times. This study was with normally sighted participants only.

The impact of font type is subtle, reflected in the small and variable results in the aforementioned studies.

Altering text presentation has limitations; it requires control over the reading materials. Reading in the world, such as signs and directions, are not suitable. However, with the ever-increasing involvement of digital technology in life, this is becoming less of a barrier.

7.1.4.3 Reading aids

Most patients with low vision require some form of magnification to read and to have good lighting. Patients with central scotomata, i.e. Age-related Macular Degeneration or Stargardt disease, typically need much higher magnification than those with some remaining central vision (Gordon E. Legge et al., 1985).

The most commonly used type of magnification is the handheld optical magnifier (Virgili, Acosta, Grover, Bentley, & Giacomelli, 2013). Optical magnification is simple and effective but does have limitations. When increasing levels of magnification are used, the field of view and depth of field decrease. A reduced field of view impacts negatively upon reading speed. To read efficiently, one needs to see a minimum number of letters between 15-30 letters (G E Legge, Pelli, Rubin, & Schleske, 1985). A reduced depth of field relates to how precisely a magnifier needs to be placed above the object of interest, which means that high magnification handheld lenses can be challenging to align correctly.

Electronic magnifiers can provide an advantage over optical magnifiers (Peterson, Wolffsohn, Rubinstein, & Lowe, 2003). They can provide a higher magnification without having such a significant impact on the field of view. However, the field of view is limited by the size of the display and camera used. Electronic magnifiers have specific drawbacks; they are often much larger and heavier than their optical counterparts and are significantly more expensive than handheld magnifiers. However, the electronic nature of these devices provides opportunities to assist with reading beyond simple magnification. Many existing electronic handheld devices allow users to amplify contrast and to change the text and background colours. Inverting contrast, for example, presenting white letters on a black background, has improved reading speed for low vision (G E Legge & Rubin, 1986).

However, changing the text's colour or background has not been specifically investigated or identified to provide any objective benefit for low vision. The most definitive work in the area has been focused on psychophysics aspects that are less representative of any 'real-world' scenarios (Gordon E. Legge & Rubin, 1986). Although there is little empirical evidence

that yellow and black improve the readability of text, black text on a yellow background has been used for years in the UK by Low vision charities for correspondence and eye hospital signage to help patients navigate around their departments.

Head-mounted displays are becoming more available as aids for those with low vision; several currently commercially products are available. They can provide magnification and image enhancement features such as contrast enhancement and colour manipulation. Table 1-1 covers the various specifications of these head-mounted displays. One significant advantage is that they are not reliant on external light sources. Also, head-mounted displays can avoid handling issues that individuals affected by arthritis may have when using handheld optical magnification.

7.1.5 Previous work HMD and reading

Previous work has investigated the impact of head-mounted displays on reading; Culham et.al (Culham, Chabra, & Rubin, 2004) compared reading with four different wearable displays. While some improvements were made with the wearable displays, they concluded that optical magnification aids remained the best devices for most tasks. Wittich et al. (Wittich, Lorenzini, et al., 2018) assessed the eSight wearable display, and although they found improvements in reading acuity, there was no evidence for improvement in reading speed. Crossland et al. (Crossland et al., 2019) evaluated the SightPlus head-mounted display with similar results to Wittich. Deemer et al. (Deemer et al., 2019) qualitatively assessed two different image processing strategies, a magnification bubble and a virtual projector screen. Evaluations were performed using the Activity Inventory questionnaire

(Massof et al., 2007). Participants reported improvements in reading and visual information processing.

Head-mounted displays enable the use of more novel strategies. Spatial remapping is an approach where an image is distorted so that no information falls within an individual's scotoma. Gupta et al. assessed the system with simulated CVL (Gupta et al., 2018) providing an advantage. However, as eye-tracking is required for any successful implementation, it is far from practical use. Eye-tracking was also used in work by Aguilar and Castet (Aguilar & Castet, 2017); they tested a CCTV (closed-circuit television) system that applied magnification only to text in line with a person's gaze. An improvement in reading comfort was observed in simulated vision loss.

The majority of the studies using HMD have been focused on specific commercial devices, and thus the results reflect the chosen device user interfaces' idiosyncrasies. A significant difficulty with wearable displays, mainly when high magnification is used, is the impact on hand-eye coordination. The cameras used are attached to the HMD, and directing focus is performed by moving the head. Although learning effects were observed in these studies (Culham et al., 2004; Wittich, Lorenzini, et al., 2018), despite this, the loss of hand-eye coordination is a major limiting factor to performance and ultimately uptake.

The methodologies used to assess reading varied; only Wittich and Crossland (Crossland et al., 2019; Wittich, Lorenzini, et al., 2018) used a standardised reading test, with the other studies either using patient-reported outcome measures or non-standardised tests.

7.2 Reading with a virtual reality display

Experience with the Epson BT-350 wearable display in chapters 5 & 6 of this thesis highlighted its limitations, the low-resolution camera and offset position. Pilot work found the unmodified wearable display unsuitable for any assessments with reading. A significant challenge was, as noted in previous work (Culham et al., 2004; Wittich, Lorenzini, et al., 2018), was the hand-eye coordination, i.e. aligning text with the camera. For these reasons, a virtual environment was chosen for reading assessments. This allowed the removal of any human interface difficulties and camera limitations and focused entirely on the impact of the display and text presentation.

This section's primary aim is to investigate if the head-mounted virtual reality display can assist with reading. Different text and background colour combinations are used to identify any advantage they may provide.

7.2.1 Methods

7.2.1.1 *Participants*

Patients were recruited from the Oxford Eye Hospital outpatient clinic after informed consent. Ethical approval was attained for the study, which was carried in accordance with the tenets of the Declaration of Helsinki. [NHS ethics reference: 17/SC/0178]. Participants were selected based on the following inclusion criteria; at least 18 years old, best-corrected visual acuity (BCVA) of 0.6 logMAR (6/24) or worse and a conversant English speaker.

7.2.1.2 *Protocol*

All subjects underwent the same testing protocol;

- Informed consent
- Visual assessments
- Radner Reading Test
- Virtual Reality Reading Test
- Virtual Reality Binocular Visual Acuity

7.2.1.3 Visual assessments

7.2.1.3.1 Binocular Visual Acuity

Best-corrected binocular visual acuity (BCVA) was measured using the Revised 2000 Series Early Treatment for Diabetic Retinopathy Standard (ETDRS) letter reading eye (Precision Vision, La Salle, Illinois, USA). BCVA was calculated as the lowest complete row read, subtracting 0.002 log units for every letter correctly identified beyond that point.

7.2.1.3.2 Contrast Sensitivity

Binocular contrast sensitivity was measured at five spatial frequencies, 1,2,4,8 and 10 cycles per degree (Cambridge Research Systems, Cambridge, UK), described in section 4.1.2.6.

7.2.1.3.3 Colour Sensitivity

Binocular colour sensitivity was measured using a version of the Cambridge Colour Test (Cambridge Research Systems, Cambridge, UK), described in section 4.1.2.7.

7.2.1.4 Radner Reading

The Radner Reading test (Radner et al., 1998) was used as the basis for the reading assessments. The Radner test contains short standardised sentences; a participant is asked

to read aloud at a fixed viewing distance, binocularly, with any habitual reading aid. The participant is asked to keep the same reading distance, resisting the urge to bring the text closer as it decreases in size. Reading distance is visually monitored by the experimenter, who prompts the participant if they noticeably move the text closer.

The testing methods and design have been described in another publication (Stifter et al., 2004). The test provides two measures, reading acuity and reading speed (words per minute). Reading acuity is reported in terms of logRAD (Log Reading Acuity Determination). This is determined as the last sentence read, adjusted by the number of reading errors and viewing distance.

7.2.1.5 Virtual Reality

The virtual reality (VR) device used was the HTC Vive (HTC Corporation, Xindian, New Taipei City, Taiwan). The HTC Vive consists of a headset, two controllers, and two infrared laser emitter units. The headsets field of view is approximately 110 degrees (around 90 per eye) through two 1080x1200 pixel OLED displays updated at 90 Hz. This gives approximately 12 pixels per degree or 5 minutes of arc per pixel (figure 7.1).

The VR tasks were implemented with the Unity Engine version 2018.1.8 (Unity Technologies, San Francisco, USA). Unity is a complete development environment and rendering platform. Custom control scripts were written in C#.



Figure 7-1 HTC Vive virtual reality headset (centre) with two handheld controllers (left & right)

7.2.1.6 Reading

A reading test was implemented within the virtual environment using the exact standardised text as the Radner test. The sentences were displayed sequentially on a virtual piece of paper, with the exact descending font sizes as the Radner test (see appendix 3.2 for implementation and calibration details). The virtual paper was attached to one of the HTC Vive controllers held by the participant (figure 7.2). Participants were asked to perform the virtual reading test, with the same methodology as the physical Radner test, binocularly and with any habitual reading aid (the HTC Vive headset can comfortably be worn on top of most spectacles). The test was repeated with four presentation formats; Black Text on a white background, White on Black, Yellow on Black and Black on Yellow.

The VR device allowed for head and hand positional information to be continuously recorded, at 90Hz, allowing for a precise viewing distance measure. The Virtual Reading test provides three measurements for each test: head & hand movement, reading acuity and

reading speed. Reading acuity is calculated using the text size and reading distance [see appendix 3.2]. Participants were also asked to rate their subjective preference of the four different presentation formats.

The VR reading distance is accurately and continuously monitored within the VR software and requires no experimenter input. In contrast, this distance is hard to measure in practice accurately when performing the Radner reading test.



Figure 7-2 – Screen capture from Unity – The virtual paper, black text on the white background, showing an example sentence. Note - The image's low resolution, and blocky text, is an unavoidable artefact of the capture method used, it is much clearer when viewed in the HMD.

7.2.1.7 Visual Acuity

Binocular Visual acuity was measured using a virtual ETDRS chart within the Unity virtual environment. The chart's viewing distance was fixed relative to the head position; if the

participant could not view any characters on the chart, the distance was reduced by 0.5-metre increments.

7.2.1.8 Statistical Analysis

All statistical analysis was performed using the R statistical package (R Development Core Team, 2008). Data were tested for normality using Shapiro-Wilk's method (Shapiro & Wilk, 1965). Linear regression and Analysis of Variance (ANOVA) or non-parametric Kruskal-Wallis tests were used. Correlations were calculated using Pearson's product-moment correlation or Spearman's rank correlation, according to the normality of the data.

7.2.2 Results

A total of 53 participants were recruited. The age range was 18 to 88 years (mean 52 ± 21), including 21 females (40%) – (table 7.2) The binocular logMAR visual acuity ranged from 0.3 to 1.7 (mean 0.87 ± 0.4). The patients had a range of retinal conditions; table 7.1 summarises the cohort.

The following measures were extracted from the Radner test and the VR reading test:

- RA = reading acuity (logRAD), the smallest text read.
- CPS = critical print size, the RA where they read the fastest.

Table 7-1 Frequency of conditions

Frequency of Conditions	
Condition	Frequency
Retinitis Pigmentosa	4
Myopic Degeneration	1
Stargardt	15
Optic Atrophy	1
Macular Dystrophy	7
AMD	7
Cone Dystrophy	4
Sorsby Fundus Dystrophy	1
Nystagmus	1
Retinitis Pigmentosa (Ushers)	1
Congenital Cataracts	1
Doyne retinal dystrophy	1
Retinal Hemangioblastomas	1
Cone-Rod Dystrophy	3
Unknown	1
Optic Neuropathy	1
CSCR	2
Pattern Dystrophy	1
Total	53

Table 7-2 Patient participants

Participants				
ID	Condition	Gender	Age	logMAR binocular acuity
SP001	Retinitis Pigmentosa	M	78	0.40
SP002	Myopic Degeneration	F	18	0.70
SP004	Stargardt	M	27	0.42
SP005	Stargardt	M	44	1.02
SP006	Optic Atrophy	M	35	0.87
SP007	Stargardt	F	25	1.02
SP008	Retinitis Pigmentosa	M	31	1.50
SP009	Macular Dystrophy	F	78	1.56
SP010	Stargardt	F	23	0.92
SP011	AMD	M	77	0.40
SP012	AMD	F	81	1.16
SP013	Stargardt	M	36	1.60
SP014	AMD	F	75	1.30
SP015	Cone Dystrophy	F	38	0.80
SP016	Stargardt	F	53	1.30
SP017	Stargardt	F	49	1.00
SP018	Stargardt	F	65	0.35
SP019	Sorsby Fundus Dystrophy	F	59	0.48
SP020	Stargardt	F	45	0.76
SP021	Nystagmus	M	45	0.98
SP022	AMD	M	88	0.58
SP023	AMD	F	65	0.68
SP024	Retinitis Pigmentosa	F	29	0.88
SP025	Retinitis Pigmentosa (Ushers)	M	50	0.58
SP026	AMD	F	78	0.58
SP027	Congenital Cataracts	M	64	0.48
SP028	Doyne retinal dystrophy	M	73	1.00
SP029	Stargardt	M	23	1.00
SP030	Retinal Hemangioblastomas	M	55	0.92
SP031	Macular Dystrophy	M	54	0.70
SP032	Cone-Rod Dystrophy	M	26	0.98
SP033	Cone-Rod Dystrophy	F	63	1.60
SP035	Unknown	M	60	0.30
SP036	Macular Dystrophy	M	50	1.44
SP038	Cone Dystrophy	F	29	1.00
SP039	Cone Dystrophy	M	40	1.08
SP040	Stargardt	F	82	0.28
SP041	Optic Neuropathy	M	21	0.78
SP042	Cone Dystrophy	F	37	0.89
SP043	CSCR	M	69	1.20
SP044	Macular Dystrophy	M	50	0.90
SP045	AMD	M	85	0.98
SP046	Stargardt	M	18	0.70
SP048	Cone-Rod Dystrophy	F	26	0.32
SP050	Retinitis Pigmentosa	M	57	0.38
SP051	Stargardt	M	23	0.78
SP052	Macular Dystrophy	M	74	0.90
SP053	CSCR	M	82	0.90
SP054	Pattern Dystrophy	M	76	0.80
SP056	Stargardt	F	43	1.68
SP057	Stargardt	M	46	1.00
SP058	Macular Dystrophy	M	71	0.50
SP059	Macular Dystrophy	M	54	0.70
		F(21) M(32)	51.75±20.8	0.87±0.40

7.2.2.1 Radner Reading

A baseline reading performance was measured with the Radner reading test. Forty-nine participants completed the test; 5 were unable to complete the test due to low vision. All 5 had visual acuities of 1.3 logMAR and below.

Reading acuity was found to be normally distributed ($p=0.7$). The mean reading acuity was 0.87 ± 0.4 logRAD, ranging from 0.36->1.5. CPS was not normally distributed ($p=0.008$), median 1.14 ± 0.2 logRAD, ranging from 0.7 -> 1.7.

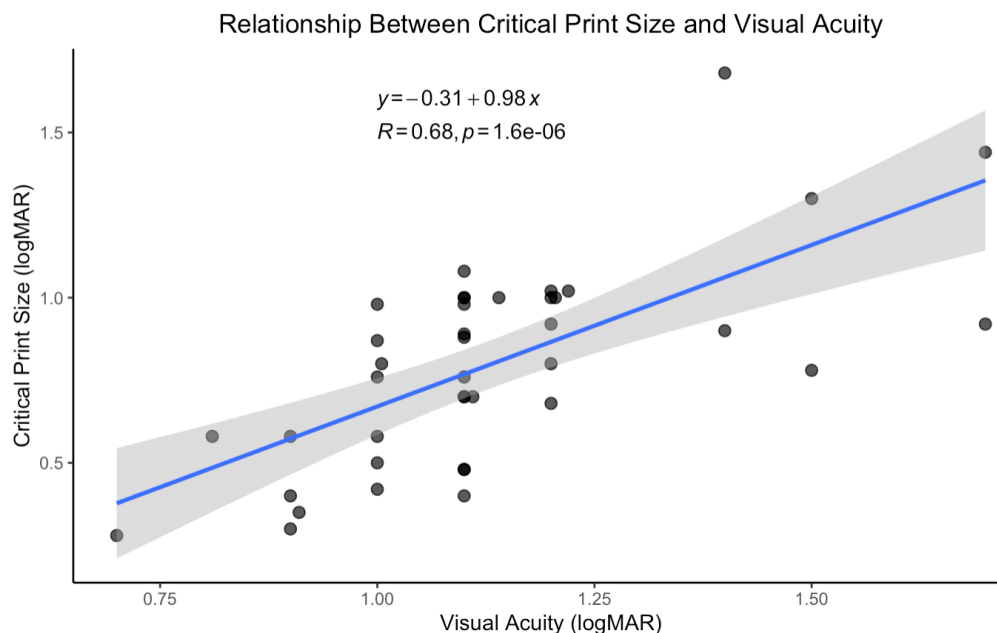


Figure 7-3 Relationship between Critical print size (CPS) and visual acuity (VA). The blue line is the regression line from the estimated coefficients. The grey band indicates the 95% confidence intervals for the regression fitting

7.2.2.1.1 Factors influencing performance

The only consistent predictor of reading performance was visual acuity. Reading acuity correlated with visual acuity [linear regression; $p < 0.0001$, $R^2 = 0.61$]. The base 10 log of Critical Print Size correlated with visual acuity [linear regression; $p < 0.0001$, $R^2 = 0.5$].

The above models all met the assumptions of homogeneity of variance, linearity and the residuals were approximately normally distributed.

7.2.2.2 Virtual Reality Reading

The VR headset's impact on visual acuity and reading was investigated, along with the effect of different colour combinations. Fifty-two participants completed the VR reading test; Only one could not complete the test due to too low vision. Figure 7.4 presents a single participant output from the VR reading task.

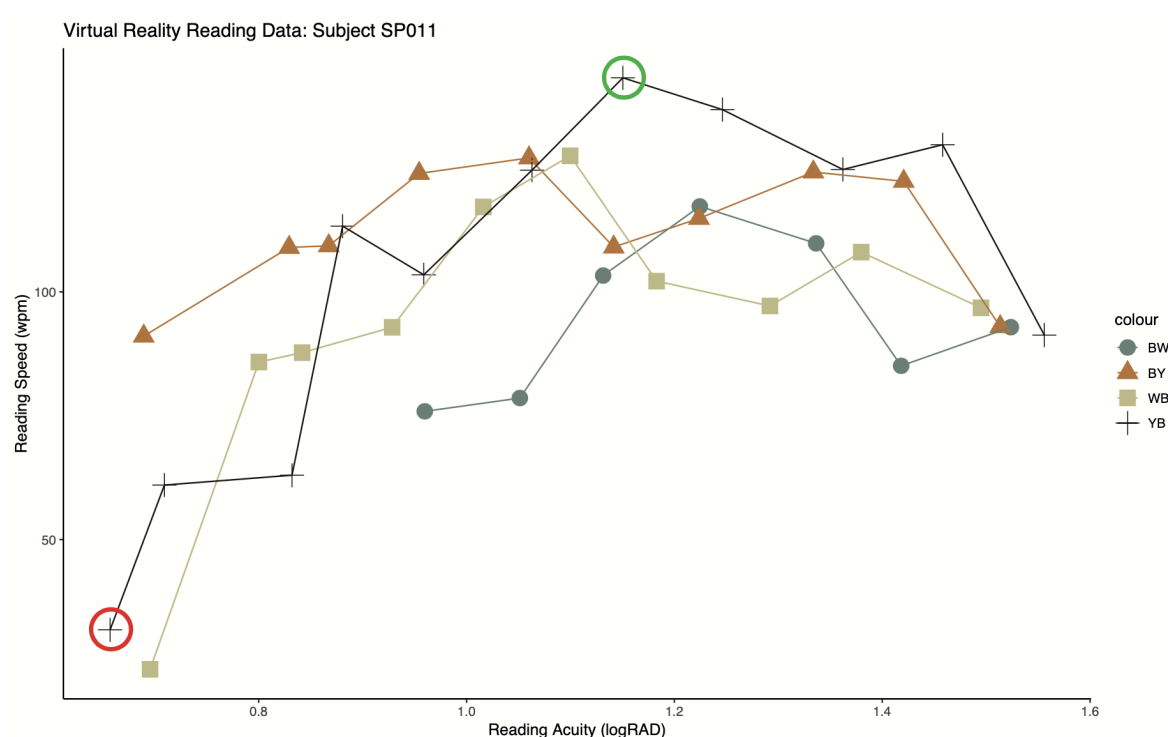


Figure 7-4 . Example data output from the four Virtual reality reading tests. The circular points represent Black on White presentation, Triangle = Black on Yellow, Square= White on Black, Cross = Yellow on Black. It can be seen that this participant smallest text read, identified by the red circle, was presented as Yellow on Black (Reading acuity = 0.66 logRAD), and the fastest read text, identified by the green circle, was also presented Yellow on Black (Critical Print Size = 1.15 logRAD).

7.2.2.3 *Virtual Reality Visual Acuity*

Comparison of the VR visual acuity and the real-world visual acuity allows for investigating any interaction of the display technology upon the participant's vision, using an established clinical assessment.

Figure 7.5 presents a comparison of the binocular visual acuity with VR binocular visual acuity. The resolution of the VR display is limited, creating a floor effect. The pixel size of the display drives the theoretical limit. For the HTC Vive, 0.7 logMAR is the lower limit. However, a few individuals were able to perform better ~ 0.6 logMAR. The lower limit of the display was calculated with the rendered text, assuming to be five times the minimum angle of resolution of the display, 5 minutes of arc.

When accounting for this floor effect, the VR BCVA and BCVA are found to be not significantly different [paired t-test; $t(19)=-0.81, p=0.42$] and correlated [linear regression; $p=0.0002, R^2=0.55$].

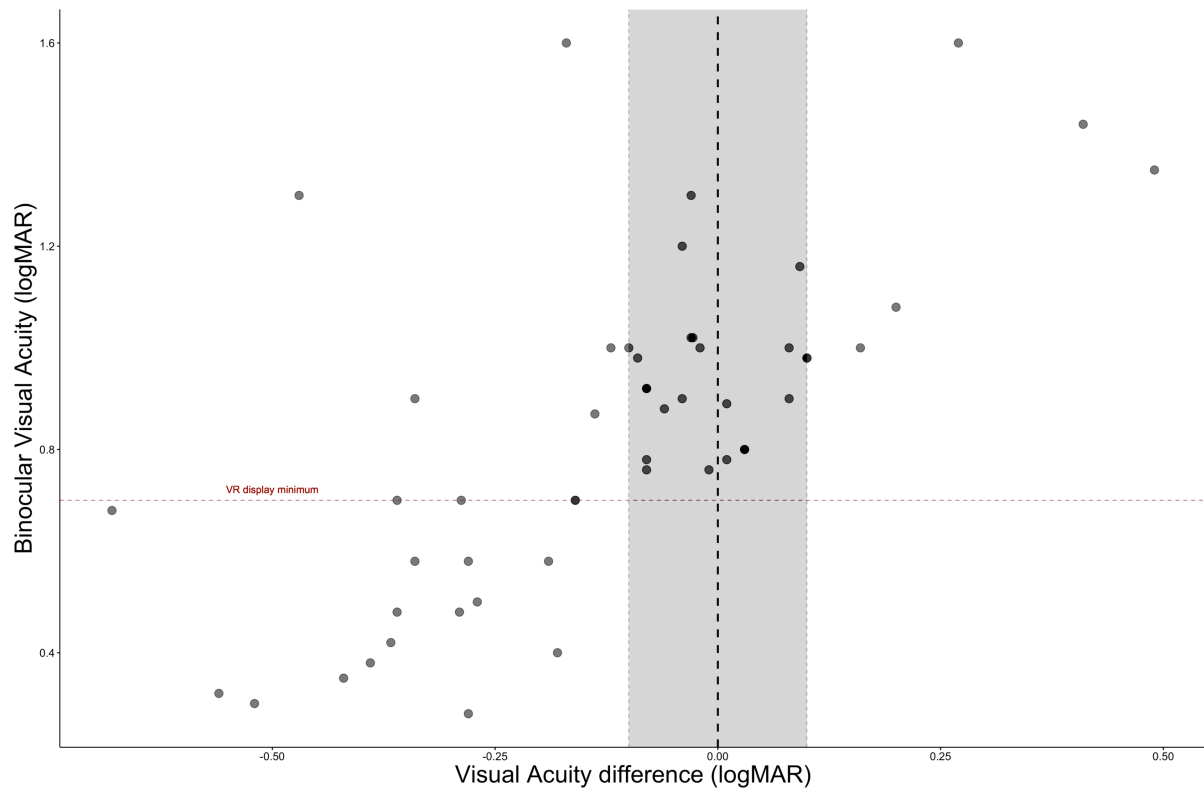


Figure 7-5 : : Presents a comparison of the binocular visual acuity with VR binocular visual acuity. The X-axis is the difference between the BVCA and virtual reality BVCA, where a negative value indicates a worse score in the VR BCVA. The red horizontal line indicates the theoretical limit of the VR display (0.7 logMAR). The grey vertical area is a boundary of ± 0.1 logMAR, which is considered a clinically relevant change.

7.2.2.4 Comparison of VR reading to Radner reading

The physical Radner reading test and the VR reading tests were compared to identify any significant differences.

Accounting for the display resolution limitation, an analysis of variance showed no significant differences with reading acuity [$F(4,109)=0.97$, $p=0.43$] between the Radner reading test and all four VR reading tests. Comparing the CPS, there was a significant difference with the Radner CPS compared to all of the VR tests CPS [$F(4,109)=9.25$, $p<0.001$]. This difference is likely due to the VR test's ability to present much larger text than the Radner. The presented Radner text's mean size was 0.98 ± 0.27 , with a 5% to 95% range of [0.51 -> 1.4] logRAD. The presented VR text's mean size was 1.28 ± 0.26 , 5% to 95% range

[0.83 -> 1.69] logRAD, showing a shift to larger print sizes. This implies that the Radner test did not fully capture the CPS for some participants due to the smaller maximum text size.

A small number of subjects could not read from the Radner test, even when bringing the test as close as possible to the eye. However, they were able to read with the VR test. One reason for this could be the increased luminance contrast of the display and the larger print size. A more likely explanation is that the VR display has a fixed focal plane. The largest text on the Radner chart is 6.3M (approximately 9.2mm high). If an individual has to bring the text very close to their eye (which was observed in some individuals, ~ 10 cm), they will be limited by their minimum accommodation distance. This, however, is not a limitation with the VR display; the accommodation distance is set to approximately 1 metre. Thus, the user can virtually bring the text close to the eye without reaching any accommodation limit.

7.2.2.5 Effect of text presentation on reading performance

The different colour combinations were investigated to understand if they had an impact on reading performance.

The VR reading data was transformed relative to the Radner reading data, allowing for inter-subject comparisons. The transformation was performed by subtracting the Radner results from the VR reading results.

A Kruskal-Wallis test was conducted to compare text/background colour's impact on reading acuity and critical print size. No significant differences were found.

Table 7.3 presents the correlation of reading acuity scores across all tests. The correlation of reading acuity within the VR reading was high [0.88 -> 0.90]. The Radner reading test's correlation is also positive but not as strong [0.63->0.70].

Table 7-3 Reading Acuity correlation

	Black on White	Black on Yellow	White on Black	Yellow on Black
Black on White	1.0	0.88	0.87	0.86
Black on Yellow		1.0	0.90	0.90
White on Black			1.0	0.90
Yellow on Black				1.0
Radner	0.73	0.74	0.63	0.69

All values are Pearson's product-moment correlations (row vs column) with p-values < 0.0001

When comparing at a group level, no significant differences were observed. However, when assessing an individual's scores, there were differences within the presentations. The participants' best and worst scores for RA and CPS across the four different coloured VR presentations were compared (figure 7.6). There were significant differences when comparing the RA best to the RA worst; paired t-test ($t(49)=-13.7$, $p<0.0001$), with an average difference of 0.15 ± 0.07 . Comparing the CPS best to the CPS worst also was significant; paired t-test ($t(49)=-11.8$, $p<0.0001$), with an average difference of 0.22 ± 0.14 .

While all participants had a best and worst performance, not all were considered significant. I considered an improvement of greater than 0.1 logRAD to be significant. A RA difference of greater than 0.1 logRAD was observed in 31 of 41 (76%) participants, and 36 of 41 (88%) had a CPS difference greater than 0.1logRAD.

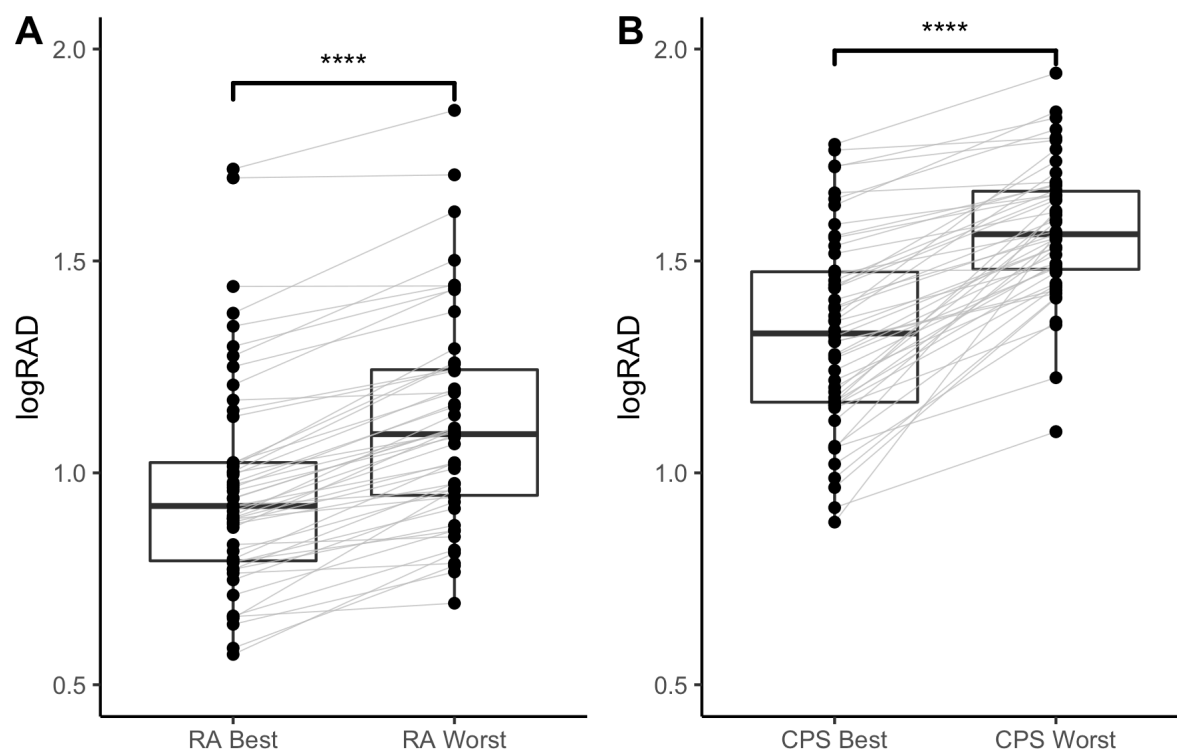


Figure 7-6 Comparison of best and worst scores. A) Presents Reading Acuity best / worst scores, B) presents critical print size (CPS) best / worst scores. The starred bars indicate levels of significance; **** = $p < 0.0001$. RA = Reading Acuity, CPS = Critical Print Size.

7.2.2.6 Preferred text presentation

Reading comfort is potentially not captured by the reading tests results, so each participant was asked to state a preferred colour presentation. Six participants expressed no preference, two said black background with no preference to text colour, and one stated black text with no preference to the background colour.

It is possible to objectively choose the best presentation method for an individual based on their test scores. There are two measures available to make this decision: Reading Acuity (RA) and Critical Print Size (CPS). RA refers to the smallest font size a participant is able to read; CPS is the font size of the fastest read. The distribution of the chosen colours using different selection methods is shown in figure 7.7 Yellow-on-black was a clear favourite, with black-on-yellow the least.

Investigating the different preference methods for similarity, using Pearson's Chi-Squared test, identified a strong association between the objective Reading Acuity selection method and the participants' subjective selection. [$p=0.002$, $\Phi=0.56$] (Where Φ is Cramer's V strength of association).

Grouping the preference by primary diagnosis is difficult with this heterogeneous group. However, the two largest groups were AMD (7) and Stargardt (15). No AMD participant preferred the black text on yellow, and the Stargardt group tends to prefer presentations with black text with either yellow or white background.

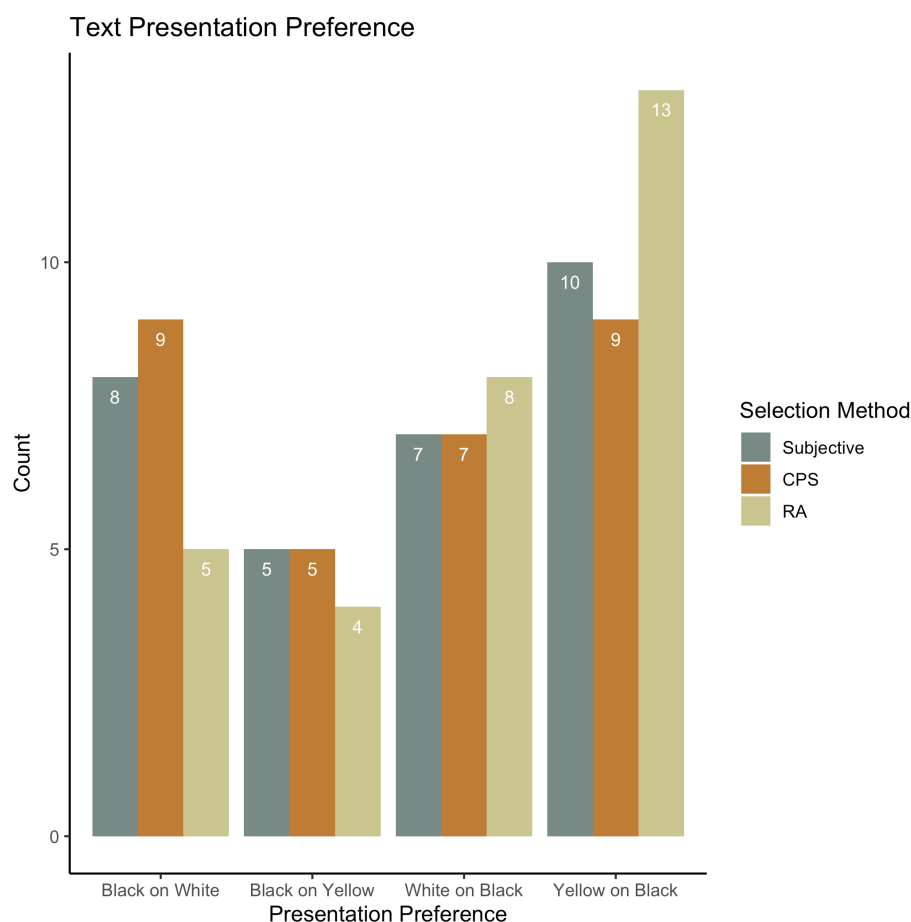


Figure 7-7 Presentation preference selection method distribution. Bar colours represent the selection method used to choose the preference. The numbers represent the totals in each group.

7.2.2.7 Factors relating to presentation preference

I attempted to understand any factors that are predictive of colour preference. A Kruskal-Wallis rank-sum was conducted to test for any difference in visual acuity across the colour preference, no significant differences were found [$H(3)=7.09$, $p=0.07$].

As the stimulus differed by colour, I investigated the colour performance of the participants. The colour test used quantifies the protan, deutan and tritan colour deficiencies. The magnitude of loss was calculated as the vector norm (Euclidian distance from the origin).

$$magnitude = \sqrt{protan^2 + deutan^2 + tritan^2}$$

Visual acuity did not correlate with colour sensitivity loss for the cohort. Narrowing by primary diagnosis, only the AMD group showed a correlation [Pearson's $R=0.93$]. However, this is a small group of 7 participants.

Figure 7.1 (Stargardt v AMD) highlights a feature of the differential diagnosis. The colour vision deficiency is significantly higher in the AMD group, notably the Tritan deficiency. The mean VA of the Stargardt group is higher than the AMD (0.91 vs 0.76, respectively), highlighting the different pathologies.

A Kruskal-Wallis rank-sum was conducted to test for any difference in colour vision deficiency across the colour preference groups [$H(3)=0.0915$, $p=0.04$]. A posthoc analysis using a Wilcoxon rank-sum test showed a significant difference between white-on-black and black-on-yellow [$W=54$, $p=0.007$] (figure 7.8).

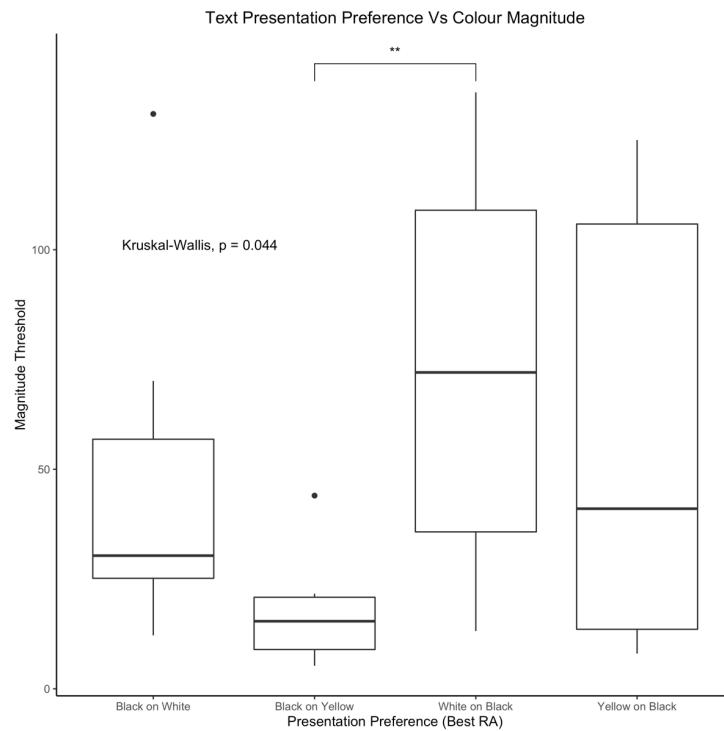


Figure 7-8 Colour deficit magnitude grouped by presentation preference. Preference is chosen using the best Reading Acuity method. There is a significant difference seen between white on black and black on yellow, The starred bars indicate levels of significance; ** : $P \leq 0.01$

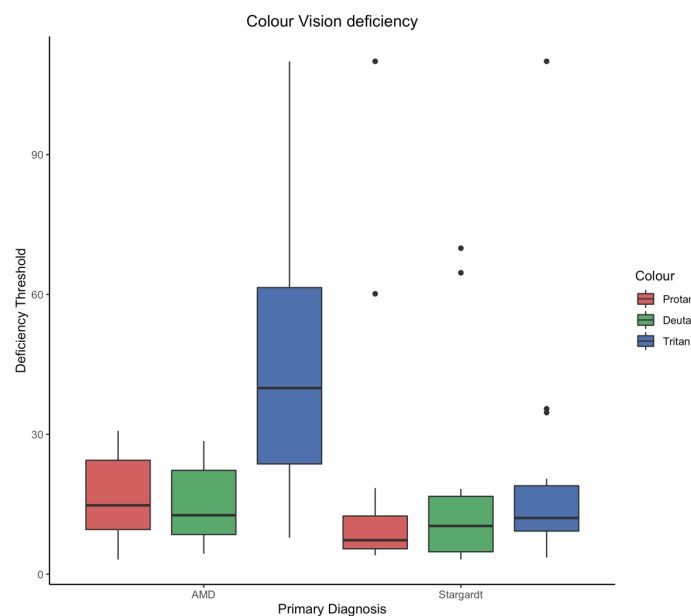


Figure 7-9 Colour vision deficiency by primary diagnosis – AMD and Stargardt. The upper and lower boundaries of the coloured boxes are the inter-quartile range. The median is the horizontal line within the box. The error bars represent 1.5 times the Interquartile range.

7.2.2.8 *Hand to Head movement*

Positional information was collected during the VR reading test, which was used to calculate a correct reading acuity. For each sentence read, head and hand positions within the virtual environment were continuously recorded. The mean distance across each sentence was chosen to calculate the reading acuity; the standard deviation was used to measure the amount of movement.

It is possible to make the following observations within this data: the mean distances for all subjects were $M=376$, $SD=90$ mm. The average reading distance correlated inversely with visual acuity [Pearson's $r = -0.65$, $p<0.001$]. During reading, movement was measured as the standard deviation, which positively correlated with the presented text's visual angle [Spearman's $\rho = 0.41$, $p<0.0001$]. The larger the text, the more movement; this was likely due to the limited visual span, as acuity reduced, and more movement was required to read the sentences.

7.2.3 Discussion

This study's primary objectives were to investigate whether a head-mounted display positively impacted reading performance and whether presenting different colour combinations of text and background had differing effects on reading performance.

When accounting for the display's resolution, the VR display had no significant adverse impacts on reading and visual acuity. Previous work has focused on what improvement this technology can make where the text is magnified using a camera. Where benefits were reported, no information was presented regarding the magnification factors used; thus, the

visual performance of the display technologies used is unclear (Culham et al., 2004; Deemer et al., 2019; Wittich, Lorenzini, et al., 2018).

This experiment measured the display technology's impact using identical text sizes, which allows one to ascertain a theoretical best performance. I deliberately used a virtual environment and calibrated the stimulus to match real-world reading as close as possible to avoid any other confounding factors. As far as I am aware, there have been no direct comparisons of reading with VR head-mounted displays and traditional paper or digital formats.

I also have identified a previously unreported potential advantage with technology, the fixed accommodation and how this can enable some participants to read text they once could not. Therefore, the device has shown the capacity for patients to use this at any age, as accommodation is not a significant issue. The fixed accommodation plane of current head-mounted displays is often seen as a disadvantage and a potential source of discomfort due to the vergence-accommodation conflict (Hoffman, Girshick, & Banks, 2015). However, here it can aid individuals with low vision. At this point, the possibility of discomfort originating from this conflict cannot be excluded. During this study, no participant complained of any discomfort from using the head-mounted display.

Text Presentation Format

The impact of the different text presentations on reading performance was significant but not consistent. The majority of participants objectively benefited from one particular colour combination over another. The two trends identified are individuals with lower visual acuity

>1.0 logMAR had a preference, both objective and subjective, towards inverted contrast (black background). These results follow what Legge et al. (G E Legge & Rubin, 1986) found with their psychophysics study. However, for people with VA from 0.3 to 1.0 logMAR, there was no consistency. Colour vision deficiency was predictive of the text presentation format. Participants with a higher overall deficit (magnitude of colour vision loss) were more likely to prefer white-on-black over black-on-yellow, with a general trend towards the inverted contrast.

I observed a difference in how colour vision deficit manifests with different diseases. AMD is more typical of Tritan deficiency (yellow-blue) and was reflected in the text presentation preference results, consistent with the understanding of colour vision deficiencies and AMD (Vemala et al., 2017). No AMD participant preferred the yellow-on-black either subjectively or using the objective measures.

The colour information in the chosen presentation could be attenuating the contrast in different ways. This may explain why patients have a strong preference, but it varies from individual to individual, particularly with this heterogeneous cohort. Another important factor to consider was the displays were not calibrated, and the stimulus was not isoluminant. Isoluminant means the luminance of the different colours was consistent, varying only the hue and saturation. The lack of preference for the yellow-on-black presentation could be from the overall lower luminance of the yellow text when contrasted to the white rather than anything related to the wavelength.

The choice of test presentation formats (yellow as text or background colour) was made to replicate what is commonly seen on signage and correspondence. I chose to limit the number of presentations to four to reduce the testing time. Another limiting factor was the sentences used; all were taken from the Radner test as they have been carefully selected and standardised. The Radner test contains 42 sentences in total, and with only four repeats of the test, a few participants re-read some sentences. It is possible that by limiting the colour presentation to four choices, the potential of this modification was not fully explored for all participants in this study. However, increasing the number of trials would require more standardised sentences or control for learning effects.

Virtual Reality and Vision Testing

In addition to assistive reading, an important and valuable role for virtual reality, head-mounted displays are in testing vision, particularly in a real-world setting. The findings from the experiments reported here indicate that using a virtual reality head-mounted display is a viable alternative to the physical Radner reading test. It allows for much more accurate viewing distance correction and a more straightforward testing protocol. The dynamic nature of the stimulus could aid with further research into the font and different image enhancement concepts. Another significant advantage is the display's occluded nature; it maintains constant illumination without having to use a lighting-controlled room.

Virtual reality head-mounted displays have great potential for mobile vision testing, effective for automated perimetry (Hollander et al., 2000; Tsapakis et al., 2017; Wroblewski, Francis, Sadun, Vakili, & Chopra, 2014). The main limiting factor is the display resolution. However, as technology improves, this will become less of an issue. Head-mounted displays

are becoming available near the limit of human visual acuity, e.g. Varjo VR-1 (www.varjo.com).

Wearable Displays for image enhancement

In the context of wearable displays and how they can improve reading, these results signify that outside of ergonomic and user interface issues, wearable displays can provide a benefit. They can provide large high-contrast readable text. The major problem with distance reading and these devices is that significant zoom factors are required, exaggerating any head movements. Also, lower VA requires higher magnification and, as observed in this study, is related to more significant head/hand movements. This is most likely due to the participants' requirement for such a large font size that they have to move their head to read a complete sentence, as the text exceeds their visual span (Gordon E. Legge et al., 2007). The challenge now is for wearable display designers to overcome the user interface issues currently limiting performance.

7.2.4 Conclusion

This experiment has shown that a head-mounted display has no adverse impact on reading performance when accounting for the display resolution. Previous work has not investigated the isolated effect of display technologies on vision. I have shown that human interface factors only limit wearable displays' performance as a low vision reading aid, and future developments must focus on this.

The heterogeneous nature of the retinal conditions in this cohort has made it challenging to identify trends regarding text presentation preference. Still, this cohort was chosen as representative of the typical patient population with central visual loss. Despite this, the clear benefit of using a modification of text presentation was observed for the majority of the study participants.

These findings could also help decision making regarding the best presentation of text by low vision services and charities for patients with visual impairment. Blanket choices should be made with caution, and any modification of text presentation ideally needs to be user-driven. However, customising text presentation to an individual may not always be an option. Thus, a recommendation from this experiment is to avoid black text with a yellow background for AMD sufferers and to offer inverted contrast if available.

The wearable HMD shows real potential as a mobile, low-cost vision testing device and warrants further development.

7.3 How vital is the display size?

A question initially posed by some of the early work with this particular HMD display was the display size, the field-of-view. AR devices usually have a significantly smaller field of view than their occluded VR counterparts. Subjectively, with face perception, this was found not to be of primary concern. However, reading is different.

This question was not directly answerable, as I used a VR display; however, using work by Atilgan et al. (Atilgan, Xiong, & Legge, 2020) it is possible to estimate the usability. Atilgan

examined the hypothesis that a critical number of characters is required for an individual to achieve their maximum reading speed, referred to as the critical character count (CCC). This count can be used to determine the minimum display size for effective reading.

7.3.1 Display size and critical character count

Atilgan et al. found that for comfortable reading, a CCC of 9 characters per line is required. Their work presented how this relates to display size and the ranges of visual acuity they will be suited.

Following on from this work, I applied the same approach to wearable displays, using the Epson BT-350 as an example.

Typically, manufacturers describe HMD size in visual angle rather than inches. For the case of the Epson display, only the diagonal visual angle is reported. This can be decomposed into its horizontal and vertical components using the displays aspect ratio.

For a diagonal visual angle of 23deg and an aspect ratio of 16:9, the horizontal and vertical visual angles are approximately 20 and 11 degrees, respectively.

For the BT350 display, the maximum CPS value for the Arial font is 1.45 logMAR, and for the courier font, 1.30 logMAR. All calculations are described in Appendix 3.4.

Knowing an individual's CPS is uncommon; VA is not so. Atilgan used linear regression to describe a relationship between CPS and VA;

$$CPS = 0.39 + 0.95 * VA$$

This is with an N=87 and is slightly different to the coefficient values from this study (figure 7-3), the slope is similar (0.98), but the intercept is different (-0.31). This is likely the smaller number of participants and less homogenous distribution, with the Atilgan study having a more even distribution of participants.

Using this relationship, the maximum VA that can theoretically achieve their CPS is:

$$VA = \frac{CPS - 0.39}{0.95}$$

For the Epson BT350 display displaying Arial font, the lowest VA is 1.11. Any lower than that, the horizontal size of the display limits the CCC. Table 7.4 shows the calculations for other commercial HMDs.

Table 7-4 Effective reading acuity range of commercial HMDs.

Product Name	Effective reading acuity range				
	eSight 3	GiveVision SightPlus	OxSight Crystal	IrisVision	NuEyes Pro
Best visual acuity (logMAR)	0.34	~0.5	0.0	0.5	0.14
Maximum VA CCC	1.72	2.05	1.45	1.99	1.62
CPS CCC	1.39	1.74	1.11	1.68	1.29
VA range (logMAR)	0.34 -> 1.39	0.5 -> 1.74	0 -> 1.11	0.5 -> 1.68	0.14 -> 1.29

VA = Visual Acuity, CPS= Critical Print Size, CCC=Critical Character Count

7.4 Discussion

The results of the reading experiment and prior work show a potential for head-mounted displays, but with some limitations. The use of a camera attached to the head is hard to control. While I have shown different presentations can improve reading acuity and reading speed for some participants, all participants will still require an element of magnification.

As discussed by Culham et al. (Culham et al., 2004) handheld magnifiers are currently the best tool for close reading due to their simplicity and ease of use. The evidence presented in this experiment, where large head-hand movements occurred, and the difficulties participants encountered when using a head attached camera, supports this. The currently available commercial technologies show that the cost to performance benefit is poor for close reading compared to traditional methods. This is predicted to change over time with further innovation and modification.

HMDs are most likely to provide a significant benefit over traditional counterparts with distance reading or in hand-busy environments. In discussion with the participants involved in this experiment, several stated that the HMD could be very useful for viewing timetables, street signs and shop names. Some of the working-age participants indicated that they could be beneficial for meetings, viewing presentations.

The principal advantage HMD provide is their digital nature. They are or can be attached to powerful computational devices. Leveraging this key to the success of HMDs for reading.

For a head-mounted display to be successful, the advantage it provides must not be outweighed by its limitations. The most significant of these are coordination issues.

7.4.1 Coordination issues.

As magnification increases, motion is exaggerated, and control difficulty increases.

Mitigating the impact of this motion is vital to providing a positive user experience.

7.4.1.1 *Photographs*

A simple approach that Iris Vision and OxSight Crystal have implemented is to pan around a photograph. This is where a user takes a picture of an object of interest, and they can zoom in and pan around using the accelerometer and gyroscope built into the HMD. This provides a much more stable viewing experience, as the movement is not amplified with zoom factor. This method requires capturing a good initial image, simple for static objects, such as signs and timetables, but is harder for moving targets such as bus numbers.

The use of smartphone cameras is a standard tool for the visually impaired. Most, if not all of the younger participants (less than 30 years old) who took part in this study described using the camera on their smartphones to aid with reading. They capture an image, zoom in and pan around to collect the information they are searching for.

7.4.1.2 *Image stabilisation*

Image stabilisation is an integral part of cinematography, with most approaches using optical or mechanical methods. Modern smartphone manufactures are increasingly using these techniques to improve photography. Attempts have been made to use image stabilisation for the visually impaired and distance viewing. Gang Luo developed an iPhone app with image stabilisation (Z. Li, Pundlik, & Luo, 2013), improving reading for both normally sighted and low vision participants (Luo & Pundlik, 2014).

Implementation of image stabilisation was applied to the system described in chapter 5. The method was to utilise the cropped space on the image when zoomed.

The process was to; detect features within an image, tracking them to the next frame received. The motion between the frames was calculated using an optical flow algorithm. A Kalman filter was then used to smooth the movement, similar to the process described and used in chapter 5. The resulting smooth transformation was used to crop the image and display it to the user.

A few participants tried the image stabilisation, while there were no quantified measurements from the participants, they reported the method perceptually improved the experience of high zoom factors. The downside to this method is its computational expense. Feature matching and tracking is a time-consuming operation and significantly slowed the frame rate and increased the display latency. Thus, the assessment with visually impaired participants was only used as a proof of concept.

7.4.1.3 Automatic text detection

The aforementioned image stabilisation approaches reduce the effect of high-frequency motion. This can significantly aid individuals, as found by Luo (Z. Li et al., 2013), but assumes that the individual can localise any target text. Visual search is known to be inhibited with reduced vision (Caldani, Chatard, Wiener-Vacher, & Bucci, 2019; Taylor, Smith, & Crabb, 2017), and automatically highlighting candidate text areas in a scene can be beneficial, i.e. in situations where an individual is searching for a direction or sign but does not know its exact location.

Text detection in real-world scenes is challenging and has been the focus of much computer vision research. A review by Han et al. (Y. Zhu, Yao, & Bai, 2016) covers the development and advances, with the deep learning approaches performing best. Examples are EAST: CNN (Zhou et al., 2017) and Textboxes++ (M. Liao, Shi, & Bai, 2018) which can reliably detect multiple orientation text in a scene. The runtime performance is increasing, with deep learning approaches typically much more computationally complex, at 10 to 15 frames per second.

Some work has been performed with the visually impaired in mind. Fei et al. (Fei et al., 2018) developed a system that spoke text to users—their method employed object detection methods, text detection, text recognition, and speech synthesis. No assessments in patients with visual impairment have been reported by this group. Pazio et al. (Pazio, Niedźwiecki, Kowalik, & Lebień, 2007) produced a similar system that read text aloud, but no experimental evidence was provided. Joan and Valli (Joan & Valli, 2017) and Ezaki et al. (Ezaki, Bulacu, & Schomaker, 2004) proposed text detection for the visually impaired; again, there is no published experimental work on the visually impaired.

Despite the number of publications that have stated visual impairment as the intended use of text detection methods, there is very little work regarding visually impaired participants. The methods proposed all have merit; however, significant practical work is required to understand the impact of such methods on a patient population.

The output of the text detectors can also be used for image stabilisation, similar to the method I used in chapter 5, where the face detector was used to stabilise the image.

7.4.2 Low-light situations

An often-described problem is reading in low-light environments. This typically happens in restaurants and bars or other people's homes where the lighting is not optimised for someone's level of vision.

Low-light is also a challenge for digital cameras. The dynamic range of a camera is typically much lower than that of the human visual system. Thus, auto-gain features are essential in digital photography, where the signal is electrically amplified. Traditional solutions for this situation are large light-capturing sensors and lenses, but they are unsuitable for wearable applications. Form factor limitations, driven by the smartphone industry, has developed sophisticated software approaches, often referred to as computational photography. They can improve image quality using various methods: image stacking, high dynamic range imaging (Hasinoff et al., 2016) and super-resolution (Walt, 2010).

7.4.3 Text to speech

Text-to-speech technologies have been available for some time, mainly in the form of computer screen-readers where the input text is known. With advances in computer vision and the improvement of text-recognition methods (Kaushik & Verma, 2018), text-to-speech in more unconstrained situations is achievable and reliable.

An online survey on assistive technologies taken by 133 low-vision participants, Yueh-Hsun et al. (Y.-H. Wu, Granquist, Gage, Crossland, & Legge, 2020) found that visual reading is of continued importance. This was mirrored in anecdotal discussions with participants in this

study. A device that reads to you is a different experience than a device that enables you to read. Participants stated that they preferred the latter of the two.

7.5 Conclusion

Head-mounted displays and colour conversion can aid individuals with reading. The advantages gained appear to be quite individual; thus, any enhancement system must have some degree of customisation. There are still significant limitations that currently inhibit adoption, but with focused development on the human-device interface, these can be overcome.

Chapter 8

General Discussion

My contributions to this chapter: The writing is my original work.

Other contributions: Professor Robert MacLaren and Professor Susan Downes assisted with proofreading.

8 General Discussion

The work described in this thesis aimed to improve reading and face perception for the visually impaired using augmented reality.

New face perception test

The first section of this thesis focused on face perception, with chapters 3 & 4 focused on assessing a new image enhancement method. I developed a new face perception test to evaluate the image enhancements – the Oxford Enhanced Vision Assessment Tool (OVAT) - as existing tests were not fit for purpose. This novel test was validated with a normal population, and repeatability scores were ascertained. The OVAT was well-received by participants, quick to perform, conceptually simple, but sensitive enough to differentiate between normally sighted participants. Limitations in recruiting older participants biased the age distribution of the normal group. Although no significant performance correlations were made with age, further assessments with a more ageing population would be ideal.

The advantages this new test provides over alternatives include photorealistic stimuli, no memory component is required, and a large sensitivity range is provided. In addition, I have devised a new method for measuring the stimulus. All face perception tests assume that the stimuli used are either pre-calibrated or of similar difficulty. Prior approaches involved using synthetic stimuli, which significantly differ from real faces (Logan et al., 2016), or careful calibration and selection of the stimulus (Bowles et al., 2009) which is labour intensive and time-consuming. I have introduced the stimuli measuring method derived from the computer vision research community based on face recognition. In computer face

recognition, the general approach is to describe a face using a specific descriptor and compare it to a reference. I used a descriptor generated by FaceNet (Schroff et al., 2015) which is robust to head pose, and lighting and can be computed quickly on any image containing a face.

The significance of this new stimuli measuring method is that it enables researchers to quickly use alternate face stimuli without going through a time-consuming calibration process. It also allows for more realistic, real-world stimuli to be used, with little curation of the stimuli, as the descriptor is designed to work for images of a face in all poses.

The OVAT has been further assessed by fifty young, normally sighted participants as part of a study by Jones et al (Jones et al., 2020). The test was included as part of a tablet-based vision test and was shown to work well, demonstrating its potential, for example, in the quick screening of patients in waiting rooms.

A challenge identified when attempting to distribute and share the face perception test is the ownership of the stimulus used. The specific dataset I used is available for academic research only and cannot be distributed. I have publicly shared the source code** for the test for other researchers to use but cannot share the face stimuli. The personal nature of the stimulus adds a complication when creating these datasets. Adhering to privacy laws can severely limit data collection and distribution; for example, the General Data Protection Regulation (<https://gdpr-info.eu>).

** <https://bitbucket.org/iainrwilson/facediscrimination/src/master/>

A solution to this is found in the latest research on generative adversarial networks (GANs) (Goodfellow et al., 2020). GANs are a type of neural network that can generate new content that did not exist before. The idea is based on two neural networks competing against each other; a generator that converts random noise into a coherent image and a discriminator that predicts if the generator's output is from the original dataset or generated. This competition helps them to mimic any distribution of data (Elgendy, 2020).

Modern GANs can generate high-quality face images (figure 8.1) that do not represent an actual individual, thus alleviating any privacy problems. Aligned with the stimuli measuring method I introduced, a test could conceivably be developed that dynamically generates face stimulus and calculates their difficulty without any limitations on the stimulus.



Figure 8-1 None of these people are real; they are generated by StyleGAN (Karras et al., 2020). Source: <https://github.com/NVlabs/stylegan>

Limitations of the face perception test

The OVAT is not without limitations. While still generally short, 7 minutes on average, the time to perform the test could still be reduced. The method uses a four-alternate forced choice, which could be optimised and explored further. Increasing the number of options, e.g. presenting eight faces at once, will reduce the test time and be used on a larger display. While visually more realistic than prior methods, the stimuli I used did not represent the distribution of faces commonly encountered. I limited the skin tone to white Caucasians to remove any colour cues. More experimentation could be performed to understand the importance of skin tone, particularly with the new stimuli measuring method and the GANs mentioned earlier, to explore a dataset more representative of life.

Ultimately the test is contrived; static images are not how we mostly see faces. Most of the visual world is dynamic. There is evidence that the visual system takes advantage of dynamic information when a static image is lacking (Dobs, Bülthoff, & Schultz, 2018). Future development of a face perception test that utilizes dynamic stimuli, whilst challenging to create, would be highly relevant in assessing and improving face perception in real-world scenarios.

Enhancing face perception

Utilizing OVAT, in chapter 4, I assessed two image enhancement methods on a patient population; the Cartoonisation method and an existing method by Peli et al. (T. Peli & Lim, 1982). The results showed that the majority of participants (60%) did not show any improvement by scoring better on the face perception test. However, a significant portion (40%) did gain an advantage. For participants who improved, the image enhancement

methods both provided an advantage, with the Cartoonisation method performing the best. The mechanisms for the improvement are most likely attributable to a contrast enhancement. The Cartoonisation method maximized the contrast in an image while removing redundant (and potentially confusing) information. The alternate method only amplified selected frequencies by an amount, leaving potentially confusing information.

The statistical correlation with the functional data collection (contrast sensitivity and VA) agrees with other authors' work (Alexander, 1988; Barnes, De L'Aune, et al., 2011; McCulloch et al., 2011; Taylor, Smith, et al., 2018). The downside to the Cartoonisation was the unconventional presentation; a significant number of participants noted that they found viewing faces with this type of enhancement applied disconcerting. An improvement to my approach would be to include colour information and use the binary cartoon filter to amplify features of the face. A further extension, performed by Peli et al. (Eli Peli et al., 1994), would be to tune the enhancements based on an individual's sensitivity loss. The limitation of the Peli et al. study was the face perception test used was not well designed and an untrustworthy outcome measure. Utilizing the more robust new test and further experimentation with a patient population could yield more conclusive results.

I investigated the possibility of predicting face perception performance using the answers to questionnaires. The results were promising and indicated that I could predict who would benefit from the visual enhancement using less than ten questions. The results need validation with a large cohort, but it could significantly impact caregiving and advice for people with central vision loss.

A limitation of the experimental design was the length of the protocol. The OVAT is cognitively demanding; the adaptive nature of the test meant that regardless of visual ability, the test was equally difficult. The patient participants repeated the test 3 times, which along with the other visual function tests (visual acuity, contrast and colour sensitivity), was tiring, particularly for the more elderly patients. I made efforts to make participants as comfortable as possible, with frequent breaks and refreshments offered. However, some missing data points occurred where the room booking time was exceeded, or the participants were unwilling or unable to continue.

With the time restrictions, I could not perform repeatability measurements of the OVAT tool with the patient population. Repeatability values were collected during the development of the OVAT, and I assumed that this would translate to the patient population. There is evidence of higher test-retest variability of visual acuity tests in patients with age-related macular degeneration (Patel et al., 2008a). It is currently unknown if this variability translates to face perception and could be the subject of future study.

Real-time image enhancement

Chapter 5 described the development of a real-time image enhancement system that integrated the Cartoonisation image enhancement method into a head-mounted display. As part of the development, I had to carefully choose the tools based on the mobile nature and real-time performance goals. I introduced an online learning approach to skin detection, which was lightweight and achieved high performance in accuracy and speed. I integrated this skin detector into a complete face segmentation system and assessed it with a small patient cohort.

The participants overwhelming did not like the image enhancements. This result was not surprising, as the subjective responses in chapter 4 found most participants thought they performed worse with the Cartoonisation method. However, they frequently performed better, implying that while helpful the Cartoonisation was less comfortable to view. This finding indicates that lab-based experiments will not necessarily lead to the best experience for users in the real world and highlight that user input in the design and development of image enhancements is essential. However, some features of the real-time enhancement system were appreciated, such as face cropping and tracking/stabilisation. Incidentally, the face tracking system was implemented from pilot participant feedback and was well received.

Perhaps if the participants had been allowed to spend more time field-testing the device in their home environment, they might have noticed an advantage. A take-home study was planned but hindered by several factors; gaining ethical approval was time-consuming, low participant uptake, and limited hardware availability. The low uptake was due to the low-quality camera on the simpler HMD. I only built one prototype with a higher quality camera, which was unsuitable for unsupervised use.

The experimentation in chapter 5 was limited to a subjective discussion; it was impractical to repeat the earlier face perception tests with the HMD. Using the HMD with a computer monitor and the OVAT created coordination issues and dramatically increased the test time. Moving the focus of the camera between identities on the screen was slow. The face stabilisation feature did not help, as it is designed to fix the participant's view onto a face and did not support multiple faces nearby. A solution would be to expand the face

stabilisation to allow the user to control the selected face manually in a scene, e.g. cycle through the identities with a button.

All of the participants in the study were shown a simpler version of the head-mounted displays, i.e. one without any image enhancement functionality. This device presented a video view from an attached camera with digital zoom. Subjective responses from participants are reported in chapter 5. The device was well received, with most participants stating that they thought it could be beneficial, but the primary limitation was the quality of the camera. Thus, a higher-quality camera was used with the real-time image enhancement version and indeed was very well received compared to the lower quality camera. Ultimately the participants wanted to zoom in and view hard-to-see objects, and the higher-quality camera enabled this to a greater degree.

The next step to assess this device further would be to include a take-home element using this simpler version. The results from the subjective discussions in chapter 5 are likely biased, as participants are likely to be overly positive. Future work integrating the developed algorithms into a device that users could use at home and feedback regarding their experience would provide a better insight into the impact of the image enhancements.

Even though the participants did not receive the image enhancements well, the real-time face segmentation system performed well. The approach I developed can be used to apply any image enhancement algorithm to a face in real-time. The methods and implementation details outlined in chapter 5 can enable future research into face perception enhancement on mobile devices.

Display Size

One of the questions inspiring this thesis was the size of the displays used. Prior work by myself and van Rheede et al. (Joram Jacob Van Rheede et al., 2016), using a similar Epson display used in this thesis, hypothesized that the poor performance in the central vision loss group was partly due to the small display size. However, I found this hypothesis to be inaccurate. The qualitative discussions with participants found that the display size was not limited when using the device for face perception. Investigations I performed in chapter 7 also concluded that reading the minimum visual acuity for comfortable reading included 75% of the patients involved in the study.

A significant challenge with designing HMD see-through displays is maintaining a large field of view. Large fields-of-view, 90 degrees, are common in occluded HMDs, but a field-of-view of 40 degrees is considered large for a see-through HMD. This knowledge will enable designers and developers to be more confident in their choice of hardware.

At-home monitoring

Visual acuity tests were performed using the Epson HMD in chapter 6, and the results showed a good agreement with the standard ETDRS charts. The wearable nature of the HMD creates a simple measurements system without the need for large charts and testing rooms.

Further work is required to validate fully as I only collected single binocular measurements and no repeated measures, as the goal of chapter 6 was to understand if participants could

read text on the Epson HMD. However, the results indicate that the HMD has potential for at-home monitoring of visual acuity. Participants could perform periodic measurements and be informed to seek medical help if any dramatic change in visual acuity is detected. It can also be used as a tool for research into longer-term disease progression.

Reading

The second section of this thesis focused on reading and how a wearable display can improve it. The control issues found when using the HMD in chapter 5, where patients struggled to direct the camera, led me to use a virtual reality display. I implemented a reading test in a virtual environment, following the design and protocol of an existing paper test. Using a VR display enabled me to decouple the control issues from the image enhancements. This allowed me to assess how the display technologies interact with reading and visual impairments.

In chapter 7, using an established reading assessment chart, I compared paper-based reading to VR reading and found no negative impact on reading ability. I further assessed the effect of different text and background colours. Different colours positively impacted; most individuals had a preferred colour combination, both subjectively and objectively.

I identified a correlation between colour vision deficit and preferred text/background colour. The greater the colour vision deficit, the more likely an individual would prefer a black background. The black text on a yellow background is the least preferable for someone with a large colour vision deficit. No participant with a primary diagnosis of AMD preferred black text with a yellow background. This was likely due to the tritan deficiency

common with AMD (Vemala et al., 2017). The AMD cohort is small; however, the observations are consistent with the literature.

The difficulty with finding any factors correlating to performance highlights the patient cohort's heterogeneous nature and how complex and multifactorial influences are involved in reading and vision loss. The important message is that one device with a specific modification is unlikely to benefit all users. Thus, any enhancement system must consider its user as an individual with specific requirements.

Further work could be carried out on a more focused cohort, i.e. with a singular diagnosis or consistent colour deficits.

Virtual Reality

The use of VR as a visual assessment tool was a secondary outcome of this thesis. It was well-received, is simple to use, and has the potential for enabling future work.

A VR environment can be used to simulate an extensive range of situations. VR has a positive impact on clinical psychology research (Riva & Cattolica, 2020), is used in anxiety research and is helpful as a therapeutic tool (Freeman et al., 2018). VR has been used to assess the impact of different visual conditions Chow-Wing-Bom et al. (Chow-Wing-Bom, Dekker, & Jones, 2020) used a VR environment to simulate vision loss to understand the effects on visual search better.

A significant barrier, which I identified during my work here and has undoubtedly been the case for others, is validating image enhancement methods in a visually impaired population. There are numerous difficulties involved, including access to patients, ethical approvals, and suitable safe testing locations, all of which inhibit research. These difficulties are expressed in the consistently small patient cohorts used in related studies.

Assessing different display technologies and enhancement methodologies has a significant technical barrier to entry. Developing efficient image processing software for mobile devices is a specialized task. Simple image processing tasks that can be performed effortlessly on a desktop computer can require optimization before using on a mobile device. Developing hardware for new display ideas has perhaps an even higher barrier to entry, leaving the most to rely on hardware manufacturers to build displays that often have substantial price tags and may not be fit for purpose for the visually impaired.

Chapter 7.4 discusses the limitations and feasibility of HMD for reading and introduces some potential solutions. However, a fully functional fit for purpose HMD will come with a significant development cost to implement on a wearable display, particularly one that can be used in the home or outside a testing environment.

The virtual reality environment and reading assessment tool developed in this thesis can be used in conjunction with software implementations vis a vis the suggestions in chapter 7.4. This will be very helpful for others working in this field and speed future research significantly. Using a VR environment will remove the hardware limitations of building new displays and simplify the software development process. This will enable researchers to

assess and further develop image enhancement ideas quickly. Building on this will allow a better understanding of the behavioural limitations, such as image stabilization and user interfaces limits.

In addition, eye tracking is a feature available in some VR systems; this enables a more accurate visual impairment simulation, as presented by Jones and Ometto (Jones & Ometto, 2018). These simulations will allow early ideas to be assessed for feasibility in a normally sighted population before assessment in a visually impaired population.

Augmented reality and the visually impaired

Augmented reality and wearable display have a clear role as aids for the visually impaired. The technology is still in its infancy and is constantly evolving, becoming smaller and more accessible.

The currently available options are expensive and relatively simple in what functionality they provide. At the time of writing, the cheapest device is over £1000, and all attempt to be a general tool. This generalisation is likely driven by the cost of development and the effort to increase their market size.

Not all participants found the experience of wearing a version of HMD to be positive, but a significant number of participants did. However, this does not directly translate to helpful. In the simplest form, they are wearable digital magnifiers, and this feature is what most individuals desire. Herein lies probably the most significant barrier. All HMDs rely on a

camera, and that camera is attached to the display. Controlling this camera and focusing on objects of interest is unnatural and challenging.

The role of image enhancements can negate the need for some magnification, improving the visibility so that magnification may not be required or the magnitude reduced. The methodologies I have presented, face tracking and background removal can help overcome the camera control barriers and improve the user experience.

During conversations with the visually impaired participants I have met during the work presented in this thesis, they simply desire to see the world like a normally sighted person, free from cumbersome aids or relying on others. Augmented reality and wearable displays have the potential to provide this experience. The hardware specifications (display and camera resolution) and computational ability of current HMDs are capable of achieving this goal, but the user interfaces are not. Focused work on improving the user experience should be the priority for researchers and manufacturers.

8.1 Conclusion

Herein I have completed the two aims of my thesis. Firstly: to improve face perception in a portion of a visually impaired population by utilizing a new image enhancement method. To address this, I developed a mobile real-time image enhancement system that can selectively apply an image enhancement to maximize its effectiveness using an augmented reality display. Secondly: To show how a wearable display can improve reading for the visually impaired using a virtual reality display, which I achieved through a series of experiments.

I also developed a face perception test that will positively impact others in this specific research field. The virtual reality vision testing environment I created shows promise as a tool to accelerate research into vision enhancements.

Bibliography

- Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, 182(April 2018), 73–83. <https://doi.org/10.1016/j.cognition.2018.09.002>
- Abudarham, N., & Yovel, G. (2016). Reverse engineering the face space: Discovering the critical features for face identification. *Journal of Vision*, 16(3), 40. <https://doi.org/10.1167/16.3.40>
- Abudarham, N., & Yovel, G. (2019). Same critical features are used for identification of familiarized and unfamiliar faces. *Vision Research*, 157(February 2018), 105–111. <https://doi.org/10.1016/j.visres.2018.01.002>
- Adelstein, B. D., Burns, E. M., Ellis, S. R., & Hill, M. I. (2005). Latency Discrimination Mechanisms in Virtual Environments: Velocity and Displacement Error Factors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(26), 2221–2225. <https://doi.org/10.1177/154193120504902601>
- Aguilar, C., & Castet, E. (2017). Evaluation of a gaze-controlled vision enhancement system for reading in visually impaired people. *PLoS ONE*, 12(4), 1–24. <https://doi.org/10.1371/journal.pone.0174910>
- Al-Atabany, W. I., Memon, M. a, Downes, S. M., & Degenaar, P. a. (2010). Designing and testing scene enhancement algorithms for patients with retina degenerative disorders. *Biomedical Engineering Online*, 9, 27. <https://doi.org/10.1186/1475-925X-9-27>
- Al-mohair, H. K., Mohamad-saleh, J., & Suandi, S. A. (2013). Impact of Color Space on Human Skin Color Detection Using an Intelligent System. *Recent Advances in Image, Audio and Signal Processing*, 178–187.
- Alexander, M. F. (1988). Assessment of Visual Function in Patients With Age-Related Macular Degeneration and Low Visual Acuity. *Archives of Ophthalmology*, 106(11), 1543. <https://doi.org/10.1001/archophth.1988.01060140711040>
- Allman, J; Miezin, F; McGuinness, E. (1985). RESPONSES FROM BEYOND FIELD : Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons.
- Amore, F. M., Fasciani, R., Silvestri, V., Crossland, M. D., de Waure, C., Cruciani, F., & Reibaldi, A. (2013). Relationship between fixation stability measured with MP-1 and reading performance. *Ophthalmic and Physiological Optics*, 33(5), 611–617. <https://doi.org/10.1111/opo.12048>
- Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications*.
- Arden, G. B., & Wolf, J. E. (2004). Colour vision testing as an aid diagnosis and management of age related maculopathy. *British Journal of Ophthalmology*, 88(9), 1180–1185. <https://doi.org/10.1136/bjo.2003.033480>
- Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research*, 45(23), 2926–2933. <https://doi.org/10.1016/j.visres.2005.06.013>
- Atilgan, N., Xiong, Y. Z., & Legge, G. E. (2020). Reconciling print-size and display-size constraints on reading. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48), 30276–30284. <https://doi.org/10.1073/pnas.2007514117>
- Bach, M. (2007). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis, 965–971. <https://doi.org/10.1007/s00417-006-0474-4>
- Bailey, I. L., & Lovie, J. E. (1980). THE DESIGN AND USE OF A NEW NEAR-VISION CHART. *Optometry and Vision Science*, 57(6), 378–387. <https://doi.org/10.1097/00006324->

198006000-00011

- Barnes, C. S., De L'Aune, W., & Schuchard, R. A. (2011). A test of face discrimination ability in aging and vision loss. *Optometry and Vision Science*, 88(2), 188–199. <https://doi.org/10.1097/OPX.0b013e318205a17c>
- Barnes, C. S., De L'Aune, W., & Schuchard, R. A. (2011). A Test of Face Discrimination Ability in Aging and Vision Loss. *Optometry and Vision Science*, 88(2), 188–199. <https://doi.org/10.1097/OPX.0b013e318205a17c>
- Bastawrous, A., Rono, H. K., Livingstone, I. A. T., Weiss, H. A., Jordan, S., Kuper, H., & Burton, M. J. (2015). Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and Community-Based Fieldwork. *JAMA Ophthalmology*, 133(8), 930–937. <https://doi.org/10.1001/jamaophthalmol.2015.1468>
- Bauma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(49), 177–178.
- Beck, R. W., Moke, P. S., Turpin, A. H., Ferris, F. L., SanGiovanni, J. P., Johnson, C. A., ... Kraker, R. T. (2003). A computerized method of visual acuity testing: Adaptation of the early treatment of diabetic retinopathy study testing protocol. *American Journal of Ophthalmology*, 135(2), 194–205. [https://doi.org/10.1016/S0002-9394\(02\)01825-1](https://doi.org/10.1016/S0002-9394(02)01825-1)
- Berkeley, U. C., Joshua, J. J., Peace, E., & Iii, W. J. (2012). The Role of External Features in Face Recognition with Central Vision Loss.
- Bethlehem, R. A. I., Dumoulin, S. O., Dalmaijer, E. S., Smit, M., Berendschot, T. T. J. M., Nijboer, T. C. W., & Van Der Stigchel, S. (2014). Decreased fixation stability of the preferred retinal location in juvenile macular degeneration. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0100171>
- Bieniek, A., & Moga, A. (2000). An efficient watershed algorithm based on connected components. *Pattern Recognition*, 33(6), 907–916. [https://doi.org/10.1016/S0031-3203\(99\)00154-5](https://doi.org/10.1016/S0031-3203(99)00154-5)
- Bittner, A. K., Ibrahim, M. A., Haythornthwaite, J. A., Diener-West, M., & Dagnelie, G. (2011). Vision test variability in retinitis pigmentosa and psychosocial factors. *Optometry and Vision Science*, 88(12), 1496–1506. <https://doi.org/10.1097/OPX.0b013e3182348d0b>
- Black, J. M., Jacobs, R. J., Phillips, G., Chen, L., Tan, E., Tran, A., & Thompson, B. (2013). An assessment of the iPad as a testing platform for distance visual acuity in adults. *BMJ Open*, 3(6), 5–7. <https://doi.org/10.1136/bmjopen-2013-002730>
- Boas, Y. A. G. V. (2013). Overview of virtual reality technologies. *Interactive Multimedia Conference, 2013*.
- Bolme, D., Beveridge, J. R., Draper, B. A., & Lui, Y. M. (2010). Visual object tracking using adaptive correlation filters. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 6, pp. 2544–2550). IEEE. <https://doi.org/10.1109/CVPR.2010.5539960>
- Boucart, M., Dinon, J. F., Despretz, P., Desmettre, T., Hladiuk, K., & Oliva, A. (2008). Recognition of facial emotion in low vision: A flexible usage of facial features. *Visual Neuroscience*, 25(4), 603–609. <https://doi.org/10.1017/S0952523808080656>
- Bourne, R. R. A., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., ... Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9), e888–e897. [https://doi.org/10.1016/S2214-109X\(17\)30293-0](https://doi.org/10.1016/S2214-109X(17)30293-0)
- Bourne, R. R. A., Jonas, J. B., Bron, A. M., Cicinelli, M. V., Das, A., Flaxman, S. R., ... Resnikoff,

- S. (2018). Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe in 2015: Magnitude, temporal trends and projections. *British Journal of Ophthalmology*, 102(5), 575–585. <https://doi.org/10.1136/bjophthalmol-2017-311258>
- Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... Yovel, G. (2009). *Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the cambridge face memory test and cambridge face perception test. Cognitive Neuropsychology* (Vol. 26). <https://doi.org/10.1080/02643290903343149>
- Boykov, Y. Y., & Jolly, M.-P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001* (Vol. 1, pp. 105–112). IEEE Comput. Soc. <https://doi.org/10.1109/ICCV.2001.937505>
- Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of ‘mass’ in line drawings of faces. *Applied Cognitive Psychology*, 6(7), 619–628. <https://doi.org/10.1002/acp.2350060705>
- Bullimore, M., Bailey, I., & Wacker, R. (1991). Face recognition in age-related maculopathy. *Investigative Ophthalmology & Visual ...*, 32(7). Retrieved from <http://www.iovs.org/content/32/7/2020.short>
- Calabrèse, A., Owsley, C., McGwin, G., & Legge, G. E. (2016). Development of a reading accessibility index using the MNREAD acuity chart. *JAMA Ophthalmology*, 134(4), 398–405. <https://doi.org/10.1001/jamaophthalmol.2015.6097>
- Caldani, S., Chatard, H., Wiener-Vacher, S., & Bucci, M. P. (2019). Visual searching capabilities in Age-Related Macular Degeneration (AMD) subjects. *Applied Neuropsychology:Adult*, 0(0), 1–8. <https://doi.org/10.1080/23279095.2019.1678158>
- Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, 41(2), 195–202. [https://doi.org/10.1016/S0028-3932\(02\)00149-5](https://doi.org/10.1016/S0028-3932(02)00149-5)
- Calle, M. L., & Urrea, V. (2011). Letter to the editor: Stability of Random Forest importance measures. *Briefings in Bioinformatics*, 12(1), 86–89. <https://doi.org/10.1093/bib/bbq011>
- Chai, D., & Ngan, K. N. (1999). Face segmentation using skin-color map in videophone applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(4), 551–564. <https://doi.org/10.1109/76.767122>
- Chaves-González, J. M., Vega-Rodríguez, M. A., Gómez-Pulido, J. A., & Sánchez-Pérez, J. M. (2010). Detecting skin in face recognition systems: A colour spaces study. *Digital Signal Processing: A Review Journal*, 20(3), 806–823. <https://doi.org/10.1016/j.dsp.2009.10.008>
- Cheng, H. D., Jiang, X. H., Sun, Y., & Wang, J. (2001). Color image segmentation: Advances and prospects. *Pattern Recognition*, 34(12), 2259–2281. [https://doi.org/10.1016/S0031-3203\(00\)00149-7](https://doi.org/10.1016/S0031-3203(00)00149-7)
- Cheong, A. M. Y., Legge, G. E., Lawrence, M. G., Cheung, S. H., & Ruff, M. A. (2008). Relationship between visual span and reading performance in age-related macular degeneration. *Vision Research*, 48(4), 577–588. <https://doi.org/10.1016/j.visres.2007.11.022>
- Chow-Wing-Bom, H., Dekker, T. M., & Jones, P. R. (2020). The worse eye revisited: Evaluating the impact of asymmetric peripheral vision loss on everyday function. *Vision Research*, 169(October 2019), 49–57. <https://doi.org/10.1016/j.visres.2019.10.012>

- Chung, S. T. L. (2002). The effect of letter spacing on reading speed in central and peripheral vision. *Investigative Ophthalmology & Visual Science*, 43(4), 1270–1276.
<https://doi.org/10.1016/j.visres.2013.03.005>
- Chung, S. T. L. (2011). Improving reading speed for people with central vision loss through perceptual learning. *Investigative Ophthalmology and Visual Science*, 52(2), 1164–1170. <https://doi.org/10.1167/iovs.10-6034>
- Chung, S. T. L. (2020). Reading in the presence of macular disease: a mini-review. *Ophthalmic and Physiological Optics*, 40(2), 171–186.
<https://doi.org/10.1111/opo.12664>
- Chung, S. T. L., & Legge, G. E. (2016). Comparing the Shape of Contrast Sensitivity Functions for Normal and Low Vision. *Investigative Ophthalmology & Visual Science*, 57(1), 198–207. <https://doi.org/10.1167/iovs.15-18084>
- Chung, S. T. L., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, 38(19), 2949–2962. [https://doi.org/10.1016/S0042-6989\(98\)00072-8](https://doi.org/10.1016/S0042-6989(98)00072-8)
- Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high pass and low pass spatial filtering on face identification. *Perception and Psychophysics*, 58(4), 602–612.
- Cotton, S. D. O. (1996). Do all human skin colours lie on a defined surface within LMS space ? *School of Computer Science*.
- Crossland, M. D., Culham, L. E., & Rubin, G. S. (2004). Fixation stability and reading speed in patients with newly developed macular disease. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 24(4), 327–333. <https://doi.org/10.1111/j.1475-1313.2004.00213.x>
- Crossland, M. D., Starke, S. D., Imielski, P., Wolffsohn, J. S., & Webster, A. R. (2019). Benefit of an electronic head-mounted low vision aid. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 39(6), 422–431.
<https://doi.org/10.1111/opo.12646>
- Culham, L. E., Chabra, A., & Rubin, G. S. (2004). Clinical performance of electronic, head-mounted, low-vision devices. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 24(4), 281–290.
<https://doi.org/10.1111/j.1475-1313.2004.00193.x>
- Curcio, C. A., Millican, C. L., Allen, K. A., & Kalina, R. E. (1993). Aging of the human photoreceptor mosaic: Evidence for selective vulnerability of rods in central retina. *Investigative Ophthalmology and Visual Science*, 34(12), 3278–3296.
- Curcio, Christine A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4), 497–523.
<https://doi.org/10.1002/cne.902920402>
- da Silva, P. B. E., Leal, A. S., & Ferraz, N. N. (2020). Usability of smartphone apps as reading aids for low vision patients. *Disability and Rehabilitation: Assistive Technology*, 0(0), 1–5. <https://doi.org/10.1080/17483107.2020.1820086>
- Das, A., Pukhrambam, M., & Saha, A. (2016). Real-time robust face detection and tracking using extended haar functions and improved boosting algorithm. *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 981–985. <https://doi.org/10.1109/ICGCIoT.2015.7380605>
- Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63(2), 180–187.
<https://doi.org/10.1037/0021-9010.63.2.180>

- Deemer, A. D., Swenor, B. K., Fujiwara, K., Deremeik, J. T., Ross, N. C., Natale, D. M., ... Massof, R. W. (2019). Preliminary evaluation of two digital image processing strategies for head-mounted magnification for low vision patients. *Translational Vision Science and Technology*, 8(1). <https://doi.org/10.1167/tvst.8.1.23>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies-a review of behavioral findings and methodology. *Frontiers in Psychology*, 9(AUG), 1–7. <https://doi.org/10.3389/fpsyg.2018.01355>
- Duchaine, Brad, & Nakayama, K. (2006). The Cambridge Face Memory Test : Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants, 44, 576–585. <https://doi.org/10.1016/j.neuropsychologia.2005.07.001>
- Duchaine, Bradley, Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419–430. <https://doi.org/10.1080/02643290701380491>
- Efron, R. (1967). THE DURATION OF THE PRESENT. *Annals of the New York Academy of Sciences*, 138(2 Interdiscipli), 713–729. <https://doi.org/10.1111/j.1749-6632.1967.tb55017.x>
- Elgendy, M. (2020). *Deep Learning for Vision Systems*. Simon and Schuster.
- Elliott, D. B. (1987). Contrast Sensitivity Decline With Ageing: a Neural or Optical Phenomenon? *Ophthalmic and Physiological Optics*, 7(4), 415–419. <https://doi.org/10.1111/j.1475-1313.1987.tb00771.x>
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8(4), 431–439. <https://doi.org/10.1068/p080431>
- Enger, C., Alexander, M. F., & Fine, S. L. (1988). Contrast Sensitivity in Age-Related Macular Degeneration. *Archives of Ophthalmology*, 106(1), 55–57. <https://doi.org/10.1001/archophth.1988.01060130061028>
- Ezaki, N., Bulacu, M., & Schomaker, L. (2004). Text detection from natural scene images: Towards a system for visually impaired persons. *Proceedings - International Conference on Pattern Recognition*, 2, 683–686. <https://doi.org/10.1109/ICPR.2004.1334351>
- Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology*, 8(1), 1–19. <https://doi.org/10.1080/02643299108253364>
- Farfadi, S. S., Saberian, M., & Li, L. J. (2015). Multi-view face detection using Deep convolutional neural networks. *ICMR 2015 - Proceedings of the 2015 ACM International Conference on Multimedia Retrieval*, 643–650. <https://doi.org/10.1145/2671188.2749408>
- Fei, L., Chen, H., Wang, K., Lin, S., Yang, K., & Cheng, R. (2018). Scene text detection and recognition system for visually impaired people in real world, (October 2018), 29. <https://doi.org/10.1117/12.2325523>
- Field, G. D., & Chichilnisky, E. J. (2007). Information processing in the primate retina: Circuitry and coding. *Annual Review of Neuroscience*, 30, 1–30. <https://doi.org/10.1146/annurev.neuro.30.051606.094252>
- Fine, E. M., & Peli, E. (1995). Enhancement of text for the visually impaired. *Optical Society*

- of America*, 12(7), 9. <https://doi.org/10.1364/JOSAA.12.001439>
- Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., ... Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, 5(8), 625–632. [https://doi.org/10.1016/S2215-0366\(18\)30226-8](https://doi.org/10.1016/S2215-0366(18)30226-8)
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Garcia, C., & Tziritas, G. (1999). Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, 1(3), 264–277. <https://doi.org/10.1109/6046.784465>
- Genead, M. A., Fishman, G. A., Stone, E. M., & Allikmets, R. (2009). The natural history of stargardt disease with specific sequence mutation in the ABCA4 gene. *Investigative Ophthalmology and Visual Science*, 50(12), 5867–5871. <https://doi.org/10.1167/iovs.09-3611>
- Giacomelli, G., Virgili, G., Giansanti, F., Sato, G., Cappello, E., Cruciani, F., ... Menchini, U. (2013). Clinical and microperimetric predictors of reading speed in low vision patients: A structural equation modeling approach. *Investigative Ophthalmology and Visual Science*, 54(6), 4403–4408. <https://doi.org/10.1167/iovs.12-10734>
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Principles of perceptual learning and development. East Norwalk, CT, US: Appleton-Century-Crofts.
- Glen, F. C., Crabb, D. P., Smith, N. D., Burton, R., & Garway-Heath, D. F. (2012). Do patients with glaucoma have difficulty recognizing faces? *Investigative Ophthalmology and Visual Science*, 53(7), 3629–3637. <https://doi.org/10.1167/iovs.11-8538>
- Golubova, E., Starke, S. D., Crossland, M. D., & Wolffsohn, J. S. (2021). Design considerations for the ideal low vision aid: insights from de-brief interviews following a real-world recording study. *Ophthalmic and Physiological Optics*. <https://doi.org/10.1111/opo.12778>
- Gooch, B., Reinhard, E., & Gooch, A. (2004). Human facial illustrations: Creation and psychophysical evaluation. *ACM Trans. Graph.*, 23(1), 27–44. <https://doi.org/10.1145/966131.966133>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Gopalakrishnan, S., Chouhan Suwalal, S., Bhaskaran, G., & Raman, R. (2020). Use of augmented reality technology for improving visual acuity of individuals with low vision. *Indian Journal of Ophthalmology*, 68(6), 1136–1142. https://doi.org/10.4103/ijo.IJO_1524_19
- Greenspan, H., Goldberger, J., & Eshet, I. (2001). Mixture model for face-color modeling and segmentation. *Pattern Recognition Letters*, 22(14), 1525–1536. [https://doi.org/10.1016/S0167-8655\(01\)00086-1](https://doi.org/10.1016/S0167-8655(01)00086-1)
- Grover, S., Fishman, G. A., & Brown, J. (1998). Patterns of visual field progression in patients with retinitis pigmentosa. *Ophthalmology*, 105(6), 1069–1075. [https://doi.org/10.1016/S0161-6420\(98\)96009-2](https://doi.org/10.1016/S0161-6420(98)96009-2)
- Gui, L., Lin, Y., Wang, S., & Ding, X. (2011). Accurate face segmentation based on fusion of

- improved GrabCut and ASM. *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011*, 3(October 2011), 1175–1179.
<https://doi.org/10.1109/CISP.2011.6100423>
- Gupta, A., Mesik, J., Engel, S. A., Smith, R., Schatzka, M., Calabrèse, A., ... Legge, G. E. (2018). Beneficial effects of spatial remapping for reading with simulated central field loss. *Investigative Ophthalmology and Visual Science*, 59(2), 1105–1112.
<https://doi.org/10.1167/iov.16-21404>
- Haibo He, & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
<https://doi.org/10.1109/TKDE.2008.239>
- Hajraoui, A., & Sabri, M. (2014). Face Detection Algorithm based on Skin Detection, Watershed Method and Gabor Filters. *International Journal of Computer Applications*, 94(6), 33–39. <https://doi.org/10.5120/16349-5695>
- Han, J., Award, G. M., Sutherland, A., & Wu, H. (2006). Automatic skin segmentation for gesture recognition combining region and support vector machine active learning. *FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006*, 237–242. <https://doi.org/10.1109/FGR.2006.27>
- Hart, P. M., Chakravarthy, U., Stevenson, M. R., & Jamison, J. Q. (1999). A vision specific functional index for use in patients with age related macular degeneration, 1115–1120.
- Hasinoff, S. W., Sharlet, D., Geiss, R., Adams, A., Barron, J. T., Kainz, F., ... Levoy, M. (2016). Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics*, 35(6), 1–12. <https://doi.org/10.1145/2980179.2980254>
- Hassell, J. B., Lamoureux, E. L., & Keeffe, J. E. (2006). Impact of age related macular degeneration on quality of life. *British Journal of Ophthalmology*, 90(5), 593–596.
<https://doi.org/10.1136/bjo.2005.086595>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009a). Linear Methods for Classification. In *The Elements of Statistical Learning* (pp. 101–137). https://doi.org/10.1007/978-0-387-84858-7_4
- Hastie, T., Tibshirani, R., & Friedman, J. (2009b). Random Forests - The Elements of Statistical Learning: Data Mining, Inference, and Prediction. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.) (pp. 587–604). New York, NY: Springer New York.
https://doi.org/10.1007/978-0-387-84858-7_15
- Hauk, N., Hüffmeier, J., & Krumm, S. (2018). Ready to be a Silver Surfer? A Meta-analysis on the Relationship Between Chronological Age and Technology Acceptance. *Computers in Human Behavior*, 84, 304–319. <https://doi.org/10.1016/j.chb.2018.01.020>
- Hoffman, D. M., Girshick, A. R., & Banks, M. S. (2015). Vergence – accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(2008), 1–30.
<https://doi.org/10.1167/8.3.33.Introduction>
- Hollander, D. a, Volpe, N. J., Moster, M. L., Liu, G. T., Balcer, L. J., Judy, K. D., & Galetta, S. L. (2000). Use of a portable head mounted perimetry system to assess bedside visual fields. *The British Journal of Ophthalmology*, 84(10), 1185–1190.
<https://doi.org/10.1136/bjo.84.10.1185>
- Howe, J. (2012). Eccentric viewing training and its effect on the reading rates of individuals with absolute central scotomas: A meta-analysis. *Journal of Visual Impairment and Blindness*, 106(9), 527–542. <https://doi.org/10.1177/0145482x1210600904>
- Huang, K. S., & Trivedi, M. M. (2004). Robust real-time detection, tracking, and pose estimation of faces in video streams. *Proceedings - International Conference on Pattern*

- Recognition*, 3, 965–968. <https://doi.org/10.1109/icpr.2004.1334689>
- Hwang, A., & Peli, E. (2014). An Augmented-Reality Edge Enhancement Application for Google Glass. *Optometry & Vision Science*, 91(8), 1021–1030. Retrieved from http://journals.lww.com/optvissci/Abstract/2014/08000/An_Augmented_Reality_Edge_Enhancement_Application.28.aspx
- Irons, J., Barnes, N., Provis, J., & Kwa, A. (2014). A new theoretical approach to improving face recognition in disorders of central vision : Face caricaturing Elinor McKone, 14, 1–29. <https://doi.org/10.1167/14.2.12.doi>
- Itseez. (2015). Open Source Computer Vision Library.
- Jackson, A. S., Valstar, M., & Tzimiropoulos, G. (2016). A CNN cascade for landmark guided semantic part segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9915 LNCS, 143–155. https://doi.org/10.1007/978-3-319-49409-8_14
- Jacobs, J. M., Hammerman-Rozenberg, R., Maaravi, Y., Cohen, A., & Stessman, J. (2005). The impact of visual impairment on health, function and mortality. *Aging Clinical and Experimental Research*, 17(4), 281–286. <https://doi.org/10.1007/BF03324611>
- Jafri, R., & Arabnia, H. R. (2009). A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, 5(2), 41–68. <https://doi.org/10.3745/JIPS.2009.5.2.041>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013a). Classification. In *An Introduction to Statistical Learning* (pp. 127–173). https://doi.org/10.1007/978-1-4614-7138-7_4
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013b). Linear Model Selection and Regularization. In *An Introduction to Statistical Learning: with Applications in R* (pp. 203–264). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7_6
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013c). Linear Regression - An Introduction to Statistical Learning: with Applications in R. In G. James, D. Witten, T. Hastie, & R. Tibshirani (Eds.) (pp. 59–126). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7_3
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013d). Resampling Methods. In *An Introduction to Statistical Learning. Springer Texts in Statistics* (pp. 175–201). https://doi.org/10.1007/978-1-4614-7138-7_5
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013e). Tree-Based Methods. In *An Introduction to Statistical Learning*. (pp. 303–335). https://doi.org/10.1007/978-1-4614-7138-7_8
- Jiang, Z., Yao, M., & Jiang, W. (2007). Skin detection using color, texture and space information. *Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, 3(Fskd), 366–370. <https://doi.org/10.1109/FSKD.2007.518>
- Joan, S. P. F., & Valli, S. (2017). An enhanced text detection technique for the visually impaired to read text. *Information Systems Frontiers*, 19(5), 1039–1056. <https://doi.org/10.1007/s10796-016-9699-x>
- Johnson, A. P., Woods-Fry, H., & Wittich, W. (2017). Effects of Magnification on Emotion Perception in Patients With Age-Related Macular Degeneration. *Investigative Ophthalmology & Visual Science*, 58(5), 2520. <https://doi.org/10.1167/iovs.16-21349>
- Jolly, J. K., Juenemann, K., Boagey, H., Nadsady, M., Bridge, H., & MacLaren, R. E. (2020). Validation of electronic visual acuity (EVA) measurement against standardised ETDRS

- charts in patients with visual field loss from inherited retinal degenerations. *British Journal of Ophthalmology*, 104(7), 924–931. <https://doi.org/10.1136/bjophthalmol-2019-315124>
- Jones, P. R., & Ometto, G. (2018). Degraded reality: Using VR/AR to simulate visual impairments. *2018 IEEE Workshop on Augmented and Virtual Realities for Good, VAR4Good 2018*, (March), 2–5. <https://doi.org/10.1109/VAR4GOOD.2018.8576885>
- Jones, P. R., Tigchelaar, I., Demaria, G., Wilson, I., Bi, W., Taylor, D. J., & Crabb, D. P. (2020). Refinement and preliminary evaluation of two tablet-based tests of real-world visual function. *Ophthalmic and Physiological Optics*, 40(1), 35–46. <https://doi.org/10.1111/opo.12658>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, 17(11), 4302–4311. <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>
- Kaushik, D., & Verma, V. S. (2018). Review on Text Recognition in Natural Scene Images BT - Innovations in Computational Intelligence : Best Selected Papers of the Third International Conference on REDSET 2016. In B. Panda, S. Sharma, & U. Batra (Eds.) (pp. 29–43). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-4555-4_3
- Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1867–1874. <https://doi.org/10.1109/CVPR.2014.241>
- Khan, R., Hanbury, A., & Stoettinger, J. (2010). Skin detection: A random forest approach. In *2010 IEEE International Conference on Image Processing* (pp. 4613–4616). IEEE. <https://doi.org/10.1109/ICIP.2010.5651638>
- Khan, R., Hanbury, A., Stöttinger, J., & Bais, A. (2012). Color based skin classification. *Pattern Recognition Letters*, 33(2), 157–163. <https://doi.org/10.1016/j.patrec.2011.09.032>
- Khattab, D., Theobalt, C., Hussein, A. S., & Tolba, M. F. (2014). Modified GrabCut for human face segmentation. *Ain Shams Engineering Journal*, 5(4), 1083–1091. <https://doi.org/10.1016/j.asej.2014.04.012>
- Kim, Y., Hwang, I., & Cho, N. I. (2017). Convolutional neural networks and training strategies for skin detection. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3919–3923). IEEE. <https://doi.org/10.1109/ICIP.2017.8297017>
- Kinaterder, M., Gualtieri, J., Dunn, M. J., Jarosz, W., Yang, X. D., & Cooper, E. A. (2018). Using an Augmented Reality Device as a Distance-based Vision Aid-Promise and Limitations. *Optometry and Vision Science*, 95(9), 727–737. <https://doi.org/10.1097/OPX.0000000000001232>
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 1755–1758.
- King, D. E. (2015). Max-Margin Object Detection. Retrieved from <http://arxiv.org/abs/1502.00046>
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. Psychophysics: A practical introduction. San Diego, CA, US: Elsevier Academic Press.
- Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loève Procedure for the Characterization of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), 103–108. <https://doi.org/10.1109/34.41390>
- Kolmogorov, V., & Zabih, R. (2002). What energy functions can be minimized via graph cuts? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*

- Intelligence and Lecture Notes in Bioinformatics*), 2352(2), 65–81.
https://doi.org/10.1007/3-540-47977-5_5
- Lane, J., Robbins, R. A., Rohan, E. M. F., Crookes, K., Essex, R. W., Maddess, T., ... McKone, E. (2019). Caricaturing can improve facial expression recognition in low-resolution images and age-related macular degeneration. *Journal of Vision*, 19(6), 18.
<https://doi.org/10.1167/19.6.18>
- Lane, J., Rohan, E. M. F., Sabeti, F., Essex, R. W., Maddess, T., Dawel, A., ... McKone, E. (2018). *Impacts of impaired face perception on social interactions and quality of life in age-related macular degeneration: A qualitative study and new community resources*. *PLoS ONE* (Vol. 13). <https://doi.org/10.1371/journal.pone.0209218>
- Lazzeri, S. (2013). Analysis of Functional Dissociations between Best Corrected Visual Acuity and Microperimetric Parameters in Neovascular Age-Related Macular Degeneration Patients Underwent to Three Monthly Ranibizumab Injections. *Journal of Clinical & Experimental Ophthalmology*, 04(04). <https://doi.org/10.4172/2155-9570.1000293>
- Leat, S. J., & Mei, M. (2009). Custom-devised and generic digital enhancement of images for people with maculopathy. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 29(4), 397–415.
<https://doi.org/10.1111/j.1475-1313.2008.00633.x>
- Leat, S. J., Omoruyi, G., Kennedy, A., & Jernigan, E. (2005). Generic and customized digital image enhancement filters for the visually impaired. *Vision Research*, 45(15), 1991–2007. <https://doi.org/10.1016/j.visres.2005.01.028>
- Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision*, 11(5), 8–8.
<https://doi.org/10.1167/11.5.8>
- Legge, G E, Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading--I. Normal vision. *Vision Research*, 25(2), 239–252. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4013091>
- Legge, G E, & Rubin, G. S. (1986). R4-Psychophysics of reading. IV. Wavelength effects in normal and low vision. *Journal of the Optical Society of America. A, Optics and Image Science*, 3(1), 40–51. <https://doi.org/10.1364/JOSAA.3.000040>
- Legge, Gordon E., Cheung, S. H., Yu, D., Chung, S. T. L., Lee, H. W., & Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *Journal of Vision*, 7(2), 1–15. <https://doi.org/10.1167/7.2.9>
- Legge, Gordon E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, 41(6), 725–743. [https://doi.org/10.1016/S0042-6989\(00\)00295-9](https://doi.org/10.1016/S0042-6989(00)00295-9)
- Legge, Gordon E., & Rubin, G. S. (1986). Psychophysics of reading IV Wavelength effects in normal and low vision. *Journal of the Optical Society of America A*, 3(1), 40.
<https://doi.org/10.1364/josaa.3.000040>
- Legge, Gordon E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading-II. Low vision. *Vision Research*, 25(2), 253–265. [https://doi.org/10.1016/0042-6989\(85\)90118-X](https://doi.org/10.1016/0042-6989(85)90118-X)
- Li, H., & Ngan, K. N. (2008). Saliency model-based face segmentation and tracking in head-and-shoulder video sequences. *Journal of Visual Communication and Image Representation*, 19(5), 320–333. <https://doi.org/10.1016/j.jvcir.2008.04.001>
- Li, H., Ngan, K. N., & Liu, Q. (2009). FaceSeg: Automatic face segmentation for real-time video. *IEEE Transactions on Multimedia*, 11(1), 77–88.

- <https://doi.org/10.1109/TMM.2008.2008922>
- Li, J. Q., Welchowski, T., Schmid, M., Mauschwitz, M. M., Holz, F. G., & Finger, R. P. (2020). Prevalence and incidence of age-related macular degeneration in Europe: A systematic review and meta-analysis. *British Journal of Ophthalmology*, 104(8), 1077–1084. <https://doi.org/10.1136/bjophthalmol-2019-314422>
- Li, Z., Pundlik, S., & Luo, G. (2013). Stabilization of magnified videos on a mobile device for visually impaired. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 54–55. <https://doi.org/10.1109/CVPRW.2013.15>
- Liao, M., Shi, B., & Bai, X. (2018). TextBoxes++: A Single-Shot Oriented Scene Text Detector. *IEEE Transactions on Image Processing*, 27(8), 3676–3690. <https://doi.org/10.1109/TIP.2018.2825107>
- Liao, S., Jain, A. K., & Li, S. Z. (2016). A Fast and Accurate Unconstrained Face Detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 211–223. <https://doi.org/10.1109/TPAMI.2015.2448075>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
- Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, 8(10), 1386–1390. <https://doi.org/10.1038/nn1538>
- Logan, A. J., Wilkinson, F., Wilson, H. R., Gordon, G. E., & Loffler, G. (2016). The Caledonian face test: A new test of face discrimination. *Vision Research*, 119, 29–41. <https://doi.org/10.1016/j.visres.2015.11.003>
- Lott, L. a, Haegerstrom-Portnoy, G., Schneck, M. E., & Brabyn, J. a. (2005). Face recognition in the elderly. *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, 82(10), 874–881. <https://doi.org/10.1097/01.opx.0000180764.68737.91>
- Lucas, B. D., & Kanade, T. (1981). Iterative Image Registration Technique With an Application To Stereo Vision., 2, 674–679.
- Lumini, A., & Nanni, L. (2018). Fair comparison of skin detection approaches on publicly available datasets. *ArXiv Preprint ArXiv:1802.02531*, (February). Retrieved from <http://arxiv.org/abs/1802.02531>
- Luo, G., & Pundlik, S. (2014). Image stabilization in smart phone magnification apps helps read distant text. *Investigative Ophthalmology & Visual Science*, 55(13), 4149.
- MacKeben, M., Nair, U. K. W., Walker, L. L., & Fletcher, D. C. (2015). Random word recognition chart helps scotoma assessment in low vision. *Optometry and Vision Science*, 92(4), 421–428. <https://doi.org/10.1097/OPX.0000000000000548>
- Mangione, C. M., Lee, P. P., Gutierrez, P. R., Spritzer, K., Berry, S., Hays, R. D., & National Eye Institute Visual Function Questionnaire Field Test Investigators. (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. *Archives of Ophthalmology (Chicago, Ill. : 1960)*, 119(7), 1050–1058. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11448327>
- Marmor, M. F. (1986). Contrast sensitivity versus visual acuity in retinal disease. *British Journal of Ophthalmology*, 70(7), 553–559. <https://doi.org/10.1136/bjo.70.7.553>
- Martiniello, N., Eisenbarth, W., Lehane, C., Johnson, A., & Wittich, W. (2019). Exploring the use of smartphones and tablets among people with visual impairments: Are

- mainstream devices replacing the use of traditional visual aids? *Assistive Technology*, 00(00), 1–12. <https://doi.org/10.1080/10400435.2019.1682084>
- Massof, R. W., Ahmadian, L., Grover, L. L., Deremeik, J. T., Goldstein, J. E., Rainey, C., ... Barnett, G. . (2007). The Activity Inventory: An Adaptive Visual Function Questionnaire. *Optometry and Vision Science*, 84(8), 763–774. <https://doi.org/10.1097/OPX.0b013e3181339efd>
- Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Mathur, R., Bhaskaran, K., Edwards, E., Lee, H., Chaturvedi, N., Smeeth, L., & Douglas, I. (2017). Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: A cohort study in the Clinical Practice Research Datalink 2004–2014. *BMJ Open*, 7(2), 1–11. <https://doi.org/10.1136/bmjopen-2016-014444>
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). Configural processing faces. *Trends in Cognitive Sciences*, 6(6), 255–260. [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4)
- McCulloch, D. L., Loffler, G., Colquhoun, K., Bruce, N., Dutton, G. N., & Bach, M. (2011). The effects of visual degradation on face discrimination. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 31(3), 240–248. <https://doi.org/10.1111/j.1475-1313.2011.00828.x>
- Mei, M., & Leat, S. J. (2009). Quantitative assessment of perceived visibility enhancement with image processing for single face images: A preliminary study. *Investigative Ophthalmology and Visual Science*, 50(9), 4502–4508. <https://doi.org/10.1167/iovs.08-3079>
- Mitchell, J., & Bradley, C. (2006). Quality of life in age-related macular degeneration: A review of the literature. *Health and Quality of Life Outcomes*, 4, 1–20. <https://doi.org/10.1186/1477-7525-4-97>
- Moret-Tatay, C., & Perea, M. (2011). Do serifs provide an advantage in the recognition of written words? *Journal of Cognitive Psychology*, 23(5), 619–624. <https://doi.org/10.1080/20445911.2011.546781>
- Murro, V., Sodi, A., Giacomelli, G., Mucciolo, D. P., Pennino, M., Virgili, G., & Rizzo, S. (2017). Reading ability and quality of life in stargardt disease. *European Journal of Ophthalmology*, 27(6), 740–745. <https://doi.org/10.5301/ejo.5000972>
- Naji, S., Jalab, H. A., & Kareem, S. A. (2019). A survey on skin detection in colored images. *Artificial Intelligence Review*, 52(2), 1041–1087. <https://doi.org/10.1007/s10462-018-9664-9>
- Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39(23), 3824–3833. [https://doi.org/10.1016/S0042-6989\(99\)00096-6](https://doi.org/10.1016/S0042-6989(99)00096-6)
- Nirkin, Y., Masi, I., Tuân, A. T., Hassner, T., & Medioni, G. (2018). On face segmentation, face swapping, and face perception. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 98–105. <https://doi.org/10.1109/FG.2018.00024>
- Norton, D., McBain, R., & Chen, Y. (2009). Reduced ability to detect facial configuration in middle-aged and elderly individuals: Associations with spatiotemporal visual processing. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 64(3), 328–334. <https://doi.org/10.1093/geronb/gbp008>
- O'Donnell, C. (2005). The greatest generation meets its greatest challenge: Vision loss and depression in older adults. *Journal of Visual Impairment and Blindness*, 99(4), 197–208.

- <https://doi.org/10.1177/0145482x0509900402>
- Ojanpää, H., & Näsänen, R. (2003). Utilisation of spatial frequency information in face search. *Vision Research*, 43(24), 2505–2515. [https://doi.org/10.1016/S0042-6989\(03\)00459-0](https://doi.org/10.1016/S0042-6989(03)00459-0)
- Osman, M., Maarof, M., & Rohani, M. (2016). Towards integrating statistical color features for human skin detection. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(2), 272–276. Retrieved from <https://zenodo.org/record/1111592>
- Owen, C. G., Jarrar, Z., Wormald, R., Cook, D. G., Fletcher, A. E., & Rudnicka, A. R. (2012). The estimated prevalence and incidence of late stage age related macular degeneration in the UK. *British Journal of Ophthalmology*, 96(5), 752–756. <https://doi.org/10.1136/bjophthalmol-2011-301109>
- Owsley, C., & Sekuler, R. (1981). Aging and low-contrast vision: face perception. *IOVS*, (August), 362–365.
- Paliy, I., Dovgan, V., Boumbarov, O., Panev, S., Sachenko, A., Kurylyak, Y., & Zagorodnya, D. (2011). Fast and robust face detection and tracking framework. *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2011*, 1(September), 430–434. <https://doi.org/10.1109/IDAACS.2011.6072790>
- Palmer, S., Logan, D., Nabili, S., & Dutton, G. N. (2010). Effective rehabilitation of reading by training in the technique of eccentric viewing: Evaluation of a 4-year programme of service delivery. *British Journal of Ophthalmology*, 94(4), 494–497. <https://doi.org/10.1136/bjo.2008.152231>
- Panda-Jonas, S., Jonas, J. B., & Jakobczyk-Zmija, M. (1995). Retinal Photoreceptor Density Decreases with Age. *Ophthalmology*, 102(12), 1853–1859. [https://doi.org/10.1016/S0161-6420\(95\)30784-1](https://doi.org/10.1016/S0161-6420(95)30784-1)
- Park, K., Kim, Y., Nguyen, B. J., Chen, A., & Chao, D. L. (2020). Quality of life assessment of severely visually impaired individuals using aira assistive technology system. *Translational Vision Science and Technology*, 9(4), 1–8. <https://doi.org/10.1167/tvst.9.4.21>
- Pascolini, D., & Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5), 614–618. <https://doi.org/10.1136/bjophthalmol-2011-300539>
- Patel, P. J., Chen, F. K., Rubin, G. S., & Tufail, A. (2008a). Intersession repeatability of visual acuity scores in age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, 49(10), 4347–4352. <https://doi.org/10.1167/iovs.08-1935>
- Patel, P. J., Chen, F. K., Rubin, G. S., & Tufail, A. (2008b). Intersession repeatability of visual acuity scores in age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, 49(10), 4347–4352. <https://doi.org/10.1167/iovs.08-1935>
- Pazio, M., Niedźwiecki, M., Kowalik, R., & Lebień, J. (2007). Text detection system for the blind. *European Signal Processing Conference, (Eusipco)*, 272–276.
- Pearson, D. E., & Robinson, J. a. (1985). Visual communication at very low data rates. *Proceedings of the IEEE*, 73(4), 795–812. <https://doi.org/10.1109/PROC.1985.13202>
- Pearson, D., Hanna, E., & Martinez, K. (1990). Computer generated cartoons. Retrieved from <http://eprints.ecs.soton.ac.uk/18484/>
- Peli, E., & Peli, T. (1984). Image enhancement for the visually impaired. *Optical Engineering*, 32(8), 2337–2350. Retrieved from

- <http://opticalengineering.spiedigitallibrary.org/article.aspx?articleid=1244416>
- Peli, Eli, Lee, E., Trempe, C. L., & Buzney, S. (1994). Image enhancement for the visually impaired: the effects of enhancement on face recognition. *Journal of the Optical Society of America A*, 11(7), 1929. <https://doi.org/10.1364/JOSAA.11.001929>
- Peli, T., & Lim, J. (1982). Adaptive filtering for image enhancement. *Optical Engineering*, V(1). Retrieved from <http://opticalengineering.spiedigitallibrary.org/article.aspx?articleid=1222267>
- Pelli, D. G., & Robson, J. G. (1988). the Design of a New Letter Chart for Measuring Contrast Sensitivity. *Clinical Vision Science*, 2(3), 187–199. <https://doi.org/10.1016/j.parkreldis.2012.11.013>
- Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research*, 90, 10–14. <https://doi.org/10.1016/j.visres.2013.04.015>
- Peterson, R. C., Wolffsohn, J. S., Rubinstein, M., & Lowe, J. (2003). Benefits of electronic vision enhancement systems (EVES) for the visually impaired. *American Journal of Ophthalmology*, 136(6), 1129–1135. [https://doi.org/10.1016/S0002-9394\(03\)00567-1](https://doi.org/10.1016/S0002-9394(03)00567-1)
- Pezzullo, L., Streatfeild, J., Simkiss, P., & Shickle, D. (2018). The economic impact of sight loss and blindness in the UK adult population. *BMC Health Services Research*, 18(1), 1–13. <https://doi.org/10.1186/s12913-018-2836-0>
- Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, 115(24), 6171–6176. <https://doi.org/10.1073/pnas.1721355115>
- Phung, S. L., Chai, D., & Bouzerdoun, A. (2001). A universal and robust human skin color model using neural networks. *Proceedings of the International Joint Conference on Neural Networks*, 4, 2844–2849. <https://doi.org/10.1109/ijcnn.2001.938827>
- Phung, Son Lam, Bouzerdoun, A., & Chai, D. (2005). Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1), 148–154. <https://doi.org/10.1109/TPAMI.2005.17>
- R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <http://www.r-project.org>
- Radner, W. (2017). Reading charts in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 255(8), 1465–1482. <https://doi.org/10.1007/s00417-017-3659-0>
- Radner, W., Willinger, U., Obermayer, W., Mudrich, C., Velikay-Parel, M., & Eisenwort, B. (1998). A new reading chart for simultaneous determination of reading vision and reading speed. *Klinische Monatsblätter Fur Augenheilkunde*, 213(3), 174–181. <https://doi.org/10.1055/s-2008-1034969>
- Rahman, N., Georgiou, M., Khan, K. N., & Michaelides, M. (2020). Macular dystrophies: Clinical and imaging features, molecular genetics and therapeutic options. *British Journal of Ophthalmology*, 104(4), 451–460. <https://doi.org/10.1136/bjophthalmol-2019-315086>
- Riva, G., & Cattolica, U. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information , (January).
- Robinson, J. L., Braimah Avery, V., Chun, R., Pusateri, G., & Jay, W. M. (2017). Usage of Accessibility Options for the iPhone and iPad in a Visually Impaired Population.

- Seminars in Ophthalmology*, 32(2), 163–171.
<https://doi.org/10.3109/08820538.2015.1045151>
- Ross, J. E., Clarke, D. D., & Bron, A. J. (1985). Effect of age on contrast sensitivity function: Uniocular and binocular findings. *British Journal of Ophthalmology*, 69(1), 51–56.
<https://doi.org/10.1136/bjo.69.1.51>
- Rosser, D. A., Cousens, S. N., Murdoch, I. E., Fitzke, F. W., & Laidlaw, D. A. H. (2003). How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Investigative Ophthalmology and Visual Science*, 44(8), 3278–3281.
<https://doi.org/10.1167/iovs.02-1100>
- Rother, C., Kolmogorov, V., & Blake, A. (2004). GrabCut - Interactive foreground extraction using iterated graph cuts. *ACM SIGGRAPH 2004 Papers, SIGGRAPH 2004*, 309–314.
<https://doi.org/10.1145/1186562.1015720>
- Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research*, 32(5), 895–902. [https://doi.org/10.1016/0042-6989\(92\)90032-E](https://doi.org/10.1016/0042-6989(92)90032-E)
- Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, 32(4), 863–881.
<https://doi.org/10.1016/j.neubiorev.2008.01.001>
- Sabanayagam, C., Banu, R., Chee, M. L., Lee, R., Wang, Y. X., Tan, G., ... Wong, T. Y. (2019). Incidence and progression of diabetic retinopathy: a systematic review. *The Lancet Diabetes and Endocrinology*, 7(2), 140–149. [https://doi.org/10.1016/S2213-8587\(18\)30128-1](https://doi.org/10.1016/S2213-8587(18)30128-1)
- Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, 32(3), 285–293. <https://doi.org/10.1068/p5027>
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. <https://doi.org/10.1109/CVPR.2015.7298682>
- Seiple, W., Rosen, R. B., & Garcia, P. M. T. (2013). Abnormal Fixation in Individuals With Age-Related Macular Degeneration When Viewing an Image of a Face. *Optometry and Vision Science*, 90(1), 45–56. <https://doi.org/10.1097/OPX.0b013e3182794775>
- Sekuler, R., Hutman, L., & Owsley, C. (1980). Human aging and spatial vision. *Science*, 209(4462), 1255–1256. <https://doi.org/10.1126/science.7403884>
- Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75(2), 221–242.
<https://doi.org/10.1111/j.2044-8295.1984.tb01895.x>
- Shapiro, A. S. S., & Wilk, M. B. (1965). Biometrika Trust An Analysis of Variance Test for Normality (Complete Samples) Published by : Oxford University Press on behalf of Biometrika Trust Stable. *Biometrika*, 52(3), 591–611. Retrieved from <https://pdfs.semanticscholar.org/1f1d/9a7151d52c2e26d35690dbc7ae8098beee22.pdf>
- Simunovic, M. P. (2016a). Acquired color vision deficiency. *Survey of Ophthalmology*.
<https://doi.org/10.1016/j.survophthal.2015.11.004>
- Simunovic, M. P. (2016b). Acquired color vision deficiency. *Survey of Ophthalmology*, 61(2), 132–155. <https://doi.org/10.1016/j.survophthal.2015.11.004>
- Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3), 519.
<https://doi.org/10.1364/josaa.4.000519>
- Sivakumar, P., Vedachalam, R., Kannusamy, V., Odayappan, A., Venkatesh, R., Dhoble, P., ...

- Narayana, S. (2020). Barriers in utilisation of low vision assistive products. *Eye (Basingstoke)*, 34(2), 344–351. <https://doi.org/10.1038/s41433-019-0545-5>
- Smith, V. C., Pokorny, J., & Diddie, K. R. (1988). Color matching and the Stiles–Crawford effect in observers with early age-related macular changes. *Journal of the Optical Society of America A*, 5(12), 2113. <https://doi.org/10.1364/josaa.5.002113>
- Sobottka, K., & Pitas, I. (1998). A novel method for automatic face segmentation, facial feature extraction and tracking. *Signal Processing: Image Communication*, 12(3), 263–281. [https://doi.org/10.1016/S0923-5965\(97\)00042-8](https://doi.org/10.1016/S0923-5965(97)00042-8)
- Solomon, S. G. (2021). *Retinal ganglion cells and the magnocellular, parvocellular, and koniocellular subcortical visual pathways from the eye to the brain. Handbook of Clinical Neurology* (1st ed., Vol. 178). Elsevier B.V. <https://doi.org/10.1016/B978-0-12-821377-3.00018-0>
- Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, 8(4), 276–286. <https://doi.org/10.1038/nrn2094>
- SPRY, P. G. D. (2001). Visual Fields. *British Journal of Ophthalmology*, 85(2), 249d – 249. <https://doi.org/10.1136/bjo.85.2.249d>
- Stevens, G. A., White, R. A., Flaxman, S. R., Price, H., Jonas, J. B., Keeffe, J., ... Bourne, R. R. A. (2013). Global prevalence of vision impairment and blindness: Magnitude and temporal trends, 1990-2010. *Ophthalmology*, 120(12), 2377–2384. <https://doi.org/10.1016/j.ophtha.2013.05.025>
- Stifter, E., König, F., Lang, T., Bauer, P., Richter-Mülsch, S., Velikay-Parel, M., & Radner, W. (2004). Reliability of a standardized reading chart system: Variance component analysis, test-retest and inter-chart reliability. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 242(1), 31–39. <https://doi.org/10.1007/s00417-003-0776-8>
- Stroud, J. M. (1967). THE FINE STRUCTURE OF PSYCHOLOGICAL TIME. *Annals of the New York Academy of Sciences*, 138(2 Interdiscipli), 623–631. <https://doi.org/10.1111/j.1749-6632.1967.tb55012.x>
- Sutherland, I. (1965). Augmented Reality: “The Ultimate Display” by Ivan Sutherland, 1965. *Multimedia: From Wagner to Virtual Reality*, 4–5.
- Szabo, S. M., Janssen, P. A., Khan, K., Lord, S. R., & Potter, M. J. (2010). Neovascular AMD: An overlooked risk factor for injurious falls. *Osteoporosis International*, 21(5), 855–862. <https://doi.org/10.1007/s00198-009-1025-8>
- Tan, W. R., Chan, C. S., Yogarajah, P., & Condell, J. (2012). A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, 8(1), 138–147. <https://doi.org/10.1109/TII.2011.2172451>
- Tanaka, J. W., & Simonyi, D. (2016). The “Parts and Wholes” of Face Recognition: A Review of the Literature. *Quarterly Journal of Experimental Psychology*, 69(10), 1876–1889. <https://doi.org/10.1080/17470218.2016.1146780>
- Taylor, D. J., Edwards, L. A., Binns, A. M., & Crabb, D. P. (2018). Seeing it differently: self-reported description of vision loss in dry age-related macular degeneration. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 38(1), 98–105. <https://doi.org/10.1111/opo.12419>
- Taylor, D. J., Hobby, A. E., Binns, A. M., & Crabb, D. P. (2016). How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. *BMJ Open*, 6(12), 1–14. <https://doi.org/10.1136/bmjopen-2016-011504>
- Taylor, D. J., Smith, N. D., Binns, A. M., & Crabb, D. P. (2018). The effect of non-neovascular

- age-related macular degeneration on face recognition performance. *Graefe's Archive for Clinical and Experimental Ophthalmology = Albrecht von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie*, 815–821.
<https://doi.org/10.1007/s00417-017-3879-3>
- Taylor, D. J., Smith, N. D., & Crabb, D. P. (2017). Searching for objects in everyday scenes: Measuring performance in people with dry age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, 58(3), 1887–1892.
<https://doi.org/10.1167/iovs.16-21122>
- Tejeria, L., Harper, R. A., Artes, P. H., & Dickinson, C. M. (2002). Face recognition in age related macular degeneration: perceived disability, measured disability, and performance with a bioptic device, 1019–1026.
- Terrillon, J. C., Shirazi, M. N., Fukamachi, H., & Akamatsu, S. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000*, 54–61.
<https://doi.org/10.1109/AFGR.2000.840612>
- Thompson, W. B., Legge, G. E., Kersten, D. J., Shakespeare, R. A., & Lei, Q. (2017). Simulating visibility under reduced acuity and contrast sensitivity. *Journal of the Optical Society of America A*, 34(4), 583. <https://doi.org/10.1364/josaa.34.000583>
- Tiddeman, B. ., Stirrat, M. R., & Perrett, D. I. (2005). Towards Realism in Facial Image Transformation: Results of a Wavelet MRF Method. *Computer Graphics Forum*, 24(3), 449–456. <https://doi.org/10.1111/j.1467-8659.2005.00870.x>
- Tin Kam Ho. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278–282). IEEE Comput. Soc. Press.
<https://doi.org/10.1109/ICDAR.1995.598994>
- Tita-Nistor, L., González, E. G., Mkowitz, S. N., & Steinbach, M. J. (2009). Plasticity of fixation in patients with central vision loss. *Visual Neuroscience*, 26(5–6), 487–494.
<https://doi.org/10.1017/S0952523809990265>
- Tsapakis, S., Papaconstantinou, D., Diagourtas, A., Droutsas, K., Andreanos, K., Moschos, M. M., & Brouzas, D. (2017). Visual field examination method using virtual reality glasses compared with the humphrey perimeter. *Clinical Ophthalmology*, 11, 1431–1443.
<https://doi.org/10.2147/OPHTH.S131160>
- Van Gemert, M. J. C., Jacques, S. L., Sterenborg, H. J. C. M., & Star, W. M. (1989). Skin optics. *IEEE Transactions on Biomedical Engineering*, 36(12), 1146–1154.
<https://doi.org/10.1109/10.42108>
- van Rheede, Joram J., Wilson, I. R., Qian, R. I., Downes, S. M., Kennard, C., & Hicks, S. L. (2015). Improving Mobility Performance in Low Vision With a Distance-Based Representation of the Visual Scene. *Investigative Ophthalmology & Visual Science*, 56(8), 4802. <https://doi.org/10.1167/iovs.14-16311>
- Van Rheede, Joram Jacob, Wilson, I. R., Di Bon-Conyers, L., Croxford, S., MacLaren, R. E., & Hicks, S. L. (2016). Smart Specs: Electronic vision enhancement in real-life scenarios. *Investigative Ophthalmology & Visual Science*, 57(12), 5165.
- Vandenbroucke, T., Buyl, R., De Zaeytijd, J., Bauwens, M., Uvijls, A., De Baere, E., & Leroy, B. P. (2015). Colour vision in Stargardt disease. *Ophthalmic Research*, 54(4), 181–194.
<https://doi.org/10.1159/000438906>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn. *GetMobile: Mobile Computing and Communications*, 19(1), 29–33.

- <https://doi.org/10.1145/2786984.2786995>
- Vemala, R., Sivaprasad, S., & Barbur, J. L. (2017). Detection of Early Loss of Color Vision in Age-Related Macular Degeneration - With Emphasis on Drusen and Reticular Pseudodrusen. *Investigative Ophthalmology & Visual Science*, 58(6), BIO247–BIO254. <https://doi.org/10.1167/iovs.17-21771>
- Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137–154. Retrieved from <http://link.springer.com/article/10.1023/B:VISI.0000013087.49260.fb>
- Virgili, G., Acosta, R., Grover, L. L., Bentley, S. A., & Giacomelli, G. (2013). Reading aids for adults with low vision. *Cochrane Database of Systematic Reviews*, 2013(10). <https://doi.org/10.1002/14651858.CD003303.pub3>
- Wallace, J. M., Chung, S. T. L., & Tjan, B. S. (2017). Object crowding in age-related macular degeneration. *Journal of Vision*, 17(1), 1–13. <https://doi.org/10.1167/17.1.33>
- Walt, S. J. van der. (2010). Super-resolution Imaging. *Thesis*, (December 2010), 156.
- Watson, a B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*. <https://doi.org/10.3758/BF03202828>
- Watson, A. B. (2014). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, 14(7), 15. <https://doi.org/10.1167/14.7.15>
- Watson, A. B. (2016). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, 14(2014), 1–17. <https://doi.org/10.1167/14.7.15.doi>
- Westheimer, G. (1979). The spatial sense of the eye. *Investigative Ophthalmology and Visual Science*, 18(9), 893–912.
- Whittaker, S. G., Budd, J., & Cummings, R. W. (1988). Eccentric fixation with macular scotoma. *Investigative Ophthalmology and Visual Science*, 29(2), 268–278.
- Wilson, H. R., Loffler, G., & Wilkinson, F. (2002a). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42(27), 2909–2923. [https://doi.org/10.1016/S0042-6989\(02\)00362-0](https://doi.org/10.1016/S0042-6989(02)00362-0)
- Wilson, H. R., Loffler, G., & Wilkinson, F. (2002b). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42(27), 2909–2923. [https://doi.org/10.1016/S0042-6989\(02\)00362-0](https://doi.org/10.1016/S0042-6989(02)00362-0)
- Wilson, I. R., Van Rheede, J. J., Campbell, A. M., MacLaren, R. E., & Hicks, S. L. (2016). A mobile image enhancement system for sight impaired individuals. *Investigative Ophthalmology & Visual Science*, 57(12), 1956.
- Wittich, W., Jarry, J., Morrice, E., & Johnson, A. (2018). Effectiveness of the Apple iPad as a Spot-reading Magnifier. *Optometry and Vision Science*, 95(9), 704–710. <https://doi.org/10.1097/OPX.0000000000001269>
- Wittich, W., Lorenzini, M. C., Markowitz, S. N., Tolentino, M., Gartner, S. A., Goldstein, J. E., & Dagnelie, G. (2018). The Effect of a Head-mounted Low Vision Device on Visual Function. *Optometry and Vision Science*, 95(9), 774–784. <https://doi.org/10.1097/OPX.0000000000001262>
- Wong, W. L., Su, X., Li, X., Cheung, C. M. G., Klein, R., Cheng, C. Y., & Wong, T. Y. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis. *The Lancet Global Health*, 2(2), e106–e116. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1)
- Wroblewski, D., Francis, B. A., Sadun, A., Vakili, G., & Chopra, V. (2014). Testing of visual

- field with virtual reality goggles in manual and visual grasp modes. *BioMed Research International*, 2014. <https://doi.org/10.1155/2014/206082>
- Wu, S. M. (1991). Input-output relations of the feedback synapse between horizontal cells and cones in the tiger salamander retina. *Journal of Neurophysiology*, 65(5), 1197–1206. <https://doi.org/10.1152/jn.1991.65.5.1197>
- Wu, Y.-H., Granquist, C., Gage, R., Crossland, M. D., & Legge, G. E. (2020). Online Survey of Digital Reading by Adults with Low Vision. *Optometry and Vision Science*, 97(4). Retrieved from https://journals.lww.com/optvissci/Fulltext/2020/04000/Online_Survey_of_Digital_Reading_by_Adults_with.4.aspx
- Xiong, Y. Z., Lorsche, E. A., Mansfield, J. S., Bigelow, C., & Legge, G. E. (2018). Fonts designed for macular degeneration: Impact on reading. *Investigative Ophthalmology and Visual Science*, 59(10), 4182–4189. <https://doi.org/10.1167/iov.18-24334>
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14(6), 737–746. <https://doi.org/10.1068/p140737>
- Yu, D., Cheung, S. H., Legge, G. E., & Chung, S. T. L. (2010). Reading speed in the peripheral visual field of older adults: Does it benefit from perceptual learning? *Vision Research*, 50(9), 860–869. <https://doi.org/10.1016/j.visres.2010.02.006>
- Yu, D., Legge, G. E., Park, H., Gage, E., & Chung, S. T. L. (2010). Development of a training protocol to improve reading performance in peripheral vision. *Vision Research*, 50(1), 36–45. <https://doi.org/10.1016/j.visres.2009.10.005>
- Yusuf, A. A., Mohamad, F. S., & Sufyanu, Z. (2017). Human Face Detection Using Skin Color Segmentation and Watershed Algorithm. *Http://Www.Sciencepublishinggroup.Com*, 1(1), 29. <https://doi.org/10.11648/J.AJAI.20170101.14>
- Zafaruddin, G. M., & Fadewar, H. S. (2018). Face recognition using eigenfaces. *Advances in Intelligent Systems and Computing*, 810, 855–864. https://doi.org/10.1007/978-981-13-1513-8_87
- Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. *Proceedings of the IEEE International Conference on Computer Vision*, 1(c), 666–673.
- Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: An efficient and accurate scene text detector. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*, 2642–2651. <https://doi.org/10.1109/CVPR.2017.283>
- Zhu, Q., Wu, C. T., Cheng, K. T., & Wu, Y. L. (2004). An adaptive skin model and its application to objectionable image filtering. *ACM Multimedia 2004 - Proceedings of the 12th ACM International Conference on Multimedia*, 56–63. <https://doi.org/10.1145/1027527.1027538>
- Zhu, Q., & Wu, K. C. C. (2004). Adaptive Learning of an Accurate Skin-Color Model, 0–5.
- Zhu, Y., Yao, C., & Bai, X. (2016). Scene text detection and recognition: recent advances and future trends. *Frontiers of Computer Science*, 10(1), 19–36. <https://doi.org/10.1007/s11704-015-4488-0>
- Zuo, H., Fan, H., Blasch, E., & Ling, H. (2017). Combining Convolutional and Recurrent Neural Networks for Human Skin Detection. *IEEE Signal Processing Letters*, 24(3), 289–293. <https://doi.org/10.1109/LSP.2017.2654803>

Appendix – 1

A1.1 Visually impaired participants and experimental protocol

This section describes the recruitment process and overview of the experimental protocol. A detailed explanation of the protocol components will be contained within the respective chapters; this section is an outline to provide context.

A1.2 Recruitment

Participants were recruited from the Oxford Eye Hospital outpatient clinic after informed consent. Ethical approval was attained for the study, which was carried in accordance with the tenets of the Declaration of Helsinki. [NHS ethics reference: 17/SC/0178].

Participants were selected based on the following inclusion criteria; at least 18 years old, best-corrected visual acuity (BCVA) of 0.6 logMAR (6/24) or worse and a conversant English speaker.

The recruitment process involved identifying viable participants, to whom an information sheet was provided (see appendix) explaining the study. If participants could not read independently, the information sheet was read to them.

A1.3 Protocol

Upon acceptance, the participant was invited to attend two experimental sessions at the Oxford Eye Hospital, John Radcliffe Hospital.

1. Face Perception Session:

- a. Informed consent*
- b. Quality of life Questionnaires*
- c. Binocular Visual Acuity*
- d. Face Perception Tests**
- e. Smart Glasses introduction**
- f. Smart Glasses Visual Acuity**
- g. Contrast Sensitivity test*
- h. Colour Sensitivity test*

2. Reading Session:

- a. Informed consent*
- b. Quality of life Questionnaires*
- c. Binocular Visual Acuity*
- d. Radner Reading Test**
- e. Virtual Reality Reading Test**
- f. Contrast Sensitivity test*
- g. Colour Sensitivity test*

The protocol was split into two sessions, which could happen on the same day. Breaks were encouraged, lunch a refreshment was provided. The difference between the sessions is highlighted in bold. Most participants performed both sessions, where they did not repeat the items in *italics*, only the items in **bold**.

A1.4 Patient Participants Ethics approval



Health Research Authority

Dr Stephen Hicks
University Research Lecturer
University of Oxford
Level 6, West Wing, John Radcliffe Hospital
Oxford
OX39DU

Email: hra.approval@nhs.net

08 May 2017

Dear Dr Hicks

Letter of HRA Approval

Study title:	Improving visual recognition and mobility with a wearable display
IRAS project ID:	198017
REC reference:	17/SC/0178
Sponsor	University of Oxford

I am pleased to confirm that HRA Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

IRAS project ID	198017
-----------------	--------

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *“After Ethical Review – guidance for sponsors and investigators”*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the [HRA website](http://www.hra.nhs.uk), and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the [HRA website](http://www.hra.nhs.uk).

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at <http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/>.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application

IRAS project ID	198017
-----------------	--------

procedure. If you wish to make your views known please use the feedback form available on the HRA website: <http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/>.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at <http://www.hra.nhs.uk/hra-training/>

Your IRAS project ID is **198017**. Please quote this on all correspondence.

Yours sincerely

Kevin Ahmed
Assessor

Telephone: 0207 104 8171
Email: hra.approval@nhs.net

Copy to: *Ms Heather House, Sponsor Contact, Oxford University Hospitals NHS
Foundation Trust*

IRAS project ID	198017
-----------------	--------

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	Version	Date
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)	1.0	15 July 2016
Instructions for use of medical device	1.1	20 April 2017
Interview schedules or topic guides for participants [Interview Schedule]	1.0	27 March 2017
IRAS Application Form [IRAS_Form_28032017]		28 March 2017
IRAS Application Form XML file [IRAS_Form_28032017]		28 March 2017
IRAS Checklist XML [Checklist_20042017]		20 April 2017
Letter from funder [EC funding letter]		10 April 2017
Letter from funder [EC proposal evaluation]		10 April 2017
Letter from sponsor	1.0	10 March 2017
Participant information sheet (PIS) [PIS - A&B]	1.1	20 April 2017
Participant information sheet (PIS) [PIS - C]	1.1	20 April 2017
Referee's report or other scientific critique report [NIHR grant application]		
Referee's report or other scientific critique report [NIHR grant outcome]		
Research protocol or project proposal	1.0	10 March 2017
Summary CV for Chief Investigator (CI) [Stephen Hicks CV]		
Validated questionnaire [MoCA-Blind]	7.1	
Validated questionnaire [VFQ-25]		01 January 2000
Validated questionnaire [DLTV]	1.0	03 January 2017
Other[Statement of Activities]	1.0	08 May 2017
Other [Schedule of Events]	1.0	08 May 2017
198017 17 SC 0178 Ltr- Favourable Opinion with Conditions-20.04.17		20 April 2017
198017 17 SC 0178 Ltr- Ack of Docs Received Following FO-24.04.17		24 April 2017

Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the, *participating NHS organisations, capacity and capability and Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: Ms Heather House
 Tel: 01865572245
 Email: ctrg@admin.ox.ac.uk

HRA assessment criteria

Section	HRA Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and consent process	Yes	No comments
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	The sponsor has submitted the HRA Statement of Activities and intends for this to form the agreement between the sponsor and study sites. The sponsor is not requesting, and does not require any additional contracts with study sites.
4.2	Insurance/indemnity	Yes	Where applicable, independent contractors (e.g. General Practitioners)

IRAS project ID	198017
-----------------	--------

Section	HRA Assessment Criteria	Compliant with Standards	Comments
	arrangements assessed		should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this research study
4.3	Financial arrangements assessed	Yes	External study funding has been secured from Horizon 2020 and is subject to a funding agreement between the Sponsor and participating NHS organisation. No study funding will be provided to sites, as detailed at Schedule 1 of the Statement of Activities.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	No comments
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

All participating NHS organisations will undertake the same study activities. There is therefore only one study site 'type' involved in the research.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local LCRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.

NHS organisations in England that are participating in the study **will be expected to formally confirm their capacity and capability** to host this research.

- Following issue of this letter, participating NHS organisations in England may now confirm to the sponsor their capacity and capability to host this research, when ready to do so. How capacity and capability will be confirmed is detailed in the Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) section of this appendix.

The Assessing, Arranging, and Confirming document on the HRA website provides further information for the sponsor and NHS organisations on assessing, arranging and confirming capacity and capability.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).

A Principal Investigator should be appointed at study sites.

GCP training is not a generic training expectation, in line with the [HRA statement on training expectations](#).

IRAS project ID	198017
-----------------	--------

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken

Where arrangements are not already in place, network staff (or similar) undertaking any of the research activities listed in A18 or A19 of the IRAS form (except for administration of questionnaires or surveys), would be expected to obtain an honorary research contract from one NHS organisation (if university employed), followed by Letters of Access for subsequent organisations. This would be on the basis of a Research Passport (if university employed) or an NHS to NHS confirmation of pre-engagement checks letter (if NHS employed). These should confirm enhanced DBS checks, including appropriate barred list checks, and occupational health clearance. For research team members only administering questionnaires or surveys, a Letter of Access based on standard DBS checks and occupational health clearance would be appropriate.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.

The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

A1.5 Patient Participant information sheet



Oxford University Hospitals 
NHS Foundation Trust

Chief Investigator: Dr Stephen Hicks
Senior Research Fellow

Level 6, West Wing, John Radcliffe Hospital,
Oxford OX3 9DU
Tel: +44(0)1865 234 635 Fax: +44(0)1865 231914
Email: stephen.hicks@ndcn.ox.ac.uk
Web: www.ndcn.ox.ac.uk

Participant Information Sheet

Improving visual recognition and mobility with a wearable display

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?

We have developed a light-weight wearable set of electronic glasses to enhance vision in people who are legally blind or partially sighted. The glasses look like a regular pair of

Information and consent – Group A&B
Improving visual recognition and mobility with a wearable display
Dr Stephen Hicks

v1.3 12th September 2017
IRAS Project Number: 198017
Page: 1 of 9

spectacles except that the lenses have been replaced by small, see-through displays. The spectacles are equipped with cameras and a small computer which captures the nearby environment and converts this into an enhanced visual image.

In this research project, we would like to compare a number of enhancement techniques that may improve the recognition of faces and objects, and improve people's ability to avoid obstacles. We will be looking at the speed and accuracy of people using our wearable display while they perform a number of short tasks such as pointing at and grasping objects, walking around obstacles, identifying faces and reading short passages of text. We might also ask you to operate the controls of the electronic glasses yourself, to make sure that the controls are easy to understand and use. This will help us to refine the software that controls the visual aid and improve it before allowing a separate set of people in the study to take the glasses home to try them out in everyday life. Finally, we would like to test how much sight loss you have using a letter chart and a visual field test, to see how different types of sight loss can be best improved with our glasses.

Why have I been invited?

You have been chosen because you have poor sight but are still able to see some brightness and maybe some colours. You attend a clinic at the John Radcliffe Hospital in Oxford and this is the way you were identified as being eligible to take part in the study.

Do I have to take part?

No, it is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not have any negative impact.

What will happen to me if I take part?

If you would like to take part, we will arrange a time for you to attend a study visit at the Oxford Eye Hospital, Level LG1 West Wing, John Radcliffe Hospital.

We will first ask you for your consent to be part of the study followed by a short interview including some questionnaires relating to your vision loss.

Then we will measure your vision using standard clinical tests to understand your level of sight loss. If you have performed any of these tests within the last year we will ask you for permission to access your records. This is to obtain relevant results so that we can avoid any unnecessary repetition of the tests.

Next you will be asked to wear a pair of electronic glasses. The glasses will boost the contrast and outlines of nearby things to help them stand out from the background. We will investigate your ability to see faces and objects using our visual enhancements on the glasses or on a computer display. We may ask you to walk around a number of objects including a small maze with and without the glasses. We will conclude with a debrief and the whole experiment.

should take about two hours, and you can have a break at any time you like.

What if I am not suitable for taking part?

There is a chance that you will be found not suitable for involvement in the study. Any information that is collected about you will be kept strictly confidential.

What are the benefits of taking part?

While there may be no direct benefit to you, the results of this study will inform the further development of the glasses.

Are there any possible disadvantages or risks to taking part?

The glasses have already been used extensively by a number visually impaired people in their homes, with no concern. There are no known disadvantages or risks in taking part in this study.

What if there is a problem?

The University of Oxford, as sponsor, has appropriate insurance in place in the unlikely event that you suffer any harm as a direct consequence of your participation in this study.

If you wish to complain about any aspect of the way in which you have been approached or treated during the course of this study, you should contact the head of this study, Stephen Hicks (01865 234635, email

stephen.hicks@ndcn.ox.ac.uk) or you may contact the University of Oxford Clinical Trials and Research Governance (CTRG) office on 01865 572224, or the head of CTRG, email ctr@admin.ox.ac.uk.

The Patient Advisory Liaison Service (PALS) is a confidential NHS service that can provide you with support for any complaints or queries you may have regarding the care you receive as an NHS patient. PALS is unable to provide information about this research study. If you wish to contact the PALS team please contact 01865 221473
Email: PALS@ouh.nhs.uk

Interviews may be recorded and anonymised quotes may be used in publications.

Will my taking part in this study be kept confidential?

All information that is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the University (e.g. publications) will have your name and address removed so that you cannot be recognised from it. The data will be kept for a maximum of five years and stored securely in the Nuffield Department of Clinical Neurosciences. Responsible members of the appropriate Regulatory bodies, the University of Oxford or the Oxford University Hospitals NHS Foundation Trust may be given access to identifiable data for monitoring and/or audit of the study to ensure we are complying with regulations.

We will ask you if we can contact you again for future, related studies. Your decision to be, or not to be contacted again will have no negative impact in any way, and you may

still participate in this research if you do not wish to be contacted again.

Will I be reimbursed for taking part?

We will reimburse you and your companion or carer, for any travel expenses. This will be paid directly into your bank account if you have one, otherwise cash will be arranged.

What will happen to the results of the research study?

The results of this study are likely to be published in scientific journals. You may request a copy of the published results if you wish by contacting Stephen Hicks (01865 234635, email Stephen.hicks@ndcn.ox.ac.uk). No personally identifiable information will be published.

Who is organising and funding this study?

This study is organised by the University of Oxford, sponsored by the Clinical Trials and Research Governance, University of Oxford and is funded from a private source, Mr Zhang.

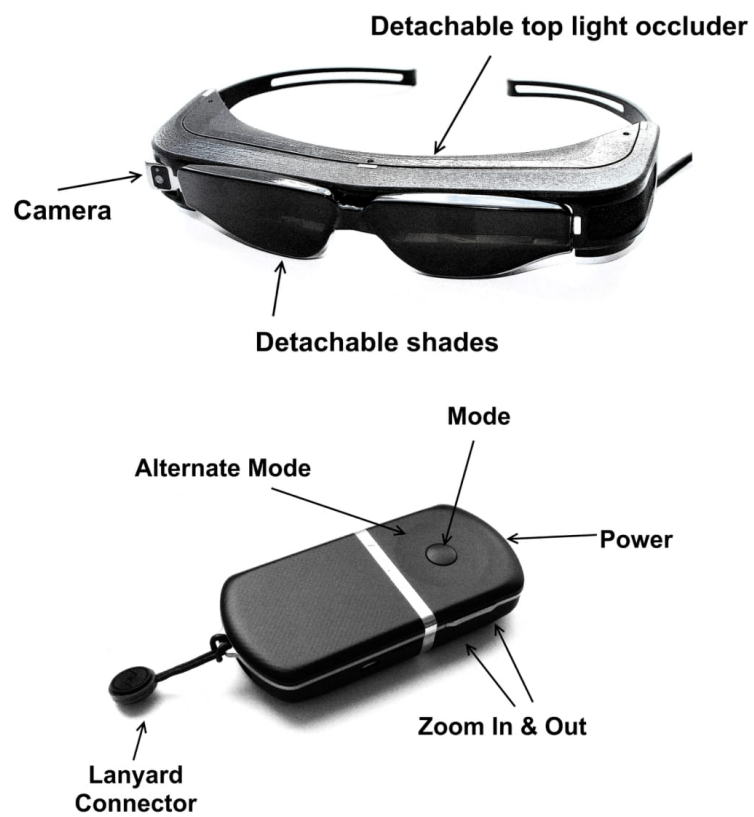
Who has reviewed this study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect participants' interests. This study has been reviewed and given favourable opinion by the Berkshire B Research Ethics Committee

Contact for further information

Dr Stephen Hicks
Tel 01865 234 635
Email stephen.hicks@ndcn.ox.ac.uk

The Electronic Glasses



A1.6 Patient Participant Consent Form



Oxford University Hospitals 
NHS Foundation Trust

Study Code:	Site ID Code:	Participant identification number:

CONSENT FORM

Improving visual recognition and mobility with a wearable display

Name of Researcher: Dr Stephen Hicks

If you agree, please initial each box

1. I confirm that I have read the information sheet dated 12/09/2017 (version 1.3) for this study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.	
3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from University of Oxford, from regulatory authorities [and from the NHS Trust(s)], where it is relevant	

to my taking part in this research. I give permission for these individuals to have access to my records.	
4. I agree to audio/video recording and the use of anonymised quotes in research reports and publications.	
5. I understand that the information collected about me may be used in an anonymous form to support other research in the future. It will not be possible for me to be identified by it.	
6. <i>OPTIONAL: I agree to be contacted about ethically approved research studies for which I may be suitable. I understand that agreeing to be contacted does not oblige me to participate in any further studies.</i>	
7. I agree to take part in this study.	

Name of Participant

Date

Signature

Name of Person taking Consent

Date

Signature

**1 copy for participant; 1 copy for researcher site file; 1 (original) to be kept in medical notes (if participant is a patient).*

A1.7 Patient Case Report Form

Case Report Form

IMAGE ENHANCEMENT FOR THE VISUALLY IMPAIRED (SMART GLASSES PROJECT)

Project registration number
17/SC/0178

Sponsor: University of Oxford

Chief Investigator: Professor Susan Downes

Date:

Subject code:

--	--	--	--	--

Clinical Trial site and PI

Oxford Eye Hospital (OEH)- Miss Susie Downes

CERTIFICATION ON CRF COMPLETION

I am confident that the information supplied in this case record form is complete and accurate data. I confirm that the study was conducted in accordance with the protocol and any protocol amendments and that written informed consent was obtained prior to the study.

Investigator's Signature: _____

--	--	--	--	--	--	--	--	--

Date of Signature: d d m m m y y y y

Visit 1:**SCREENING**

INCLUSION CRITERIA	YES	NO
1. Central Vision Loss (e.g AMD/Stargardt's)	☐	☐
2. Visual Acuity between 6/24 and Count Fingers (in at least one eye)	☐	☐
3. Age equal to or above 18 years	☐	☐
4. History of visual impairment longer than 2 years	☐	☐
5. Conversant English Speaker	☐	☐

EXCLUSION CRITERIA

1. Cognitive Impairment	☐	☐
2. Unable to perform baseline tests	☐	☐

Investigator's Initials

Date

Demographics:

1. Date of Birth

d	d	m	m	y	y	y	y

2. Years since diagnosis

3. Gender

MALE

FEMALE

Investigator's Initials

Date

--	--	--	--	--	--	--	--

YES NO

RECRUITMENT FORMS

- | | | |
|--|--------------------------|--------------------------|
| 1. Patient information sheet read and understood | ☐ | ☐ |
| 2. Research participant consent form signed | ☐ | ☐ |

Investigator's Initials

--

Date

--	--	--	--	--	--	--	--

Medical History/ Notes

--

Investigator's Initials

--

Date

--	--	--	--	--	--	--	--

FACE DISCRIMINATION EXPERIMENT

- | | |
|-------------------------------|--------------------------|
| 1. Face Perception (Baseline) | ☐ |
| a. Retest | ☐ |

- 2. Face Perception (Modified)
 - a. Peli
 - b. Cartoon

☐
☐

Investigator's Initials									
Date	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>								

QUESTIONNAIRES

- 1. DLTV
- 2. VFQ- 25

☐

Investigator's Initials									
Date	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>								

VISION CHARACTERISATION

- 1. Binocular Visual Acuity (ETDRS)

--
- 2. Cambridge Contrast Sensitivity
- 3. Cambridge Colour Sensitivity

☐
☐

Investigator's Initials									
Date	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>								

VISIT 2:

Date:

TEXT READING EXPERIMENT

1. Radner Text Reading

☐
2. Binocular Visual Acuity (ETDRS) with Smart Glasses
3. Smart Glasses Text Reading

• Not applicable ☐

☐
4. Virtual Reality Text Reading

☐

Investigator's Initials

Date

--	--	--	--	--	--	--	--

IMAGING

1. Microperimetry

☐
2. Fundus Autofluorescence Imaging

☐
3. Fundus photography

☐
4. Spectral Domain OCT Imaging

☐

Investigator's Initials

Date

--	--	--	--	--	--	--	--

General Notes Visit 2

--

Investigator's Initials	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td></tr></table>								
Date	<table border="1"><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>								

A1.8 Normally sighted participant ethics approval

MEDICAL SCIENCES INTER-DIVISIONAL RESEARCH ETHICS COMMITTEE

Research Services, University of Oxford, Wellington Square, Oxford, OX1 2J
Tel: +44(0)1865 616577 Fax: +44(0)1865 280467
ethics@medsci.ox.ac.uk



CONFIDENTIAL

Ref: R50271/RE001

Mr Iain Wilson
Nuffield Laboratory of Ophthalmology
University of Oxford
Level 6 West Wing
John Radcliffe Hospital
Oxford

7th March 2017

Dear Iain

Research Ethics Approval - CUREC 1

Study title: Face Discrimination enhancement: Simulated Visual Impairment

The above application has been considered on behalf of the Medical Sciences Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for Ethical Approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and approval has been granted for a period of **2 years**, commencing on **7th March 2017**. The reference number for this study is **R50271/RE001**.

This is **subject to**:

- a) **it is your responsibility to comply with the requirements for administering any tests or questionnaires and, if in doubt, to contact the publisher of those tests or questionnaires.**
- b) **if new staff are engaged MS IDREC should be informed of their names, status and ethics training.**

I would like to remind you that your study may be selected for review by the MS IDREC during an annual audit.

Amendments

Should there be any subsequent changes to the study, which raise ethical issues not covered in the original application, you should submit details to the IDREC for consideration and approval. If there are significant alterations then the Committee will require a succinct letter explaining these. If your study received University Sponsorship through CTRG, then the amendment will first need to be submitted to CTRG for approval of continued Sponsorship.

Please do not hesitate to contact me if you have any queries.

Yours Sincerely

A handwritten signature in black ink, appearing to read 'H. Barnby-Porritt'.

Dr. Helen Barnby-Porritt
Research Ethics Manager, Medical Sciences

A1.9 Normally sighted participant information sheet

Nuffield Laboratory of Ophthalmology
Nuffield Department of Clinical Neurosciences.
Level 6, West Wing, John Radcliffe Hospital,
Oxford OX3 9DU



PI: Professor Robert Maclaren
PI Email: robert.maclaren@eye.ox.ac.uk
Iain Wilson MSc(Res)
01865 234635
iain.wilson@ndcn.ox.ac.uk

Face Discrimination Enhancement: Simulated Visual Impairment

PARTICIPANT INFORMATION SHEET

Ethics Approval Reference: R50271/RE001

1. *Background and aims of the study*

In the UK there are approximately 370,000 people registered blind or partially sighted. Central Vision Loss affects the vast majority of the visually impaired population. For most of these sufferers there is little a clinician can do to help. Central vision loss leads to difficulties in tasks such as reading, watching television, driving and recognising faces. There is considerable evidence that central vision loss leads to a reduction of quality of life.

This study aims to better understand impact of image enhancements on face discrimination for people who suffer from central vision loss. To do this we are developing a test to measure a person's ability to perceive faces and validating it with a simulation of central vision loss, using an eye-tracker and computer display.

This study is funded by the National Institute for Health Research.

2. *Why have I been invited to take part?*

You have been invited because we require visually healthy adults to aid in the development of a face discrimination test.

The inclusion criterion is that you are over 18 years old and do not have a visual impairment or have any face perception impairment (such as prosopagnosia aka. face blindness). Requiring glasses or contact lenses will not exclude you from participation.

3. *Do I have to take part?*

No. You can ask questions about the study before deciding whether or not to participate. If you do agree to participate, you may withdraw yourself and your data from the study at any time, without giving a reason and without penalty, by advising the researchers of this decision.

4. *What will happen in the study?*

[Delete as appropriate]

Study 1:

If you are happy to take part in the study, you will be asked to attend one visit lasting up to one hour. We will go through the information sheet with you and then ask you to sign a consent form.

During this visit, you will first be asked to perform a visual acuity test. This will be an automated test run on a computer, similar to what an optician uses.

Next, you will be asked to perform the face discrimination test. This will involve the presentation of four faces on a computer screen, where one is different from the other three. Your task will be to identify the odd one out. The task will repeat many times, with varying difficulty.

This study will take place either in Wadham College or in the Nuffield Laboratory of Ophthalmology, John Radcliffe Hospital.

Study 2:

If you are happy to take part in the study, you will be asked to attend one visit lasting up to one hour. We will go through the information sheet with you and then ask you to sign a consent form.

During this visit, you will first be asked to perform a visual acuity test. This will be an automated test run on a computer, similar to what an optician uses.

Next, you will be asked to perform the face discrimination test with simulated central vision loss. This will involve the presentation of four faces on a computer screen, where one is different from the other three. Your task will be to identify the odd one out. You will view the computer screen with an eye tracker, which will monitor where you are looking, and mask a small area on the computer screen to simulate central vision loss. The task will repeat many times, with varying difficulty.

This study will take place either in the Nuffield Laboratory of Ophthalmology (John Radcliffe Hospital) or in the Sherrington Building (Department of Physiology, Anatomy and Genetics Sherrington Road, Oxford, OX1 3PT)

Study 3:

If you are happy to take part in the study, you will be asked to attend five visits, on consecutive days at the same time of day. Each visit will last up to one hour. We will go through the information sheet with you and then ask you to sign a consent form.

On your first visit, you will first be asked to perform a visual acuity test. This will be an automated test run on a computer, similar to what an optician uses.

Next, you will be asked to perform the face discrimination test with simulated central vision loss. This will involve the presentation of four faces on a computer screen, where one is different from the other three. Your task will be to identify the odd one out. You will view the computer screen with an eye tracker, which will monitor where you are looking, and mask a small area on the computer screen to simulate central vision loss. The task will repeat many times, with varying difficulty.

On the next 4 visits, you will be asked to repeat face discrimination test with simulated central vision loss.

This study will take place either in the Nuffield Laboratory of Ophthalmology (John Radcliffe Hospital) or in the Sherrington Building (Department of Physiology, Anatomy and Genetics Sherrington Road, Oxford, OX1 3PT)

5. Are there any potential risks in taking part?

There are no known risks from taking part in this study.

6. Are there any benefits in taking part?

There will be no direct benefit to you from taking part in this research, however you will be contributing to the development of technologies to aid the visually impaired.

7. Expenses and payments

You will receive £10 per visit, upon completion of all required visits.

8. What happens to the data provided?

The **research data** will be stored confidentially in a secure location within the Nuffield Laboratory of Ophthalmology computer systems. Your responses will be anonymised and you will not be identifiable from any of the data stored electronically.

Personal / sensitive data will be stored confidentially in a locked filing cabinet within the department separate from the research data.

Only the primary researcher and principle investigator will have access to the research data.

All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research.

9. Will the research be published?

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published open access

10. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number: R50271/RE001).

11. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please speak to the relevant researcher Iain Wilson on 01865 234635 or their supervisor Professor Robert MacLaren on 01865 223380, who will do their best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how they intend to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

Chair, **Medical Sciences Inter-Divisional Research Ethics Committee**; Email: ethics@medsci.ox.ac.uk;
Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

12. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Iain Wilson
Nuffield Laboratory of Ophthalmology,
Level 6 West Wing, John Radcliffe Hospital

Oxford, OX3 9QH
Tel: 01865 234635
Email: ian.wilson@ndcn.ox.ac.uk

A1.10 Normally sighted participant consent form

Nuffield Laboratory of Ophthalmology
Nuffield Department of Clinical Neurosciences.
Level 6, West Wing, John Radcliffe Hospital,
Oxford OX3 9DU



PI: Professor Robert MacLaren
PI Email: robert.maclaren@eye.ox.ac.uk
Iain Wilson MSc(Res)
01865 234635
iain.wilson@ndcn.ox.ac.uk

PARTICIPANT CONSENT FORM

CUREC Approval Reference: R50271/RE001

Face Discrimination Enhancement: Simulated Visual Impairment

Purpose of Study:

This study aims to better understand impact of image enhancements on face discrimination for people who suffer from central vision loss.

*Please initial each
box*

- | | | |
|---|--|----------------------|
| 1 | I confirm that I have read and understand the information sheet version <u>1.0</u> dated January 2017 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. | <input type="text"/> |
| 2 | I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without any adverse consequences or academic penalty. | <input type="text"/> |
| 3 | I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this study. I give permission for these individuals to access my data. | <input type="text"/> |
| 4 | I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. | <input type="text"/> |
| 5 | I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project. | <input type="text"/> |
| 6 | I understand how this research will be written up and published. | <input type="text"/> |
| 7 | I understand how to raise a concern or make a complaint. | <input type="text"/> |
| 8 | I agree to take part in the above study. | <input type="text"/> |

Name of Participant

Date

Signature

Name of person taking consent

Date

Signature

A1.11 Normally sighted participant recruitment poster

Poster Advert 1.0 January 2017



CUREC Approval reference: R50271/RE001

Nuffield Laboratory of Ophthalmology
Nuffield Department of Clinical Neurosciences
Level 6 West Wing, John Radcliffe Hospital
OX3 9DU

DO YOU RECOGNIZE THIS PERSON???



Face Discrimination enhancement: Simulated Visual Impairment

We are looking for healthy volunteers, aged 18 and above, with normal vision (glasses/contacts are okay) for a study investigating face recognition ability in a simulated visual impairment condition. Sessions should last no more than 1 hour.

You will be compensated for your participation

If you are interested and would like more information, please contact Iain Wilson, Nuffield Laboratory of Ophthalmology, John Radcliffe Hospital, OX3 9DU. 01865 234635 or by email: iain.wilson@ndcn.ox.ac.uk

[illegible]

A1.12 DLTV Questionnaire

How much difficulty do you have...

		<i>No difficulty</i>	<i>A little difficulty</i>	<i>A lot of difficulty</i>	<i>Cannot see to do</i>
1	Distinguishing a person's features across the room	4	3	2	1
2	Noticing objects off to either side	4	3	2	1
3	Watching TV programs	4	3	2	1
4	Seeing steps and using them	4	3	2	1
5	Enjoying the scenery if out for a drive	4	3	2	1
6	Reading road signs/street names	4	3	2	1
7	Distinguishing a person's features across the street	4	3	2	1
8	Recognizing seasonal changes in the garden	4	3	2	1
9	Distinguishing a person's features at arm's length	4	3	2	1
10	Pouring yourself a drink	4	3	2	1
11	Cutting up food on your plate	4	3	2	1
12	Cutting your finger nails	4	3	2	1
13	Using kitchen appliances	4	3	2	1
14	Adjusting to darkness after being in the light	4	3	2	1
15	Adjusting to the light after being in darkness	4	3	2	1

How confident do you feel in your ability to walk around

		<i>Extremely</i>	<i>Somewhat</i>	<i>Barely</i>	<i>Not at all</i>
16	In your immediate neighbourhood	4	3	2	1
17	Outside your immediate neighbourhood	4	3	2	1

With your near glasses on how much difficulty do you have

		<i>No difficulty</i>	<i>A little difficulty</i>	<i>A lot of difficulty</i>	<i>Cannot see to do</i>
18	Reading normal sized newspaper print	4	3	2	1
19	Reading newspaper headlines	4	3	2	1
20	Reading correspondence (e.g., bills, letters, cards)	4	3	2	1
21	Signing documents (checks, etc.)	4	3	2	1
22	Identifying money from purse or wallet	4	3	2	1

How would you rate

		<i>Excellent</i>	<i>Good</i>	<i>Fair</i>	<i>Poor</i>
23	Your overall distance vision	4	3	2	1
24	Your overall near vision (i.e., for close work)	4	3	2	1

A1.13 VFQ-25 Questionnaire

PB/IA

National Eye Institute Visual Functioning Questionnaire - 25 (VFQ-25)

version 2000

(INTERVIEWER ADMINISTERED FORMAT)

January 2000

RAND hereby grants permission to use the "National Eye Institute Visual Functioning Questionnaire 25 (VFQ-25) July 1996, in accordance with the following conditions which shall be assumed by all to have been agreed to as a consequence of accepting and using this document:

1. Changes to the NEI VFQ-25 - July 1996 may be made without the written permission of RAND. However, all such changes shall be clearly identified as having been made by the recipient.
2. The user of this NEI VFQ-25 - July 1996 accepts full responsibility, and agrees to hold RAND harmless, for the accuracy of any translations of the NEI VFQ-25 Test Version - July 1996 into another language and for any errors, omissions, misinterpretations, or consequences thereof.
3. The user of this NEI VFQ-25 - July 1996 accepts full responsibility, and agrees to hold RAND harmless, for any consequences resulting from the use of the NEI VFQ-25.
4. The user of the NEI VFQ-25 - July 1996 will provide a credit line when printing and distributing this document or in publications of results or analyses based on this instrument acknowledging that it was developed at RAND under the sponsorship of the National Eye Institute.
5. No further written permission is needed for use of this NEI VFQ-25 - July 1996.

7/29/96

© RAND 1996

Instructions:

I'm going to read you some statements about problems which involve your vision or feelings that you have about your vision condition. After each question I will read you a list of possible answers. Please choose the response that best describes your situation.

Please answer all the questions as if you were wearing your glasses or contact lenses (if any).

Please take as much time as you need to answer each question. All your answers are confidential. In order for this survey to improve our knowledge about vision problems and how they affect your quality of life, your answers must be as accurate as possible. Remember, if you wear glasses or contact lenses for a particular activity, please answer all of the following questions as though you were wearing them.

Visual Functioning Questionnaire - 25

PART 1 - GENERAL HEALTH AND VISION

1. In general, would you say your overall health is*:

(Circle One)

READ CATEGORIES:	Excellent.....	1
	Very Good	2
	Good	3
	Fair	4
	Poor.....	5

2. At the present time, would you say your eyesight using both eyes (with glasses or contact lenses, if you wear them) is excellent, good, fair, poor, or very poor or are you completely blind?

(Circle One)

READ CATEGORIES:	Excellent.....	1
	Good	2
	Fair	3
	Poor.....	4
	Very Poor.....	5
	Completely Blind.....	6

* Skip Question 1 when the VFQ-25 is administered at the same time as the SF-36 or RAND 36-Item Health Survey 1.0

3. How much of the time do you worry about your eyesight?

(Circle One)

READ CATEGORIES:	None of the time	1
	A little of the time	2
	Some of the time	3
	Most of the time.....	4
	All of the time?.....	5

4. How much pain or discomfort have you had in and around your eyes (for example, burning, itching, or aching)? Would you say it is:

(Circle One)

READ CATEGORIES:	None.....	1
	Mild.....	2
	Moderate	3
	Severe, or	4
	Very severe?.....	5

PART 2 - DIFFICULTY WITH ACTIVITIES

The next questions are about how much difficulty, if any, you have doing certain activities wearing your glasses or contact lenses if you use them for that activity.

5. How much difficulty do you have reading ordinary print in newspapers? Would you say you have:
(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all	1
A little difficulty.....	2
Moderate difficulty	3
Extreme difficulty	4
Stopped doing this because of your eyesight	5
Stopped doing this for other reasons or not interested in doing this	6

6. How much difficulty do you have doing work or hobbies that require you to see well up close, such as cooking, sewing, fixing things around the house, or using hand tools? Would you say:

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
 A little difficulty..... 2
 Moderate difficulty 3
 Extreme difficulty 4
 Stopped doing this because of your eyesight 5
 Stopped doing this for other reasons or not
 interested in doing this 6

7. Because of your eyesight, how much difficulty do you have finding something on a crowded shelf?

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
 A little difficulty..... 2
 Moderate difficulty 3
 Extreme difficulty 4
 Stopped doing this because of your eyesight 5
 Stopped doing this for other reasons or not
 interested in doing this 6

8. How much difficulty do you have reading street signs or the names of stores?

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
 A little difficulty..... 2
 Moderate difficulty 3
 Extreme difficulty 4
 Stopped doing this because of your eyesight 5
 Stopped doing this for other reasons or not
 interested in doing this 6

9. Because of your eyesight, how much difficulty do you have going down steps, stairs, or curbs in dim light or at night?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

10. Because of your eyesight, how much difficulty do you have noticing objects off to the side while you are walking along?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

11. Because of your eyesight, how much difficulty do you have seeing how people react to things you say?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

12. Because of your eyesight, how much difficulty do you have picking out and matching your own clothes?

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
A little difficulty..... 2
Moderate difficulty 3
Extreme difficulty 4
Stopped doing this because of your eyesight 5
Stopped doing this for other reasons or not
interested in doing this 6

13. Because of your eyesight, how much difficulty do you have visiting with people in their homes, at parties, or in restaurants ?

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
A little difficulty..... 2
Moderate difficulty 3
Extreme difficulty 4
Stopped doing this because of your eyesight 5
Stopped doing this for other reasons or not
interested in doing this 6

14. Because of your eyesight, how much difficulty do you have going out to see movies, plays, or sports events?

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all 1
A little difficulty..... 2
Moderate difficulty 3
Extreme difficulty 4
Stopped doing this because of your eyesight 5
Stopped doing this for other reasons or not
interested in doing this 6

15. Now, I'd like to ask about driving a car. Are you currently driving, at least once in a while?

(Circle One)

Yes..... 1 Skip To Q 15c

No 2

- 15a. IF NO, ASK: Have you never driven a car or have you given up driving?

(Circle One)

Never drove..... 1 Skip To Part 3, Q 17

Gave up..... 2

- 15b. IF GAVE UP DRIVING: Was that mainly because of your eyesight, mainly for some other reason, or because of both your eyesight and other reasons?

(Circle One)

Mainly eyesight..... 1 Skip To Part 3, Q 17

Mainly other reasons..... 2 Skip To Part 3, Q 17

Both eyesight and other reasons 3 Skip To Part 3, Q 17

- 15c. IF CURRENTLY DRIVING: How much difficulty do you have driving during the daytime in familiar places? Would you say you have:

(Circle One)

No difficulty at all 1

A little difficulty..... 2

Moderate difficulty 3

Extreme difficulty 4

16. How much difficulty do you have driving at night? Would you say you have: (READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
A little difficulty..... 2
Moderate difficulty 3
Extreme difficulty 4
Have you stopped doing this because
of your eyesight 5
Have you stopped doing this for other
reasons or are you not interested in
doing this 6

- 16a. How much difficulty do you have driving in difficult conditions, such as in bad weather, during rush hour, on the freeway, or in city traffic? Would you say you have:
(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
A little difficulty..... 2
Moderate difficulty 3
Extreme difficulty 4
Have you stopped doing this because
of your eyesight 5
Have you stopped doing this for other
reasons or are you not interested in
doing this 6

PART 3: RESPONSES TO VISION PROBLEMS

The next questions are about how things you do may be affected by your vision. For each one, I'd like you to tell me if this is true for you all, most, some, a little, or none of the time.

(Circle One On Each Line)

READ CATEGORIES:	All of the time	Most of the time	Some of the time	A little of the time	None of the time
17. <u>Do you accomplish less</u> than you would like because of your vision?	1	2	3	4	5
18. <u>Are you limited</u> in how long you can work or do other activities because of your vision?.....	1	2	3	4	5
19. How much does pain or discomfort in or around <u>your eyes</u> , for example, burning, itching, or aching, keep you from doing what you'd like to be doing? Would you say:	1	2	3	4	5

For each of the following statements, please tell me if it is definitely true, mostly true, mostly false, or definitely false for you or you are not sure.

(Circle One On Each Line)

	Definitely True	Mostly True	Not Sure	Mostly False	Definitely False
20. I <u>stay home most of the time</u> because of my eyesight.	1	2	3	4	5
21. I feel <u>frustrated</u> a lot of the time because of my eyesight.	1	2	3	4	5
22. I have <u>much less control</u> over what I do, because of my eyesight.	1	2	3	4	5
23. Because of my eyesight, I have to <u>rely too much on</u> <u>what other people tell me...</u>	1	2	3	4	5
24. I <u>need a lot of help</u> from others because of my eyesight.	1	2	3	4	5
25. I worry about <u>doing things</u> <u>that will embarrass myself</u> <u>or others</u> , because of my eyesight.	1	2	3	4	5

That's the end of the interview. Thank you very much for your time and your help.

Appendix of Optional Additional Questions

SUBSCALE: GENERAL HEALTH

- A1. How would you rate your overall health, on a scale where zero is as bad as death and 10 is best possible health?

(Circle One)

0	1	2	3	4	5	6	7	8	9	10
Worst					Best					

SUBSCALE: GENERAL VISION

- A2. How would you rate your eyesight now (with glasses or contact lens on, if you wear them), on a scale of from 0 to 10, where zero means the worst possible eyesight, as bad or worse than being blind, and 10 means the best possible eyesight?

(Circle One)

0	1	2	3	4	5	6	7	8	9	10
Worst					Best					

SUBSCALE: NEAR VISION

- A3. Wearing glasses, how much difficulty do you have reading the small print in a telephone book, on a medicine bottle, or on legal forms?

Would you say:

(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all	1
A little difficulty.....	2
Moderate difficulty	3
Extreme difficulty	4
Stopped doing this because of your eyesight	5
Stopped doing this for other reasons or not interested in doing this.....	6

A4. Because of your eyesight, how much difficulty do you have figuring out whether bills you receive are accurate?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

A5. Because of your eyesight, how much difficulty do you have doing things like shaving, styling your hair, or putting on makeup?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

SUBSCALE: DISTANCE VISION

A6. Because of your eyesight, how much difficulty do you have recognizing people you know from across a room?

(READ CATEGORIES AS NEEDED)

(Circle One)

- No difficulty at all 1
- A little difficulty..... 2
- Moderate difficulty 3
- Extreme difficulty 4
- Stopped doing this because of your eyesight 5
- Stopped doing this for other reasons or not
interested in doing this 6

- A7. Because of your eyesight, how much difficulty do you have taking part in active sports or other outdoor activities that you enjoy (like golf, bowling, jogging, or walking)?
(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all	1
A little difficulty.....	2
Moderate difficulty	3
Extreme difficulty	4
Stopped doing this because of your eyesight	5
Stopped doing this for other reasons or not interested in doing this.....	6

- A8. Because of your eyesight, how much difficulty do you have seeing and enjoying programs on TV?
(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all	1
A little difficulty.....	2
Moderate difficulty	3
Extreme difficulty	4
Stopped doing this because of your eyesight	5
Stopped doing this for other reasons or not interested in doing this.....	6

SUBSCALE: SOCIAL FUNCTION

- A9. Because of your eyesight, how much difficulty do you have entertaining friends and family in your home?
(READ CATEGORIES AS NEEDED)

(Circle One)

No difficulty at all	1
A little difficulty.....	2
Moderate difficulty	3
Extreme difficulty	4
Stopped doing this because of your eyesight	5
Stopped doing this for other reasons or not interested in doing this.....	6

SUBSCALE: DRIVING

A10. [This items, “driving in difficult conditions”, has been included as item 16a as part of the base set of 25 vision-targeted items.]

SUBSCALE: ROLE LIMITATIONS

A11. The next questions are about things you may do because of your vision. For each item, I'd like you to tell me if this is true for you all, most, some, a little, or none of the time.
(READ CATEGORIES AS NEEDED)

(Circle One On Each Line)

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a. <u>Do you have more help</u> <u>from others because of</u> <u>your vision?.....</u>	1	2	3	4	5
b. <u>Are you limited in the</u> <u>kinds of things you can do</u> <u>because of your vision?.</u>	1	2	3	4	5

SUBSCALES: WELL-BEING/DISTRESS (#A12) and DEPENDENCY (#A13)

The next questions are about how you deal with your vision. For each statement, please tell me if it is definitely true, mostly true, mostly false, or definitely false for you or you don't know.

(Circle One On Each Line)

	Definitely True	Mostly True	Not Sure	Mostly False	Definitely False
A12. I am often <u>irritable</u> because of my eyesight.....	1	2	3	4	5
A13. I <u>don't go out of my home</u> <u>alone</u> , because of my eyesight.	1	2	3	4	5

Appendix - 2

A2.1 Wearable Display

The target display is the Epson Moverio BT-35E binocular see-through smart glasses (figure A2.1). These provide the wearer with a binocular 23-degree field of view display equivalent to a 40inch television display at 2.5 meters.



Figure A2- 1 BT35E Headset. The headset is identical to the Epson BT350; the difference is the display interface. The BT350 has a small handset containing a mobile processing unit, and the BT35E has an HDMI connection for interfacing with a separate computer.



Figure A2- 2 BT-350 Handset – this connects to the same display in figure A2.1 and contains the processing unit and battery.

A similar display by the same manufacturer has been used successfully with the visually impaired, Epson Moverio BT-200 (I. R. Wilson et al., 2016), with a similar form factor and display size. An essential and unique feature of the BT-35E version is controlling the display from any computer via an HDMI connection.

Display Specifications:

- Silicon – Organic Light Emitting Diode.
- 0.413-inch panel (16:9 aspect ratio)
- Resolution: 1280 x 720
- Approximately 23 degrees field of view. 30Hz refresh rate.
- Headset weight 119 grams.

A2.2 Control Computer

Most wearable displays can only be controlled from a mobile handset provided by the manufacturer (including the other Epson devices BT-200, BT-300 and BT-350), which can restrict development. Typically, these devices have a low-powered mobile CPU. While capable of running complex computer vision algorithms, a significant amount of time and effort is required in optimisations to achieve real-time performance. This optimisation step can significantly slow the development.

The use of a high-performance laptop to control the display allows for a more efficient development process; algorithms can be quickly assessed for suitability without focusing on optimisation.

The controlling computer used; Apple MacBook Pro, 15-inch 2017. Intel i7. All image processing was implemented in C++. The display renderer used is OpenGL 4.1.

A2.3 Camera

High resolution camera: Econ Systems See3Cam_130

Capable of the following resolutions and framerates (fps).

- 4208x3120 @ 20fps
- 4K Cinema (4096 x 2160) @30fps
- 4K Ultra HD (3840x2160) @30fps
- Full HD (1080p) @ 60fps

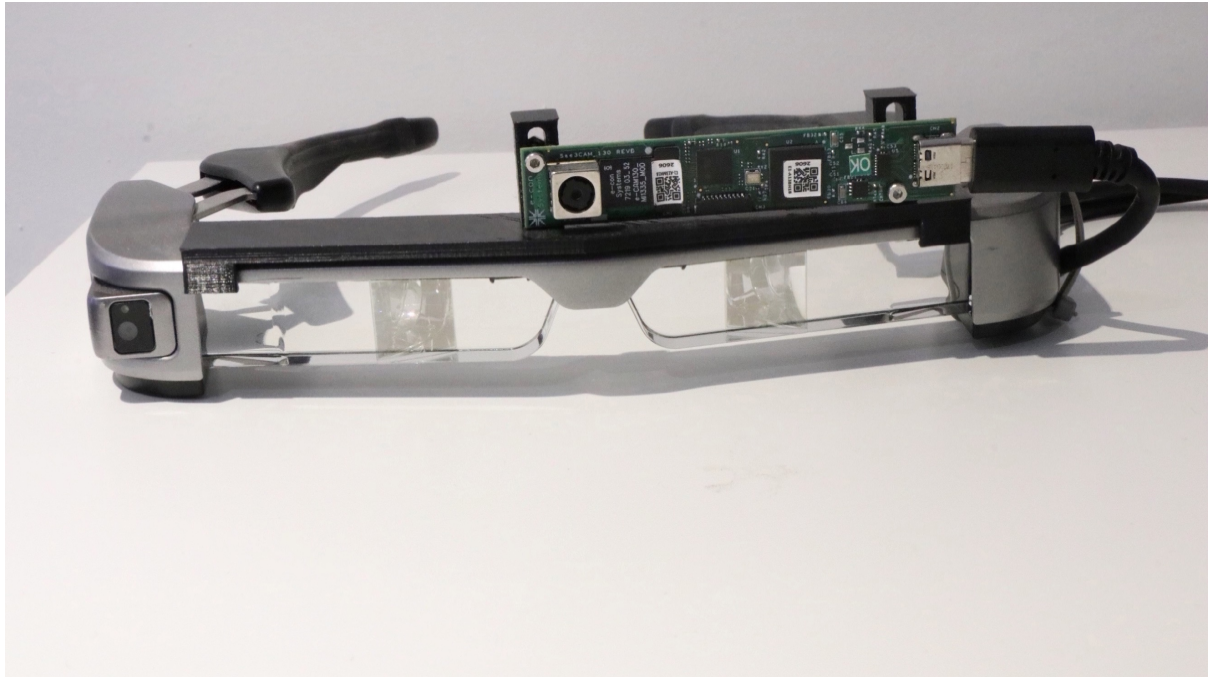


Figure A2- 3 BT35E headset with 3D printed mount and high-resolution Econ Systems camera.

A2.4 Latency reduction

I force a minimum time delay within the main display loop to best reduce the system's perceived latency. The display loop is the code that is repeatably called to update the display with an image from the camera.

This is a simple display loop in pseudocode;

```
1. While(true){  
2.     Image = read_from_camera();  
3.     Enhanced_Image = process_images(Image);  
4.     render_to_display(Enhanced_Image);  
5. }
```

The display loop is run continuously for as long as the application is running.

The function "process_images()" is where all of the image enhancement work is performed. It accepts an image captured from the camera and returns an enhanced image ready for display.

If this function takes a long time to complete, it will impact the latency of the system. The target hardware has a minimum delay of approximately 33 milliseconds, the fastest this loop can run (30Hz). If the function exceeds this 33ms time, this will also reduce the display's frequency or frame rate, waiting for the next display cycle.

Typically, complex computer vision and image processing algorithms are not deterministic, meaning that it is hard to predict the amount of time they will take to complete.

Experimentally I can state an average or an expected time to end, but there may be a significant number of slower cases. Any periodic slowdowns in the above code example will negatively impact the wearer, causing a visual stutter.

The proposed system must prioritise displaying an image to the screen. In its simplest form, the system just presents an image of the scene onto the display. This is the priority, and if any extra image processing – the face enhancement, for example – should be secondary. The hypothesis is that a slight delay in applying an image enhancement is more tolerable than a delay in displaying the whole scene.

```
1. process_image(Image){  
2.     Enhanced_Image = apply_image_enhancement(Image)  
3.     Return Enhanced_Image  
4. }
```

This approach prioritises displaying the image to the display, ensuring that the framerate never drops below a predefined value.

```
1. process_images(Image){  
2.     start_image_enhancement(Image)  
3.     wait(maximum_wait_time);  
4.     if(enhancement_is_complete)  
5.         return Enhanced_image;  
6.     else  
7.         return Image  
8. }
```

A2.5 Face detection

I choose four available for use in the OpenCV (Itseez, 2015) and Dlib (King, 2009) open-source software packages. Each was tested using the Epson Moverio BT35E camera; performance can be summarised as below.

- 1) OpenCV Harr cascade (Viola-Jones (Viola & Jones, 2004))
 - a. Pros:
 - i. Very efficient (fast)
 - ii. Detect faces at different scales
 - b. Cons:
 - i. High False Positive rate
 - ii. Fails on non-frontal faces
 - iii. Fails under occlusion

2) OpenCV DNN - Deep learning based on a Single Shot Multi-box Detector (Liu et al., 2016)

- a. Pros:
 - i. Detects faces at different scales
 - ii. Detects faces with different poses
 - iii. Robust to occlusion
- b. Cons:
 - i. Not as fast as Harr cascade

3) Dlib HoG and SVM based (King, 2009)

- a. Pros
 - i. Very efficient (fast)
 - ii. Slightly robust to different poses (frontal and slightly non-frontal)
 - iii. Robust to small occlusions
- b. Cons:
 - i. Fails for side-faces
 - ii. Fails under substantial occlusion

4) Dlib CNN based on a Max-Margin Object Detector (King, 2015)

- a. Pros:
 - i. Robust to different face orientations
 - ii. Robust to occlusion
- b. Cons:
 - i. Slow on CPU (fast on GPU)
 - ii. Fails on small faces

Of the candidate face detectors, the best performing on the experimental setup was the OpenCV DNN. This reliably detected faces, and notably, it performed quickly, less than 30ms.

The execution time of the OpenCV DNN is independent of image size, where the Harr and HoG implementations execution time increases with image size. This does mean that the HoG and Harr methods can be faster on smaller images. However, the target execution time is 30Hz (33.3ms). The OpenCV DNN approach achieves this execution time and has the best performance in terms of accuracy. This allows for flexibility with larger images and more reliable detection of small faces.

Appendix – 3

A3.1 Optotype Conversions

A3.1.1 M Units

Snellen defined an optotype for visual acuity measurement, he defined the unit M as the size of his base type, where an optotype of 1M subtends 5 minutes of arc at 1 meter. Visual acuity calculations are all based on this unit.

The minimum angle of resolution (MAR) can be calculated by measuring the height of the optotype used (h) and the viewing distance (d), as shown in the equation below.

Eq 3.1)

$$\text{MAR} = \theta = 2\arctan\left(\frac{h}{2d}\right)$$

Snellen type M is font size that subtends 5 minutes of arc at 1 meter, and the gap i.e. the hole in a letter “c” subtends 1 minute of arc at 1 meter. Thus, the height of the Snellen M optotype, at 1 meter is 1.454mm calculated as below.

Eq 3.2)

$$\text{height} = 2d \tan\left(\frac{\theta}{2}\right)$$

$$M = \frac{\text{height}}{1.454}$$

A3.1.2 Computer display font size

The size of font on a digital display is often defined as the DTP point (Desktop Publishing point), which is defined as 1/72 of an inch equivalent to 0.3527mm.

Eq 3.3)

$$M = \frac{DTP * F}{1.454}$$

Eq 3.4)

$$DTP = \frac{M * (1.454)}{F}$$

Using Equation 3.4 and 0.3527 as a scaling factor F, I can convert the Snellen M optotype size to a value that is interpreted by the computer, DTP. This shows that 1M is equivalent to approximately 4 DTP.

A3.2 Virtual Optotype Size

The main challenge with implementing any optotype based vision test within a virtual, or augmented environment is the precise measurement of the visual angle of any rendered text.

To correctly draw an optotype on a traditional computer monitor a conversion from DTP to M requires an extra step; the physical size of the display pixels, or dot pitch. Knowledge of the physical pixel size may not always be known, and the requested DTP point size may not always appear as expected due to complexities in the rendering steps. However, the relationship between DTP and Snellen M can be confirmed with a simple physical measurement of the rendered optotype on the screen.

This approach is not possible with a virtual reality display. To overcome this, I performed the measurement with a camera, capturing an image of the rendered text through the headset optics.

It is possible to estimate the size of an object from an image if the magnification factor of the camera lens and distance to the object is known.

The rendered height of the font within the headset was measured using a fixed focus camera, pointed at the right eye display. The camera used was a Point Grey FL3-U3-32S2C-CS USB3, with a Fujinon YV2.8x2.8SA-2 vari-focal lens. The camera was calibrated using a standard calibration pattern (Zhang, 1999), which provide the conversion from pixels to height and corrected for radial lens distortions.

The camera sensor pixel size was obtained from the camera datasheet (2.5um). The magnification factor (Mag), which is the ratio of the image size (h_i) to the object size (h_o) was obtained using equation below.

Eq 3.5)

$$Mag = \frac{h_i}{h_o}$$

$$h_i = \text{image height (pixels)} * \text{pixel size (mm)}$$

A3.2.1 HTC Vive

Using equation 3.5 experimentally the inverse of the magnification factor of the camera was found to be 348.96. The inverse is shown here as it is simpler to present.

The text drawn in the VR display was set to 41-point size at a distance of 1 meter. The height of this text was measured to be 25 pixels. Rearranging and applying equation 5) we can calculate the object height:

Eq 3.6)

$$h_o = \frac{h_i}{M}$$

$$h_o = h_i M^{-1}$$

$$h_o = (25 * 0.0025) * 348.96^{-1}$$

$$h_o = 21.55mm$$

I can now generate a new scaling factor, used by equation 3.4 to convert the VR DTP to Snellen M, which is 0.532, giving a Snellen M of 15.

A3.2.2 Epson BT350

Using equation 3.5, and a reference of known height, the inverse of the magnification factor of the camera was found to be 311.2.

Note, this value is different from the HTC Vive as the lens used has a variable focal length, and was set slightly differently for each experiment.

Capturing through the headset, the magnification factor was 305.1. The slight difference in the magnification factor is negligible, as this equates to approximately one-pixel difference. This result confirmed the manufactures claims, that the displays are able to provide 6/6 levels of resolution, 5 minutes of arc at one meter.

A3.3 Critical Character Count

Calculations used to define the critical character count on any given HMD.

A single character of logMAR 0 has a width equal to 5 minutes of arc, 0.083 degrees or 0.0014 radians.

Atilgan et al (Atilgan et al., 2020) used the following calculation to define the average character width;

Eq 3.7)

$$\text{Average character width} = 0.00145 * 10^P K$$

Where P is the visual angle in logMAR of the text, K is the ratio of text width to height; 0.96 for Arial and 1.33 for Courier.

For the Epson BT350 display, the character count per line is expressed as:

Eq 3.8)

$$N = \frac{w}{0.00145 * 10^P K}$$

Where w is the angular width in radians = 0.35 (20 degrees)

Rearranging if the number of characters is known, N the angular print size P can be calculated.

Eq 3.9)

$$P = \log_{10} \left(\frac{w}{0.00145 K N} \right)$$

For the Arial font the BT350 display, $w=0.35$, $k=0.96$ (Arial) and $N=9$ has an angular font size $P = 1.45$, and for the courier font ($k=1.33$) $P = 1.30$.